

*Handbooks in
Finance*



HANDBOOK of SPORTS and LOTTERY MARKETS

Donald B. Hausch
William T. Ziemba
Editors



North-Holland

DOWNLOAD FULL VERSION HERE: <https://bettingbooks.org/12>

**HANDBOOK OF SPORTS
AND LOTTERY MARKETS**

Download More Sports Betting Books at: www.bettingbooks.org

HANDBOOKS IN FINANCE

Series Editor

WILLIAM T. ZIEMBA

Advisory Editors

**KENNETH J. ARROW
GEORGE C. CONSTANTINIDES
B. ESPEN ECKBO
HARRY M. MARKOWITZ
ROBERT C. MERTON
STEWART C. MYERS
PAUL A. SAMUELSON
WILLIAM F. SHARPE**



ELSEVIER

AMSTERDAM • BOSTON • HEIDELBERG • LONDON
NEW YORK • OXFORD • PARIS • SAN DIEGO
SAN FRANCISCO • SINGAPORE • SYDNEY • TOKYO

North-Holland is an imprint of Elsevier



HANDBOOK OF SPORTS AND LOTTERY MARKETS

Edited by

Donald B. Hausch
University of Wisconsin, Madison

William T. Ziemba
University of British Columbia
Oxford University
University of Reading



ELSEVIER

AMSTERDAM • BOSTON • HEIDELBERG • LONDON
NEW YORK • OXFORD • PARIS • SAN DIEGO
SAN FRANCISCO • SINGAPORE • SYDNEY • TOKYO

North-Holland is an imprint of Elsevier



North-Holland is an imprint of Elsevier

Radarweg 29, PO Box 211, 1000 AE Amsterdam, The Netherlands
Linacre House, Jordan Hill, Oxford OX2 8DP, UK
30 Corporate Drive, Suite 400, Burlington, MA 01803, USA
525 B Street, Suite 1900, San Diego, California 92101-4495, USA

Copyright © 2008, Elsevier B.V. All rights reserved.

No part of this publication may be reproduced or transmitted in any form or by any means, electronic or mechanical, including recording, photocopying, or otherwise, without the prior written permission of the publisher.

Permissions may be sought directly from Elsevier's Science & Technology Rights Department in Oxford, UK: phone (+44) (0) 1865 843830; fax (+44) (0) 1865 853333; email: permissions@elsevier.com. Alternatively you can submit your request online by visiting the Elsevier Web site at www.elsevier.com/locate/permissions, and selecting *Obtaining permission to use Elsevier material*.



Recognizing the importance of preserving what has been written, Elsevier prints its books on acid-free paper whenever possible.

British Library Cataloguing in Publication Data

A catalogue record for this book is available from the British Library.

Library of Congress Cataloging-in-Publication Data

Hausch, Donald B.

Handbook of sports and lottery markets / edited by Donald B. Hausch.
p. cm. – (Handbooks in finance)

Includes bibliographical references and index.

ISBN 978-0-444-50744-0 (hardcover)

1. Sports betting. 2. Lotteries. 3. Gambling systems. I. Ziembra, W. T. II. Title.
GV717.H38 2008
796–dc22

2008012459

ISBN: 978-0-444-50744-0

For information on all Elsevier publications,
visit our Web site at www.books.elsevier.com.

Printed in the United States of America

08 09 10 11 12 13 9 8 7 6 5 4 3 2 1

Working together to grow
libraries in developing countries

www.elsevier.com | www.bookaid.org | www.sabre.org

ELSEVIER

BOOK AID
International

Sabre Foundation

To the memory of my father, Robert C. Hausch, who would have enjoyed this volume.

—D. B. H.

To my snow princess and her wonderful development into an outstanding researcher of the world's economic and financial activities.

—W. T. Z.

DOWNLOAD FULL VERSION HERE: <https://bettingbooks.org/12>

This page intentionally left blank

Download More Sports Betting Books at: www.bettingbooks.org

Contents

<i>List of Contributors</i>	<i>xvii</i>
<i>Preface</i>	<i>xix</i>
<i>Introduction to the Series</i>	<i>xxiii</i>
Part I Industry Studies	1
1 Pari-Mutuel Horse Race Wagering—Competition from Within and Outside the Industry	3
<i>Richard Thalheimer and Mukhtar M. Ali</i>	
1. Introduction	4
2. Competition from Casino Gaming	5
3. Competition from State Lotteries	7
4. Competition from Professional Sports	8
5. Competition from Live Racing	8
6. Competition from Simulcast Wagering	11
7. Summary and Conclusions	13
<i>References</i>	14
2 Modeling Money Bet on Horse Races in Hong Kong	17
<i>John Bacon-Shone and Alan Woods</i>	
1. Introduction	18
2. Variables examined	18
2.1. Outcome Variables	18
2.2. Independent Variables	19
3. Results and Discussion	19
4. Conclusion	21
<i>References</i>	22
Appendix: 31 Independent Variables Examined (Excluding Quadratic Terms)	23

Part II Utility, Probability, and Pace Estimation	25
3 Empirical Evidence on the Preferences of Racetrack Bettors	27
<i>Bruno Jullien and Bernard Salanié</i>	
1. Introduction	28
2. Some Stylized Facts	30
3. Expected Utility	33
4. Distortions of Probabilities	36
5. Reference Points and Asymmetric Probability Weights	39
6. Heterogeneous Preferences	42
7. Exotic Bets	45
8. Concluding Remarks	46
<i>References</i>	47
4 Approximating the Ordering Probabilities of Multi-Entry Competitions by a Simple Method	51
<i>Victor S. Y. Lo and John Bacon-Shone</i>	
1. Introduction	52
2. Theoretical Results of the Limiting Cases	53
3. A Simple Approximation	56
3.1. Empirical Analysis for the Approximated Henery Model	56
3.2. Empirical Analysis for the Stern Model	57
4. Conclusion	59
<i>References</i>	60
Appendix: Proof of Theorem 1	61
5 Modeling Distance Preference and Pace Character in Thoroughbred Turf Racing	67
<i>David Edelman</i>	
1. Background	68
2. Case Study 1: Sha Tin (Hong Kong, SAR, PRC)	69
3. Case Study 2: Randwick (Sydney, Australia)	71
4. Qualitative Questions	77
4.1. Do Distance Specialists Exist?	77
4.2. Pace, Class, and Time: The Central Paradox of Racing	78
4.3. Jockeys: Distance or Pace Preference?	79
5. Discussion	80
<i>References</i>	80

Part III Favorite-Longshot Bias in the Win Market **81**

6 The Favorite-Longshot Bias: An Overview of the Main Explanations **83**

Marco Ottaviani and Peter Norman Sørensen

1. Introduction	84
2. Notation	87
3. Misestimation of Probabilities	87
4. Market Power by Informed Bettors	89
5. Preference for Risk	90
6. Heterogeneous Beliefs	93
7. Market Power by Uninformed Bookmaker	95
8. Limited Arbitrage	96
9. Simultaneous Betting by Insiders	97
10. Timing of Bets	99
10.1. Early Betting	99
10.2. Late Betting	99
References	100

7 Examining Explanations of a Market Anomaly: Preferences or Perceptions? **103**

Erik Snowberg and Justin Wolfers

1. Introduction	104
2. Preferences—Expected Utility Models with Linear Probabilities	108
3. Perceptions—The Weighting of True Probabilities	111
4. Perceptions—Informational Effects	112
5. Definition of Models and Implications for Combinatoric Bets	114
6. Using Combinatoric Markets to Test the Models	119
6.1. Testing Conditional Independence	123
6.2. Relaxing Conditional Independence Further	125
7. Conclusion	129
References	130
Appendix A: Pricing of Combinatoric Bets Using Conditional Independence	134
Appendix B: Data	135

8 Unifying the Favorite-Longshot Bias with Other Market Anomalies **137**

Russell S. Sobel and Matt E. Ryan

1. Introduction	138
2. Biases Found in the Previous Literature	138
3. What Causes the Favorite-Longshot Bias at the Racetrack?	140
3.1. The Casual Bettor	143

3.2. <i>The Serious or Regular Bettor</i>	144
3.3. <i>The Arbitrageur</i>	148
4. Is it Risk or Information?	150
5. Can the Model Explain the Biases in Other Markets?	156
6. Conclusion	157
<i>References</i>	158
9 The Favorite-Longshot Bias in S&P 500 and FTSE 100 Index Futures Options: The Return to Bets and the Cost of Insurance	161
<i>Robert G. Tompkins, William T. Ziemba, and Stewart D. Hodges</i>	
1. Introduction	162
2. Methodology	166
3. Results	170
3.1. <i>Results for Quarterly Options on Stock Index Futures</i>	174
3.2. <i>Results for Monthly Options on Stock Index Futures</i>	175
4. Conclusions	178
<i>References</i>	179
Part IV Weak Market Efficiency	181
10 Efficiency of Racing, Sports, and Lottery Betting Markets	183
<i>William T. Ziemba</i>	
1. Introduction	184
2. Extent of Gambling in the U.S.	187
3. Racetrack Betting Markets	190
3.1. <i>Introduction to Racetrack Betting</i>	190
3.2. <i>Win Market</i>	190
3.3. <i>Place and Show Markets</i>	195
3.4. <i>Place and Show Probabilities</i>	196
3.5. <i>Optimal Capital Growth</i>	201
3.6. <i>Implementing the System and Empirical Results</i>	203
3.7. <i>Does the System Still Provide Profits?</i>	206
3.8. <i>Exotic Markets</i>	207
3.9. <i>Cross-Track Betting</i>	208
4. The Football Betting Market	209
5. The Basketball Betting Market	211
6. Lotteries	212
6.1. <i>Introduction to Lotteries</i>	212
6.2. <i>Inefficiencies with Unpopular Numbers</i>	214
<i>References</i>	217

Contents

xi

11 Point Spread and Odds Betting: Baseball, Basketball, and American Football 223

Hal S. Stern

1. Introduction 224
2. Efficiency of Odds Betting Markets 224
 - 2.1. Horse Race Betting 224
 - 2.2. Baseball 226
3. Efficiency of Point Spread Betting Markets 227
 - 3.1. American Football 227
 - 3.2. Basketball 230
4. Relationship of Point Spread and Odds Betting 230
 - 4.1. Normal Distribution Result 231
 - 4.2. Applications of the Normal Approximation 233
5. The Normal Model and Mid-Event Wagering 234
6. Summary 236
- References 237

12 Comparing Efficiency of the Over/Under-Bets on NFL and NBA Games 239

Joseph Golec and Maurry Tamarkin

1. Introduction 240
2. The Sports Betting Market: Setting Point Spreads and Over/Unders 240
3. NFL and NBA Betting Market Efficiency 242
4. Conclusion 249
- References 251

13 Arbitrage and Risk Arbitrage in Team Jai Alai 253

Daniel Lane and William T. Ziemba

1. Introduction 254
2. The Arbitrage 258
3. Risk Arbitrages 261
4. Final Remarks 269
- References 271

Part V Semi-Strong Form Efficiency 273

14 Semi-Strong Form Information Efficiency in Horse Race Betting Markets 275

M. Sung and J. E. V. Johnson

1. Introduction 276
2. Semi-Strong Form Efficiency in Horse Race Betting Markets:
Single Variable Models 277
 - 2.1. Arbitrage Between Parallel Markets 278
 - 2.2. Professional Predictions 285

2.3. <i>Betting Volume</i>	289
2.4. <i>Post Position</i>	290
2.5. <i>Pedigree</i>	291
2.6. <i>Distance Preference</i>	292
2.7. <i>Single Variable Models: Overview</i>	292
3. <i>Semi-Strong Form Efficiency in Horse Race Betting Markets: Multiple Variable Models</i>	293
3.1. <i>Distribution-Based Methods</i>	294
3.2. <i>Distribution-Free Methods</i>	298
3.3. <i>Multiple Variable Models: Overview</i>	300
4. <i>Semi-Strong Form Efficiency in Horse Race Betting Markets: Conclusion</i>	301
<i>References</i>	302
15 The Dosage Breeding Theory for Horse Racing Predictions	307
<i>Marshall Gramm and William T. Ziemba</i>	
1. <i>Introduction</i>	308
2. <i>The Racetrack as a Sequence of Markets</i>	310
3. <i>The Dosage Index and Performance Measures</i>	311
4. <i>Data Acquisition</i>	314
5. <i>Application of Breeding Information and Performance Measures to Refine Estimated Win Probabilities for the Kentucky Derby</i>	314
6. <i>The Kelly Betting Model</i>	320
7. <i>The Kentucky Derby, 1981–2007</i>	322
8. <i>The Preakness Stakes, 1946–2006</i>	327
9. <i>The Belmont Stakes, 1946–2006</i>	329
10. <i>Conclusions</i>	330
<i>References</i>	331
<i>Appendix A: Data Sources</i>	333
<i>A.1 Public's Wagering</i>	333
<i>A.2. Pedigrees</i>	333
<i>A.3. Chef-de-Race Listings</i>	333
<i>A.4. Experimental Free Handicap Listings</i>	334
<i>A.5. Results of the Kentucky Derby and Major Races Prior to the Kentucky Derby</i>	334
<i>Appendix B: Kentucky Derby, Preakness, and Belmont Winners, 1946–2006</i>	334
16 Efficiency in Horse Race Betting Markets: The Role of Professional Tipsters	341
<i>Bruno Deschamps and Olivier Gergaud</i>	
1. <i>Introduction</i>	342
2. <i>The Model</i>	343
3. <i>Data</i>	345
3.1. <i>Tips and Rewards Rules</i>	345

Contents

xiii

3.2. <i>Public Information</i>	346
3.3. <i>Measuring Forecast Originality</i>	347
4. Results	348
4.1. <i>Results of Frequency Tests</i>	348
4.2. <i>Originality and Accuracy</i>	348
4.3. <i>Anti-Herding and Excess Originality</i>	350
5. Conclusion	351
<i>References</i>	351
Appendix: Proof of Equation (5)	353

Part VI Prediction Markets 355

17 Index Betting for Sports and Stock Indices 357

John Haigh and Leighton Vaughan Williams

1. Background	358
2. How Index Betting Operates	360
3. Setting Spreads	362
4. Spreads in Performance Indices	365
5. Advantageous Bets	368
6. Regulation, Taxation, and Biases in Spread Betting Markets	372
<i>References</i>	374
Appendix A	377
Appendix B	378
Appendix C	380

18 Prediction Markets: From Politics to Business (and Back) 385

Erik Snowberg, Justin Wolfers, and Eric Zitzewitz

1. Overview	386
2. The First Prediction Markets	388
3. Markets in the Lab	391
4. Current Uses of Prediction Markets: Business and Policy	393
5. Future Directions: Decision Markets	397
6. Potential Pitfalls	398
7. Conclusion	400
<i>References</i>	401

19 Betting Exchanges: A Technological Revolution in Sports Betting 403

Michael A. Smith and Leighton Vaughan Williams

1. Introduction	404
2. The Operation of Betting Exchanges	404

3. Empirical Models and Evidence Concerning Weak-Form Information Efficiency in Betting Exchanges	409
4. New Evidence on the Degree of Bias in Betting Exchange Odds	412
5. Conclusions	416
<i>References</i>	417

Part VII Soccer 419

20 Soccer Betting in Britain 421

David Forrest

1. Introduction	422
2. Development of Soccer Betting	423
3. Transactions Costs in the Fixed Odds Market	426
4. Early Study of Market Efficiency	429
5. Tipsters	431
6. Fundamental Analysis as an Aid to Soccer Betting	432
7. Technical Analysis	437
8. Sentiment	440
9. The Future of Research on Soccer Betting	443
<i>References</i>	444

21 Efficiency of Soccer Betting Odds—Evidence from a Pan-European Electronic Market 447

Stephan Kossmeier and Simon Weinberger

1. Introduction	448
2. The Soccer Betting Market	449
3. Data Description	450
4. Efficiency Tests	451
4.1. Statistical Tests	451
4.2. Economic Tests	454
5. Conclusion	455
<i>References</i>	456

Part VIII Lotteries 457

22 How to Design a Lottery 459

Ian Walker

1. Introduction	460
2. The Odds of Winning a (Pari-Mutuel) Lottery	464
3. The Expected Value Calculation	465

Contents

xv

4. Higher Moments of the Prize Distribution	467
5. Econometric Methodology, Data, and Estimates	469
6. Game Design Simulations	474
7. Conclusion	476
<i>References</i>	477
Appendix: The Expected Value Formula	479

23 The Statistics of Lotteries	481
---------------------------------------	-----

John Haigh

1. Introduction	482
2. Prize Structure and Winning Chances	485
3. Tests of Randomness	491
4. Gambler Choices	496
<i>References</i>	502

24 U.S. Lotto Markets	503
------------------------------	-----

Victor Matheson and Kent Grote

1. Introduction	504
2. Differences Between American and European Lotteries	509
3. Fungibility of Lottery Revenues	512
4. Efficiency of Lottery Markets—Part 1	516
5. Efficiency of Lottery Markets—Part 2	519
6. Efficiency of Lottery Markets—Part 3	521
7. Conclusions	522
<i>References</i>	523

Subject Index	525
----------------------	-----

Author Index	533
---------------------	-----

DOWNLOAD FULL VERSION HERE: <https://bettingbooks.org/12>

This page intentionally left blank

Download More Sports Betting Books at: www.bettingbooks.org

List of Contributors

Mukhtar Ali, Department of Economics, University of Kentucky, Lexington, KY, USA

John Bacon-Shone, Social Sciences Research Centre, University of Hong Kong, Hong Kong

Bruno Deschamps, School of Management, University of Bath, Claverton Down, Bath, UK

David Edelman, Banking & Finance Unit, University College Dublin, Blackrock, County Dublin, Ireland

David Forrest, Centre for the Study of Gambling, University of Salford, Salford, UK

Olivier Gergaud, Department of Economics, University of Reims, Champagne-Ardenne, France

Joseph Golec, Department of Finance, School of Business, University of Connecticut, Storrs, CT, USA

Marshall Gramm, Rhodes College, Memphis, TN, USA

Kent Grote, Department of Business and Economics, Lake Forest College, Lake Forest, IL, USA

John Haigh, Mathematics Department, University of Sussex, Falmer, Brighton, UK

Donald B. Hausch, School of Business, University of Wisconsin, Madison, WI, USA

Stewart D. Hodges, Finance Group, Warwick Business School, University of Warwick, Coventry, UK

J. E. V. Johnson, Centre for Risk Research, School of Management, Highfield, University of Southampton, Southampton, UK

Bruno Jullien, Toulouse School of Economics, Toulouse, France

Stephan Kossmeier, Institute for Advanced Studies/Institut für Höhere Studien (IHS), Vienna, Austria

Daniel Lane, Telfer School of Management, University of Ottawa, Ottawa, ON, Canada

Victor S. Y. Lo, Fidelity Investments, Boston, MA, USA

Victor Matheson, Department of Economics, College of the Holy Cross, Worcester, MA, USA

Marco Ottaviani, Department of Management and Strategy, Kellogg School of Management, Northwestern University, Evanston, IL, USA

Matt E. Ryan, Department of Economics, West Virginia University, Morgantown, WV, USA

Bernard Salanié, Department of Economics, Columbia University, New York, NY, USA

Michael A. Smith, Leeds Business School, Leeds Metropolitan University, Leeds, UK

Erik Snowberg, Division of Humanities and Social Sciences, California Institute of Technology, Pasadena, CA, USA

Russell S. Sobel, Department of Economics, West Virginia University, Morgantown, WV, USA

Peter Norman Sørensen, Department of Economics, University of Copenhagen, Copenhagen, Denmark

Hal S. Stern, Department of Statistics, University of California, Irvine, CA, USA

M. Sung, Centre for Risk Research, School of Management, Highfield, University of Southampton, Southampton, UK

Maurry Tamarkin, Graduate School of Management, Clark University, Worcester, MA, USA

Richard Thalheimer, Thalheimer Research Associates, Inc., Lexington, KY, USA

Robert G. Tompkins, Centre for Practical Quantitative Finance, Hochschule für Bankwirtschaft, Germany

Leighton Vaughan Williams, Betting Research Unit, Nottingham Business School, Nottingham Trent University, Nottingham, UK

Ian Walker, Department of Economics, University of Warwick, Coventry, UK

Simon Weinberger, Institute for Advanced Studies/Institut für Höhere Studien, Vienna, Austria

Justin Wolfers, Business and Public Policy Dept., The Wharton School, University of Pennsylvania, Philadelphia, PA, USA

Alan Woods, Deceased

William T. Ziemba, Sauder School of Business, UBC, Vancouver, Canada, Mathematical Institute, Oxford University, and ICMA Centre, University of Reading, UK

Eric Zitzewitz, Department of Economics, Dartmouth College, Hanover, NH, USA

Preface

This volume surveys the broad subject of sports and lotto investments. The various chapters cover many sports, such as soccer, NFL and college football, baseball, basketball, Jai Alai, and lotto markets. We do not discuss casino gambling nor the statistics of sports; rather, we focus on the financial markets associated with legal betting on sports and lotto events. All the chapters are newly written academic surveys that we commissioned for this volume. In certain areas, this volume updates to 2008 our earlier edited volume [Hausch, D. B., V. Lo, and W. T. Ziemba (HLZ), 1994, *Efficiency of Racetrack Betting Markets*, Academic Press, San Diego, CA]. That volume became not only a classic, but a cult item as it helped usher in professional racetrack betting. While small in comparison with hedge funds, the various syndicates across the world have made about \$10 billion using computerized betting strategies. This volume continues with some of the basic research behind such investment teams and the academic theory of investment in sports and lotto markets. HLZ reprinted older classic papers and complemented them with new original work. This current volume is entirely composed of newly written chapters that build on earlier papers. So, in our view, HLZ (which was reprinted in its entirety with no changes except a new preface as HLZ, 2008, 2nd ed., World Scientific, Singapore) and this volume are companion books in this field. Other books that discuss similar topics are Vaughan Williams, L., 2003, *The Economics of Gambling*, Routledge, London; Vaughan Williams, L., 2005, *Information Efficiency in Financial and Betting Markets*, Cambridge University Press, Cambridge, UK, which are highly recommended; and our trade books, Ziemba, W. T., and D. B. Hausch, 1984, *Beat the Racetrack*, Harcourt, Brace and Jovanovich, New York; Ziemba, W. T., and D. B. Hausch, 1986, *Betting at the Racetrack*, Dr Z Investments, Inc., San Luis Obispo, CA; Ziemba, W. T., and D. B. Hausch, 1987, *Dr Z's Beat the Racetrack*, William Morrow, New York.

The volume is organized into eight parts. *Part I* discusses the industry side of the racetrack and other betting markets. Ali and Thalheimer discuss the effects of competition from casinos, lotteries, professional sports, and horse racing and wagering on racetrack handle. Bacon-Shone and Woods empirically study the factors that influence both the extent of the public's wagering and its allocation across the betting pools. One factor addressed is the partial rebate sometimes available to losers of large wagers.

Part II studies the bettors and horses in a race. Jullien and Salanié survey the literature dealing with the empirical estimation of the bettor's utility function, including addressing issues such as representative bettors versus heterogeneous beliefs, and expected versus unexpected utility. Lo and Bacon-Shone devise probabilities for multientry competitions. Edelman empirically studies the running patterns of race horses, finding distance preferences and establishing pace characteristics somewhat at odds with established physiological results on optimal running.

Part III discusses the well-established favorite-longshot bias in horse racing, which is the tendency for favorites to be underbet and longshots to be overbet. Ottaviani and Sørensen present various theoretical constructs that generate this bias. Snowberg and Wolfers use massive data sets to empirically estimate the recent favorite-longshot bias in the U.S., Australia, and other locales. They argue that the anomaly is based more on perceptions than preferences. That means that bettors overestimate the chances of low probability events. They also show that extreme favorites no longer have positive expected value as Ziemba and Hausch found in 1986. This updates the studies surveyed in Ziemba and Hausch (1986); see also the updated graph in Ziemba's chapter in Part IV and Ziemba, W. T., 2004, *Behavioral Finance, Racetrack Betting and Options and Futures Trading*, Mathematical Finance Seminar, Stanford University, Palo Alto, CA. Busche, K., and C. Hall, 1988, An Exception to the Risk Preference Anomaly, *Journal of Business* 61, 337–346; Busche, K., 1994, Efficient Market Results in an Asian Setting, in HLZ; Vaughan Williams, L., and D. Paton, 1998, Why Are Some Favourite-Longshot Biases Positive and Others Negative? *Applied Economics* 30, 1505–1510 discuss reverse biases in Asia and other locales. Sobel and Ryan document a pattern of public betting that varies by the day of the week. Different levels of casual and serious bettors at the track on different days of the week can explain this variation and provides a basis for understanding the favorite-longshot bias. Tompkins, Ziemba, and Hodges show that there are similar biases in the S&P 500 and FTSE 100 index put and call options.

Part IV discusses weak form market efficiency in racing and various sports events. Ziemba discusses efficiency in racing and other sports as well as in lotto games. He describes the place and show betting system that arises because these markets are more complex than the win market. The original 1981 system, which was popularized in the trade books by Ziemba and Hausch (1984, 1986, and 1987), still basically produces profits but needs rebates to do this because, currently, so much of the public's wagers do not enter the pools until after the race has started. This is because about 87% of the typical track's handle is bet off that track by other bettors and by professional syndicates. Ziemba also discusses cross-track betting, NFL, and NBA games and provides an introduction to lotteries. The latter topic is discussed in three chapters in Part VIII. Stern studies point spread and odds betting in U.S. college and professional baseball, basketball, and football and how these two betting concepts are related. He also investigates whether point spread betting is as efficient in these sports as it is in horse racing. Also, he shows how to estimate the odds of winning midway during a game based on the current score and the original odds line. Golec and Tamarkin analyze the market for over/under bets on NFL and NBA games. Lane and Ziemba discuss pure, no risk arbitrage and risk arbitrage in team jai alai. Most of their results generalize to other sports betting games

such as those covered by betting exchanges like London's Betfair as discussed in Part VI as well as some financial market applications. The arbitrage conditions are utility free and the risk arbitrage investments are based on the Kelly capital growth log utility criterion.

Part V discusses semi-strong form efficiency where public information is added to prices. Johnson and Sung provide a comprehensive survey of this subject in various racing markets. Gramm and Ziemba discuss the application of the breeding theory called dosage to the U.S. triple crown races, the Kentucky Derby, the Preakness, and the Belmont Stakes. The key idea is that, since the horses have never raced $1\frac{1}{4}$ miles before the Kentucky Derby or $1\frac{1}{2}$ miles before the Belmont Stakes, the public does not have direct evidence of a horse's speed and stamina for these distances. However, a horse's pedigree might provide indirect evidence, particularly if the pedigree includes stallions whose offspring exhibit consistent levels of speed and stamina. This chapter studies whether this indirect pedigree evidence can be used to profitably revise the public's win odds. And indeed this is the case, with the greatest gains associated with the $1\frac{1}{2}$ mile Belmont, followed by the $1\frac{1}{4}$ mile Kentucky Derby. For the $1\frac{1}{16}$ mile Preakness, however, pedigree offered no gains. Since many horses in the Preakness have raced $1\frac{1}{16}$ miles before the Preakness (including some who ran $1\frac{1}{2}$ miles in the Derby just two weeks earlier), the odds established by the public have incorporated direct evidence of the each horse's speed and stamina for this distance, making indirect pedigree evidence of limited incremental value. Gergaud and Deschamps investigate the effectiveness of the recommended horses from tipsters in Paris.

Part VI presents three chapters related to the recent explosion of interest in betting exchanges. Haigh and Vaughan Williams discuss index betting for sports and stock indices. The idea is that the house sets an x and a y for each event where $y > x$. The difference $y - x$ is their profit and expenses and reserve for risk. Then those that go long do it at y and those who go short do it at x . If the number of longs and shorts is about equal, then the house gets the $y - x$. But if it is not, then the house bears some risk. The payoff to longs is the final price $z - y$ and to the shorts it is $x - z$. This activity is tax free in the UK, so it is a popular way to bet on index futures. The spread cost is balanced by the tax savings in regular futures accounts by investors. Snowberg, Wolfers, and Zitzewitz investigate the prediction ability of internet wagering markets in various areas including, especially, politics. The contracts available form a fascinating description of modern times from elections, to war, to sports betting, all of which induce good probabilities that are generally superior to political polls and other estimates. Putting real money on the line tends to generate good market forecasts. Finally, Smith and Vaughan Williams provide a comprehensive survey of betting exchanges. In these cases, person A is betting against person B and the house has no risk and garners its profits from a small commission on the net winning bets. Betfair in London is the largest such betting exchange offering wagers on a vast variety of contests from sports to Academy Awards to politics and other areas around the world. From a beginning in 1999, these exchanges, especially Betfair, now match millions of bets per week.

Part VII presents two chapters on the efficiency of soccer betting markets. Forrest discusses British markets and Kossmeier and Weinberger discuss Austrian markets. Soccer

is probably the world's most popular sport, and betting on games with European, Asian, South American and other punters is huge. These two papers update the reader on this key topic. Since scores are low, there are various types of exotic wagers as well.

Finally, *Part VIII* discusses lotteries. (Lotteries, especially Canadian games, are briefly discussed in the Ziemba chapter in Part IV.) Walker discusses economic issues facing UK lotto organizations. He studies where the moneys (the sales) come from based on factors such as jackpot size and rollover policies and where they go and various microeconomic analyses that can tilt the situation. His chapter provides an outline to the statistical, economic, and practical considerations to designing lottery games. Haigh, also working in the UK, presents a marvelous analysis of the statistics of lotto games. Finally, Matheson and Grote give a comprehensive survey of their work and that of others and that from the lottery organizations regarding U.S. lottos. In the U.S., the payments are usually spread over 20 years and are taxable. So, in comparison to Canada and the UK, the U.S. prizes are worth about a third as much; see Ziemba, W. T., S. L. Brumelle, A. Gautier, and S. L. Schwartz, 1986, *Dr Z's 6/49 Lotto Guidebook*, Dr Z Investments, Inc., San Luis Obispo, CA, for additional calculations as this and other lotto questions such as growth in carryovers and sales from the optimality sometimes to buy the pot, that is, buy all the numbers and profit from the winners. Hence, they are not as good a deal despite their popularity in the U.S. Lotteries remain a dream hope of poor people who yearn for those million dollar or pound payoffs.

Donald B. Hausch, Madison
William T. Ziemba, Vancouver
February 2008

Introduction to the Series

Advisory Editors:

Kenneth J. Arrow, Stanford University; George C. Constantinides, University of Chicago; B. Espen Eckbo, Dartmouth College; Harry M. Markowitz, University of California, San Diego; Robert C. Merton, Harvard University; Stewart C. Myers, Massachusetts Institute of Technology; Paul A. Samuelson, Massachusetts Institute of Technology; and William F. Sharpe, Stanford University.

The Handbooks in Finance are intended to be a definitive source for comprehensive and accessible information. Each volume in the series presents an accurate, self-contained survey of a subfield of finance, suitable for use by finance and economics professors and lecturers, professional researchers, and graduate students and as a teaching supplement. The goal is to have a broad group of outstanding volumes in various areas of finance.

William T. Ziemba
University of British Columbia

DOWNLOAD FULL VERSION HERE: <https://bettingbooks.org/12>

This page intentionally left blank

Download More Sports Betting Books at: www.bettingbooks.org

DOWNLOAD FULL VERSION HERE: <https://bettingbooks.org/12>

PART I: Industry Studies

Download More Sports Betting Books at: www.bettingbooks.org

DOWNLOAD FULL VERSION HERE: <https://bettingbooks.org/12>

This page intentionally left blank

Download More Sports Betting Books at: www.bettingbooks.org

CHAPTER 1

Pari-Mutuel Horse Race Wagering—Competition from Within and Outside the Industry

Richard Thalheimer

President, Thalheimer Research Associates, Lexington, KY, USA.

Mukhtar M. Ali

Emeritus Professor of Economics, Department of Economics, University of Kentucky, Lexington, KY, USA.

1. Introduction	4
2. Competition from Casino Gaming	5
3. Competition from State Lotteries	7
4. Competition from Professional Sports	8
5. Competition from Live Racing	8
6. Competition from Simulcast Wagering	11
7. Summary and Conclusions	13
<i>References</i>	14

Abstract

From 1960 through 2002, real dollar wagering on pari-mutuel horse racing (handle) in North America peaked in 1977, declining 44% through the end of the period. A number of factors have been identified as contributing to this decline in handle. Most significant among these are increased competition from casino and state lottery gaming venues. The presence of casino gaming in a racetrack's market area has been found to reduce pari-mutuel handle by 31–39%. Competition from casino gaming within the pari-mutuel industry takes the form of the integration of slot machines at a pari-mutuel racing facility (racino). Slot machines and video lottery terminals (VLT's), when placed under the auspices of a state lottery, have been estimated to reduce the pari-mutuel handle at a racino from 24% to 39%, varying with the number of machines. State lotteries have resulted in estimated pari-mutuel handle reductions ranging from 10% to 36%. The presence of professional sports in a racetrack's market area has also been found to reduce pari-mutuel wagering, although to a lesser extent than competition from casino gaming or a state lottery. Finally, within-industry competition from other pari-mutuel wagering venues has had a negative effect on pari-mutuel wagering.

JEL Classifications: L83

Keywords: pari-mutuel horse race wagering, pari-mutuel wagering, horse race wagering, casino, racing, lottery

1. INTRODUCTION

From 1960 through 2002, real dollar wagering (handle) on horse racing in North America peaked in 1977 before declining 44% through the end of the period.¹ A number of factors have been identified as contributing to the decline in pari-mutuel handle. Among the major causes of this decline are increased levels of competition from gaming venues such as casinos and state lotteries. There has been tremendous growth in the gaming industry over the past several decades. Non-pari-mutuel real dollar gaming handle (casino-type gaming, lottery gaming, and charitable gaming) increased 355% from 1982 through 2002.² Competition from professional sports has also contributed to the decline in pari-mutuel handle. Finally, competition from other pari-mutuel racetracks has been found to have a negative impact on individual racetrack handle. This study is an attempt to summarize the findings on how competition both within and outside the industry have affected horse race wagering. A comprehensive review of the literature is made to ascertain the effects of such competition.

¹Total horse racing handle obtained from Association of Racing Commissioners, Inc., *Pari-mutuel Racing, A Statistical Summary* (annual issues). Adjustment to real dollars using the consumer price index (CPI) obtained from the U.S. Department of Labor, Bureau of Labor Statistics.

²Christiansen and Sinclair (2001) gaming handle statistics. U.S. Department of Labor, Bureau of Labor Statistics, CPI statistics.

The chapter is organized as follows. Section 2 reports results of studies that examined the effect of competition from casino gaming on the demand for pari-mutuel wagering. Section 3 reports the results of studies that examined the effect of competition from state lotteries on the demand for pari-mutuel wagering. Section 4 reports the results of studies that examined the effect of competition from professional sporting events on the demand for pari-mutuel wagering. Section 5 reports the results of studies that examined the effects of live race competition from other racetracks on the demand for wagering at a subject racetrack. Section 6 reports the results of studies that examined the effect of competition from simulcast wagering on the demand for wagering at a subject racetrack. Section 7 presents the summary and conclusions of the chapter.

2. COMPETITION FROM CASINO GAMING

Casino gaming has been found to be a strong substitute for pari-mutuel horse race wagering. Ali and Thalheimer (1997) examined the demand for pari-mutuel horse race wagering in the presence of casino gaming. Two demand functions were estimated, one each for thoroughbred and harness horse racetracks in New Jersey over the period of 1960–1988. Casino gaming was introduced during the study period in Atlantic City, New Jersey in 1978. Initially, there was one casino, increasing to 12 by 1988, the end of the sample period. It was estimated that the presence of 12 casino gaming facilities in Atlantic City had a significant and negative impact on the demand for wagering and resulted in a 32% decrease in live and total (live plus intrastate simulcast) pari-mutuel horse race wagering at the New Jersey thoroughbred and harness racetracks.

Another estimate of the relationship between casino gaming and pari-mutuel horse race wagering is given in Thalheimer and Ali (1995a). Separate straight and exotic pari-mutuel wagering demand equations were estimated for two New Jersey thoroughbred racetracks, Atlantic City Race Course and Monmouth Park Racetrack, over the 1960–1990 sample period.³ For each racetrack, the impact of casino gaming was found to be significant and negative. Casino gaming was estimated to have reduced wagering at Atlantic City Race Course and Monmouth Park Racetrack by 31% and 39%, respectively.⁴

Thalheimer (1998) examined the relationship of casino gaming and pari-mutuel horse race wagering when both are located at the same racetrack (racino). The location was Mountaineer Racetrack and Resort, a pari-mutuel racetrack in West Virginia. This was the first racetrack in the United States to offer casino gaming on a significant scale to its patrons. In June 1990, a limited number of electronic gaming devices under the auspices of the State Lottery, referred to as video lottery terminals (VLT's), were made available, on an experimental basis, to customers of the racetrack. VLT's are a form of slot machine and are perceived as such by the customer. Three wagering

³A straight and exotic wagering demand model was also estimated for a harness racetrack in Kentucky. Since Kentucky did not have casino gaming, the results of this model are not discussed further.

⁴Evaluated at the non-zero mean of 8.9 casinos, weighted by mean values of exotic and straight handle shares.

demand models were estimated: one for live handle, one for full-card simulcast handle⁵ (wagering on simulcasts of races taken from locations outside the state), and one for VLT handle. Each of the three demand models contained its own-demand variable and also cross-demand variables from the other two. Data were taken daily over the period of 1990–1991.

VLT's were found to have a significant and negative impact on pari-mutuel wagering. At the 1991 average VLT level of 114 machines, wagering on pari-mutuel horse races was estimated to have been reduced by 24%. On the other hand, total wagering (pari-mutuel plus VLT) was estimated to have increased 21%. That is, wagering on the 114 VLT's offset the 24% reduction in pari-mutuel wagering and added an additional 21% to the handle. Total revenue or takeout (handle less payout to customers), however, increased only 4% due to a lower takeout on VLT handle relative to the pari-mutuel handle which it replaced. One conclusion of the paper was that in order to generate sufficient VLT handle to produce positive net revenue, a minimum number of VLT's is required. A second conclusion was that customers who wagered on the VLT's were not likely to bet on the horse races being offered while those who wagered on those pari-mutuel horse races were also likely to bet on the VLT's.

In an update and expansion of the Thalheimer (1998) study, Thalheimer (2008) again examined the relationship of casino gaming and pari-mutuel horse race wagering at Mountaineer Racetrack and Resort. Pari-mutuel and VLT data were taken weekly from fiscal year 1994, when VLT's were first permitted at racetracks by state law on a permanent basis, through 2002. Over this period, the number of VLT's was increased from 400 to 3,000. On-track live and simulcast pari-mutuel handle were estimated to have decreased 39% as a result of the increase in the number of VLT's. As in the earlier study (Thalheimer, 1998), VLT handle was found to have increased when the pari-mutuel product was made available. VLT handle increased 22% as a result of the presence of year-round live horse racing. Since revenue (handle less payout to customers) was \$2.6 billion for the VLT's in 2002 compared to on-track pari-mutuel revenue of \$39 million, the increase in VLT revenue from the presence of live racing was far greater than the reduction in pari-mutuel revenue from the increased number of VLT's. The number of simulcast races from other racetracks offered to Mountaineer customers was also found to have a positive effect on VLT handle. As in the earlier study Thalheimer (1998), there was little or no crossover of VLT customers to the pari-mutuel wagering product while there was a significant crossover of pari-mutuel customers to the VLT product.

The studies reviewed so far examined the impact of casino competition on pari-mutuel wagering under two situations: (1) casinos were permitted in an existing pari-mutuel wagering market but at locations other than racetracks, and (2) casino gaming, in the form of VLT's, was permitted at existing pari-mutuel wagering facilities. Not mentioned above is the effect of competition from pari-mutuel venues on casino gaming demand when both are located in the same market area. In a study of the demand for casino gaming, Thalheimer and Ali (2003) examined this relationship. The demand

⁵Full-card simulcasting of races from locations outside West Virginia was introduced in June, 1990.

for slot machine gaming was estimated for 24 riverboat casinos and three racetrack-slot machine casinos (racinos) in the states of Illinois, Iowa, and Missouri from 1991 to 1998. One of the demand factors was competition from pari-mutuel horse race wagering venues (live and simulcast locations) measured by the distance-related accessibility of riverboat/racino customers to these alternative wagering locations. Access to pari-mutuel wagering venues was found to have a negative but insignificant effect on the demand for casino slot machine gaming.

3. COMPETITION FROM STATE LOTTERIES

State lottery wagering products have been found to be strong substitutes for pari-mutuel horse race wagering. Simmons and Sharp (1987) employed an econometric model of pari-mutuel wagering using 89 of the 100 U.S. thoroughbred race meets in various U.S. counties for the year 1982. State lotteries were estimated to have resulted in a 36% loss in pari-mutuel handle.

Vasche (1990) estimated the impact of the introduction of the California State Lottery in 1985 on pari-mutuel horse race wagering. An econometric model of California pari-mutuel thoroughbred horse race wagering developed with data prior to introduction of the lottery in 1985 is employed to estimate handle from 1985 to 1990.⁶ Estimated handle was then compared to actual handle from 1985 to 1990, the period beginning with the introduction of the lottery. The California State Lottery was estimated to have reduced pari-mutuel horse race handle by 20–30%.

Thalheimer and Ali (1995c) estimated three pari-mutuel wagering demand models, one each for two racetracks in southern Ohio: Lebanon Raceway (harness) and River Downs (thoroughbred); and another for a nearby thoroughbred racetrack in Northern Kentucky: Turfway Park. The data were annual over the period of 1960–1987. The Ohio State Lottery was introduced in 1974. The effect of competition from the Lottery was found to be significant and negative. The payout rate of the Ohio Lottery at the end of the sample period was 55.2%, resulting in an estimated 27% reduction in pari-mutuel handle for each of the three racetracks.

In Thalheimer and Ali (1995b), an econometric model of pari-mutuel wagering demand was estimated for six Kentucky racetracks using daily race meet data from 1986 to 1990. The demand model included the Kentucky State Lottery (introduced April 4, 1989) as one of its determinants. Introduction of the Kentucky State Lottery was found to have a significant and negative impact on the demand for wagering at each of the racetracks whose demands were being estimated. The reduction in pari-mutuel wagering at these racetracks ranged from 10.3% to 32.6% and averaged 18.4%.⁷

⁶The model used is not given in the article.

⁷The lottery impacts were not reported in the article but were easily computed using the lottery coefficients and the value of the fully implemented lottery ($= 1.0$).

4. COMPETITION FROM PROFESSIONAL SPORTS

Professional sporting events that compete with pari-mutuel horse racing have been found to have an adverse effect on pari-mutuel horse race wagering. Coate and Ross (1974) estimated the demand for thoroughbred and harness horse wagering in New York using daily data from 1970 to 1972. Competition from three professional sports, basketball (the New York Knickerbockers), hockey (the New York Rangers), and football (NFL Monday Night Football), was estimated to have a significant and negative effect on pari-mutuel harness horse wagering.

Thalheimer and Ali (1992) estimated the demand for wagering at a pari-mutuel harness horse racetrack in Louisville, Kentucky over the period of 1970–1987. There were two race meets per year over this period. The Louisville Redbirds, a minor league professional baseball team, began operations in Louisville in 1982. The presence of the Redbirds for a given horse race meet was measured as the number of home games over that meet. There were an average of 19.1 home games per race meet from 1982 forward. The impact of competition from the presence of the professional baseball team was estimated to have resulted in a 5.3% loss in handle at Louisville Downs.

In another study by Thalheimer and Ali (1995b), one of the pari-mutuel wagering demand determinants for Turfway Park, a thoroughbred racetrack in northern Kentucky, was the presence of major league professional sports in nearby Cincinnati, Ohio on a race day. Over the estimation period, Cincinnati was home to two professional sports teams, the Cincinnati Redlegs (Reds) baseball team and the Cincinnati Bengals football team. A professional baseball game, offered on the same day as pari-mutuel wagering at the racetrack, was found to have a negative but insignificant effect on wagering there. On the other hand, the presence of a professional football game in Cincinnati was estimated to have a significant and negative (–9.7%) effect on pari-mutuel wagering at the racetrack on that day.

In Thalheimer and Ali (1995c), one determinant of the demands for wagering at two racetracks in southern Ohio and for a nearby thoroughbred racetrack in northern Kentucky was the presence of competition from professional sports in Cincinnati, Ohio. In this case, sports competition was measured as the weighted total number of days over a year when home games were available for three professional sports, football, baseball, and basketball. Attendance at these events was used to derive weights from which the weighted average number of competing days was computed. The impact of competition from professional sports on the three racetracks was found to be significant and equal. An additional 10 days of competition from professional sports in 1987 was estimated to result in a 4% reduction in handle at the three racetracks.

5. COMPETITION FROM LIVE RACING

Several studies have found that wagering at a pari-mutuel racetrack is reduced when there is increased competition within the industry from other pari-mutuel racetracks in the market area. Morgan and Vasche (1979) estimated the demand for wagering for

three Southern California thoroughbred racetracks conducting four race meets (Santa Anita, Oak Tree Racing at Santa Anita, Del Mar, and Hollywood Park) using a pooled time-series cross-section database over the period of 1958–1978. In Morgan and Vasche (1982) the dataset was updated to include two more years, 1979 and 1980. Both demand models had identical specifications, the only difference being the two-year update of the sample in the 1982 study. In each demand model (Morgan and Vasche, 1979, 1982), the presence of competition from nighttime non-thoroughbred (i.e., quarter horse and harness) racing was found to be significant and to result in reduction in pari-mutuel wagering at the Southern California thoroughbred racetracks. Another competition variable, the combined number of harness and quarter horse racing days, was not found to be statistically significant. Morgan and Vasche (1979) state that, in addition to the two non-thoroughbred competition variables included in their models (i.e., day to night harness and quarter horse racing and combined number of harness and quarter horse days), “a number of alternative variable specifications to reflect horse racing competition were considered, such as number of thoroughbred days which face overlaps with other day and night racing in Southern California. Regardless of the exact form of the variable tested, evidence of competition between different types of racing consistently appeared.”⁸

Church and Bohara (1992) examined the demand for horse race wagering in New Mexico using data for seven racetracks over the period of 1964–1988. Four of the racetracks were operating in 1964 at the beginning of the sample period, one opened in 1971, one in 1985, and the last one in 1986. Competition for a subject racetrack in a given year was measured by an interaction term of the product of the race days for the subject racetrack and the difference in total race days for all racetracks less the number of race days for the subject racetrack. Competition was found to have the expected negative sign for six of the seven racetracks, three of which were significant. Although the competition variable for the seventh racetrack was found to be positive, it was not found to be significant.

In Thalheimer and Ali (1995a), the two New Jersey thoroughbred racetracks whose demands were estimated, Atlantic City Race Course and Monmouth Park Racetrack, had overlapping dates and as a result were competitors with each other for that period. The two racetracks also faced competition from Freehold Raceway (harness), while Monmouth Park also faced competition from the nearby Meadowlands racetrack (harness meet). Competition in a given year was measured as the number of overlapping days between racetracks for that year. Competition between Monmouth Park and Atlantic City began in 1975 and averaged 64.8 days over the period of 1975–1990.⁹ As a result of this competition, straight and exotic wagering demands for the Atlantic City racetrack were estimated to have been reduced 51.4% and 43.2%, respectively. On the other hand, competition from the Atlantic City racetrack was found to have a statistically insignificant effect on exotic wagering demand at Monmouth Park. However, it did have a significant effect on straight wagering demand, reducing that demand

⁸Morgan and Vasche (1979), p. 190.

⁹See Thalheimer and Ali (1994) for means of the competition variables.

by 17.5%.¹⁰ The difference in these relative effects may be due to the difference in the quality of racing offered by the two racetracks. Monmouth Park had higher purses and thus a higher quality of racing over the estimation period.

Competition from Freehold Raceway was not found to have had a statistically significant impact on the demand for wagering at Monmouth Park. The demand for wagering at Atlantic City Race Course was found to be only slightly reduced by competition with Freehold Raceway. Exotic wagering demand was reduced 3.1% while straight wagering demand was not found to be significantly affected. Finally, straight wagering demand at Monmouth Park was not significantly impacted by competition from the harness race meet at the nearby Meadowlands, while exotic wagering demand was found to have been reduced by 12.2%.

In Thalheimer and Ali (1995b), the effect on wagering at a racetrack as a result of competition with other racetracks in its market area was estimated. In this study of four thoroughbred racetracks and two harness racetracks in Kentucky, using daily data, the competition variable took the value of one, on a day when there was wagering at a competing racetrack, otherwise it was zero. Of the six Kentucky racetracks whose demands were estimated, four faced competition from other racetracks at some time during the year. Three of the racetracks faced competition from one racetrack while one racetrack faced competition from two racetracks. Competition from four of the five competing racetracks was found to have a significant and negative effect on wagering at the racetrack whose demand was being affected. The opportunity to wager on races offered at a competing racetrack was found to result in an estimated 8.1–23.0% reduction in wagering at the subject racetrack.¹¹

Ali and Thalheimer (1997) examined the effect of competition on the aggregate of all thoroughbred racetracks and of all harness racetracks in New Jersey from racetracks located in the bordering states of Delaware, New York, and Pennsylvania. The time period of the analysis was 1960–1988. Unlike in earlier studies, where competition was measured as days of overlapping race meets, the degree of competition was measured using a specially constructed visit cost (VC) variable. Visit cost to competing racetracks was computed as the cost of a trip multiplied by the average traveling distance (ATD) to a competing out-of-state racetrack. The cost of a trip was computed as the average of the New Jersey wage cost index (opportunity cost) and the transportation component of the New Jersey consumer price index. The distance for a New Jersey pari-mutuel racetrack patron to travel to a competing out-of-state racetrack site was computed using information on the distance from each New Jersey market area population center to each competing out-of-state racetrack. Since wagering opportunities were not available at every wagering site on every day of the year, the distance, for the same consumer, could vary from day to day. For each day of the year, the traveling distance

¹⁰The significance of the competition variables was reported in Thalheimer and Ali (1995a) but not the magnitude of the impacts. The magnitude of the impacts reported here was computed from information provided in the paper.

¹¹Magnitude of impacts computed using competition variable coefficients as reported in Thalheimer and Ali (1995b).

from each New Jersey population center (14 metropolitan statistical areas) to the out-of-state wagering sites was determined. The minimum of the distances from that center to each of the out-of-state wagering sites was chosen as the traveling distance for that day. Daily distances for each population center were then averaged over the year. The population-weighted average of the average daily distances for that year was denoted as the average traveling distance (ATD) for that year.

A decrease in the visiting cost (VC) for New Jersey pari-mutuel thoroughbred or harness racetrack patrons to wager at racetracks of the same breed in Delaware, Pennsylvania, and New York, resulted in a reduction in wagering at the New Jersey racetracks. A 1% decrease in VC to out-of-state thoroughbred racetracks from its 1988 level was estimated to result in a 0.14% decrease in New Jersey thoroughbred wagering. A 1% decrease in VC to out-of-state harness racetracks was estimated to result in a 0.13% decrease in New Jersey harness wagering. Had there been no competition from out-of-state harness racetracks, the New Jersey harness handle would have been 14.3% greater.

In addition to examining the effect of own-breed, out-of-state competition on New Jersey pari-mutuel wagering, Ali and Thalheimer (1997) also estimated the effect of cross-breed competition from both in-state and out-of-state racetrack locations. A decrease in VC to thoroughbred racetracks, both inside and outside of New Jersey, was found to reduce New Jersey harness handle. Specifically, a 1% decrease in VC from its 1988 level was estimated to result in a 0.10% decrease in New Jersey harness wagering. Had there been no competition from in-state and out-of-state thoroughbred racetracks, the New Jersey harness handle would have been 10.5% greater. New Jersey thoroughbred wagering was not found to be significantly impacted by competition from harness racetracks located inside and outside of New Jersey.

6. COMPETITION FROM SIMULCAST WAGERING

There are several forms of simulcast wagering products available to pari-mutuel wagering customers. One of these is intrastate intertrack wagering (ITW), the simulcasting of the live race product offered by a racetrack, simultaneously, to other racetrack locations within the state in which that racetrack is located. It is expected that by offering ITW, total wagering on a racetrack's live race product will increase, while wagering at that racetrack's location may decrease. The expected increase in total (live plus ITW) wagering can be attributed to the reduction in the cost of attending the races due to the increase in availability of the product to customers in the market area. On the other hand, on-track wagering at the location in which the live races are being offered may decrease since the locations at which the simulcasts of the live races are received are essentially competitors of the live race site. The degree of competition with the on-track live races is expected to increase as the travel distance to alternative wagering sites to bet on those races is decreased.

In Thalheimer and Ali (1995b), the presence of ITW was measured as the weighted average traveling distance to the nearest wagering site (including the racetrack conducting the live races). The weights used to compute the weighted average distance are the income for each of the eight major population centers from which a racetrack customer must travel to attend and bet on the races, divided by total income of all the population centers. As the number of ITW racetrack sites is increased, the weighted average traveling distance is decreased. The demands for wagering were estimated for four thoroughbred and two harness racetracks in Kentucky. It was found that ITW had an insignificant effect on the on-track handle for four of the six racetracks, while handle decreased 5% for one thoroughbred racetrack and 8% for one harness racetrack. On the other hand, as expected, total (live plus ITW) wagering at all six racetracks was found to increase. The increase in the total handle ranged from 12% to 77% over all six racetracks.

The introduction of ITW may also create new cross-breed competition for existing racetracks. For example, thoroughbred (harness) racetracks may face increased competition from cross-breed harness (thoroughbred) ITW wagering sites. In Kentucky, where the demand for the harness racing product is not as strong as for the thoroughbred racing product, the cross-breed competition from ITW harness racing on thoroughbred wagering was not found to be significant. On the other hand, the demand for harness wagering was reduced by 4% at one harness racetrack (Red Mile) and by 24% at the other harness racetrack (Louisville Downs) due to competition from ITW thoroughbred racing.

Another form of simulcast wagering is wagering on simulcasts of live races at off-track betting sites (OTB's) which, themselves, do not offer live pari-mutuel racing. Coate and Ross (1974) estimated the effect of the introduction of off-track betting in New York City in 1971 on live pari-mutuel thoroughbred and harness horse wagering there. Off-track betting was found to have a significant and negative effect on both thoroughbred and harness horse wagering.

Yet another form of simulcast wagering is referred to as full-card (whole-card) simulcasting. A customer at an in-state wagering location is able to wager on entire live race programs from one or more out-of-state racetracks that simulcast those races to the in-state location. In a recent study (Ali and Thalheimer, 2002), the wagering demands of a number of racetracks whose simulcast races were imported at a single racetrack location and offered to patrons there, along with live racing conducted at the racetrack, was estimated. The subject of the analysis was Garden State Park, a New Jersey racetrack that offered wagering on its own live thoroughbred and harness horse racing, and on simulcast racing from both in-state and out-of-state locations. The study period covered each day that live, simulcast, or a combination of live and simulcast racing was offered in 1995. Over this period, there were 49 racetracks whose simulcast programs were taken by Garden State Park at various times over the year. In addition, Garden State Park offered its own live thoroughbred and harness horse race meetings. In total, 51 racetrack products were offered, 49 of which were simulcasts of live race products from other locations to Garden State Park and two of which were live race products conducted at Garden State Park. To overcome data and statistical limitations, the 51 racetracks were aggregated by geographical location into eight thoroughbred, simulcast

racetrack groups, five simulcast harness racetrack groups, one live thoroughbred group, and one live harness group, for a total of 15 racetrack groups.

The wagering demand for each racetrack group was specified as a function of its own number of races, takeout rate, average purse per race, and average field size per race. In addition, this demand was taken as a function of the number of races offered, takeout rate, average purse per race, and average field size per race for the racetrack groups with which it competed on a given day. A number of other product specific characteristic variables for different racetrack groups were also included.

Wagering on a racetrack group's races was found to decrease as a result of competition from competing racetrack groups on a given day. The number of races offered by competing racetrack groups was found to have a significant and negative effect on wagering for nine of the 15 racetrack groups. The median number of races' cross-elasticity for those racetrack groups where number of races was found to be significant was -0.67 . Field size associated with competing racetrack groups was found to have a significant and negative effect on wagering for six of the 15 racetrack groups. The median field size cross-elasticity for those racetrack groups where field size was found to be significant was -1.07 . While average purse per race of competing racetracks was found to have a significant and negative impact on wagering, this was true for only three of the 15 racetrack groups. Thus, wagering demand for a particular racetrack group is not much affected by average purse per race of competing racetrack groups on a given day. The takeout rate associated with competing racetrack groups was also found to have a significant positive effect on wagering for five of the 15 racetrack groups, indicating that the competing racetrack groups were substitutes for a particular racetrack group whose demand was being determined. The median takeout rate cross-elasticity for those racetrack groups where takeout rate was found to be significant was 1.07 .

The presence of live thoroughbred or harness racing at Garden State Park among the competing racetrack groups was found to have a statistically significant effect for only three of the racetrack groups. In those instances where the presence of live racing was found to be significant, it was a substitute for same-breed and a complement for cross-breed racetrack groups. The fact that offering thoroughbred (harness) live races on a given day results in higher wagering on harness (thoroughbred) simulcast racing can possibly be explained by increased attendance of New Jersey harness (thoroughbred) horsemen on those days. These horsemen are familiar with all New Jersey (harness and thoroughbred) racing and may bet both breeds when attracted to the betting site on days when there is betting offered on their own-breed New Jersey live race product.

7. SUMMARY AND CONCLUSIONS

Competition from casino gaming offered at locations outside pari-mutuel horse racing sites was found to have a significant and negative impact on pari-mutuel horse race wagering. Casino gaming in Atlantic City, New Jersey was estimated to have resulted in a 32% reduction in wagering over all New Jersey thoroughbred and harness racetracks.

On an individual racetrack basis, casino gaming in Atlantic City was found to have resulted in a 31% decline in wagering at Atlantic City Race Course and a 39% decline in wagering at Monmouth Park Racetrack.

Competition from casino gaming within the pari-mutuel industry takes the form of integration of slot machines (VLT's) at a pari-mutuel racing and wagering facility (racino). Such devices are installed at the facility to broaden its product line. The introduction of casino gaming in the form of the placement of VLT's at a racetrack was found to reduce pari-mutuel wagering by 24%, at a level of 114 VLT's and 39% at a level of 3,000 VLT's. The VLT handle was found to have increased as much as 22% as a result of the presence of year-round live horse racing. At the level of 3,000 VLT's, the increase in VLT revenue from year-round live horse racing was found to be far greater than the decrease in pari-mutuel revenue from the VLT's. Offering simulcast races from other racetracks was also found to have a positive effect on VLT handle.

The impact of state lottery gaming on pari-mutuel horse race wagering was estimated for several individual state lotteries (Kentucky, Ohio, and New Jersey) as well as for the lotteries in the U.S. as a whole. The reduction in pari-mutuel wagering due to competition from state lotteries was estimated to range from 10% to 36%. With respect to non-gaming competition, a professional sporting event, held on the same day that wagering is conducted at a pari-mutuel racetrack, was found to have a significant and negative impact on wagering at that racetrack.

Within-industry competition between racetracks conducting live race meets was found to result in significant reductions in wagering in many cases. With respect to simulcast wagering, in several cases the introduction of intrastate intertrack simulcast wagering was found to result in reduced on-track wagering at the site conducting the live races while simulcasting those races to other racetrack sites in the state. On the other hand, total pari-mutuel handle (live plus simulcast) for the racetrack was estimated to increase in every case. Where more than one breed of racehorse conducts racing in a state (e.g., thoroughbred and harness), the introduction of ITW in that state may create new cross-breed competition for each breed. That is, wagering on thoroughbred races will face competition from ITW sites taking harness racing and vice versa. This ITW cross-breed competition may result in reduced wagering at existing wagering locations for the other breed. Competition from wagering on simulcasts at off-track betting locations was found to result in reduced wagering at racetracks in the same market area. At a single racetrack location where the customer may choose among a variety of simulcast pari-mutuel wagering products from other locations, wagering on an individual simulcast racetrack's product was found to be reduced by competition from other simulcast racetrack products being offered on the same day.

References

- Ali, M. M., and R. Thalheimer. 2002. Product Choice for a Firm Selling Related Products: A Parimutuel Application, *Applied Economics* 34, 1251–1271.
- Ali, M. M., and R. Thalheimer. 1997. Transportation Costs and Product Demand: Wagering on Parimutuel Horse Racing, *Applied Economics* 29, 529–542.

- Christiansen, E. M., and S. Sinclair. 2001. The Gross Annual Wager of the United States, 2000 (revised). Report available from Christiansen Capital Advisors at www.cca-i.com.
- Church, A. M., and A. K. Bohara. 1992. Incomplete Regulation and the Supply of Horse Racing, *Southern Economic Journal* 58, 732–742.
- Coate, D., and G. Ross. 1974. The Effects of Off-Track Betting in New York City on Revenues to the City and State Governments, *National Tax Journal* 27, 63–69.
- Morgan, W. D., and J. D. Vasche. 1979. Horse Racing Demand, Parimutuel Taxation and State Revenue Potential, *National Tax Journal* 32, 185–194.
- Morgan, W. D., and J. D. Vasche. 1982. A Note on the Elasticity of Demand for Wagering, *Applied Economics* 14, 469–474.
- Simmons, S., and R. Sharp. 1987. State Lotteries' Effects on Thoroughbred Horse Racing, *Journal of Policy Analysis and Management* 6, 446–448.
- Thalheimer, R., 1998. Parimutuel Wagering and Video Gaming: A Racetrack Portfolio, *Applied Economics* 30, 531–543.
- Thalheimer, R., 2008. Government Restrictions and the Demand for Casino and Parimutuel Wagering, *Applied Economics* 40, 1–19.
- Thalheimer, R., and M. M. Ali. 1992. Demand for Parimutuel Horse Race Wagering with Specific Reference to Telephone Betting, *Applied Economics* 24, 137–142.
- Thalheimer, R., and M. M. Ali. 1994. Exotic Betting, Pricing Policy and the Demand for Parimutuel Horse Race Wagering. Working Paper, Department of Equine Administration, College of Business and Public Administration, University of Louisville, Louisville, KY.
- Thalheimer, R., and M. M. Ali. 1995a. Exotic Betting Opportunities, Pricing Policies and the Demand for Parimutuel Horse Race Wagering, *Applied Economics* 27, 689–703.
- Thalheimer, R., and M. M. Ali. 1995b. Intertrack Wagering and the Demand for Parimutuel Horse Racing, *Journal of Economics and Business* 47, 369–383.
- Thalheimer, R., and M. M. Ali. 2003. The Demand for Casino Gaming, *Applied Economics* 35, 907–918.
- Vasche, J. D. 1990. The Net Revenue Effect of California's Lottery, *Journal of Policy Analysis and Management* 9, 561–564.

DOWNLOAD FULL VERSION HERE: <https://bettingbooks.org/12>

This page intentionally left blank

Download More Sports Betting Books at: www.bettingbooks.org

CHAPTER 2

Modeling Money Bet on Horse Races
in Hong Kong

John Bacon-Shone

Social Sciences Research Centre, The University of Hong Kong, Hong Kong.

Alan Woods

Deceased.

1. Introduction	18
2. Variables examined	18
2.1. <i>Outcome Variables</i>	18
2.2. <i>Independent Variables</i>	19
3. Results and Discussion	19
4. Conclusion	21
<i>References</i>	22
Appendix: 31 Independent Variables Examined (Excluding Quadratic Terms)	23

Abstract

This chapter examines factors that affect the total money bet on races in Hong Kong and also how that money is distributed between different pools. The conclusion is that quite different processes are at work, but good predictive models are feasible, enabling the detection of the impact of relatively small changes in the betting tax, particularly on how the money is distributed across pools and also enabling us to detect what attracts gamblers to bet more or shift money across pools.

Keywords: pool size, handle size, betting tax changes, horse race betting

1. INTRODUCTION

Horse races in Hong Kong provide a rich source of data given the large amount of money bet and the evidence that the favorite-longshot bias found elsewhere is not found there (Busche and Hall, 1988; Lo et al., 1995). This study was triggered by the claim of the Hong Kong Jockey Club (which controls racing and pari-mutuel betting in Hong Kong) that the money bet on horse races in Hong Kong had rebounded as a result of the introduction at the beginning of the 2006 season of a rebate in some pools¹ to better compete with illegal bookmakers. We examine a total of 5,271 races from the beginning of the 2000 season until the middle of the 2007 season, during which 482 billion HK dollars were bet (about U.S.\$61B). Thalheimer and Ali (1992, 1995a, 1995b, 1997, 1998) have examined a number of predictors of money bet on races across tracks and years. Their analysis provides insight into competitive elements at the track level, but it is a crude analysis in that it does not enable us to understand factors that affect money bet within years, let alone within meetings. Ray (2002) examined the predictors of money bet on individual races across a one-month period at a single track. In that analysis, the short time period and single track limits the generalizability of the results, while this study examines more than seven seasons at two tracks with large betting volume.

2. VARIABLES EXAMINED

2.1. Outcome Variables

LTPCPI: Log of total betting pool in HK\$ on a race, adjusted for inflation using the Consumer Price Index (A). The log transform ensures a symmetric distribution. The Consumer Price Index (A) is the CPI with broadest base in Hong Kong (covering 50% of households). This variable provides the broadest measure of demand for gambling on horse races in Hong Kong, excluding only illegal bookmakers.

¹The rebate takes the form of a 10% rebate on losing bets of at least HK\$10,000 in the win, place, quinella, and quinella-place pools.

LRREBNREB: Log ratio of the total bets on pools eligible for rebate (from 2006 onward) to the total for all other pools for each race. The log ratio follows the standard approach for compositional data analysis of Aitchison (1986), to avoid the constraints imposed by proportions, which distort standard statistical analysis. This variable should provide the most sensitive assessment of the impact of the rebate, but also provides a good summary of the demand for standard versus exotic bets.²

2.2. Independent Variables

The full list of 31 independent variables (covering economic, race, weather, pool availability, and money distribution) appears in Appendix A.

3. RESULTS AND DISCUSSION

All the independent variables and quadratic extensions were examined as possible predictors of the two dependent variables, with a requirement that the variables should be marginally significant at 1%. The use of 1% significance rather than 5% takes into account the overall Type II error, that is, the risk of bias due to selecting from a large set of variables and also the large sample size.

This analysis yields a model for the total money bet as shown in Table 1 in decreasing order of t -statistics. This model has an adjusted R^2 of 83.8% and root mean standard error (RMSE) of 0.045, which is quite remarkable compared to the 64% of Ray's model. Decreasing seasonal trend, decreasing class trend, increasing race number trend, increasing number of horses trend, preference for middle distance races, decreasing day of week trend, increasing start time trend, increasing trend by month within season, decreasing surface trend, decreasing rain trend, increasing unemployment trend, and increasing atmospheric pressure trend are all stronger marginal predictors of amount bet than the introduction in bet rebates, although the rebate impact is still strongly significant. All of these variables have an impact in the expected direction, with the exception of unemployment, although the presence of a seasonal trend in the opposite direction suggests caution in interpretation of this effect.

This model is much more detailed than those of Thalheimer and Ali, but it is interesting that the positive impact of exotic betting opportunities they found does not show up in Hong Kong. It is difficult to make a direct comparison with Ray's model given the different time scale and more detailed list of variables, however, the increasing trend with race number is found in Hong Kong, but not the tailing off toward the end of the meeting. In Hong Kong, there is a preference for Sunday, not Saturday, compared to the general preference for the weekend in Ray's model. Like Thalheimer and Ali, she finds an increase when exotic bets are available. In this case, the difference in Hong Kong may be that there are always some exotic bets available, it is just a matter of how many pools are open for a specific race.

²Win, place, quinella, and quinella-place pools versus tierce, trio, first four, double, and treble.

TABLE 1 Summary of Best Model Predicting Total Pool per Race

LTPCPI	Coefficient	Standard error	<i>t</i>
SEASON	-0.01611	0.00080	-20.2
CLASS2	-0.00351	0.00017	-20.1
RACENO	0.01641	0.00085	19.3
NUMHSE	0.01135	0.00074	15.4
CLASS	-0.01015	0.00069	-14.8
SEA2	0.00633	0.00045	14.0
DIST2	0.00000	0.00000	-12.9
NUMH2	-0.00356	0.00028	-12.9
DAYWEEK	-0.00368	0.00030	-12.3
STARTTIME	0.42798	0.03587	11.9
SEMONTH	0.00256	0.00024	10.6
SURFACE	-0.02154	0.00206	-10.5
RACE2	0.00097	0.00009	10.3
DISTANCE	0.00002	0.00000	8.7
SEM2	-0.00121	0.00014	-8.5
SQRAIN	-0.00278	0.00034	-8.2
LUP	0.11817	0.01480	8.0
MEANPRESS	0.00186	0.00023	7.9
TIMEID	-0.06269	0.00894	-7.0
LWR2	0.03092	0.00493	6.3
NUMREB	0.09561	0.02514	3.8
BETCHANGE	0.01232	0.00371	3.3
DAYW2	-0.00090	0.00029	-3.1
MEANDP	0.00058	0.00021	2.7

Overall, betting on horse races in Hong Kong is generally becoming less popular, but races with more horses, racing at middle distance, and in higher classes are preferred, with increasing money bet per race across the meeting. However, looking at the distribution of betting money across pools eligible for rebate (versus not eligible) shows a different story. Table 2 shows the summary of the best prediction model, which has an adjusted R^2 of 64.6% and root mean standard error of 0.035. This shows that the single most important marginal predictor is the number of different pools available in the non-rebate category, with increased bet options moving money away from the rebate pools. After the race number, the next most important variables are the increase in money toward the rebate pools across seasons and the switch toward the rebate pools, when the rebate was introduced. Interestingly, both of the variables relating to the distribution of money across horses show very

TABLE 2 Summary of Best Model Predicting Ratio of Total Pools per Race

LRREBNREB	Coefficient	Standard error	<i>t</i>
NUMNREB	-0.08369	0.00187	-44.7
RACE2	-0.00226	0.00007	-30.9
SEASON	0.01409	0.00046	30.7
BETCHANGE	0.06858	0.00228	30.1
LWR1	-0.08684	0.00298	-29.1
CLASS2	-0.00260	0.00013	-19.7
LWR2	-0.04477	0.00402	-11.1
LSUP	-0.06495	0.00666	-9.8
NUMNR2	0.01153	0.00128	9.0
NUMH2	-0.00162	0.00020	-8.2
SURFACE	-0.01130	0.00156	-7.2
DIST2	0.00000	0.00000	-6.3
DISTANCE	0.00001	0.00000	5.4
DAYWEEK	0.00113	0.00022	5.3
CLASS	0.00221	0.00052	4.2
NUMHSE	0.00194	0.00047	4.2
SEMONTH	-0.00073	0.00018	-4.1
RACENO	-0.00102	0.00026	-3.9
NUMREB	0.06166	0.01913	3.2

strong effects, with strong first and second favorites making the exotic pools more attractive.

Overall, the introduction of the rebate has shifted money away from the exotic pools and the other major determinant is that strong first and second favorites move money toward the exotic pools, both of which seem very rational responses from gamblers.

4. CONCLUSION

While the impact of the introduction of a rebate can be clearly seen, most of the impact is on a shift in money away from exotic pools, with a much weaker impact on increasing the total money bet. Gamblers clearly prefer larger fields, middle distances, and higher class races. They respond to strong first and second favorites by shifting money to exotic bets. Modeling the distribution of money bet across races shows potential in understanding the behavior of gamblers and provides guidance on what attracts them to bet more or shift money across pools.

References

- Aitchison, J. 1986. *The Statistical Analysis of Compositional Data*. Chapman and Hall, London.
- Busche, K., and C. D. Hall. 1988. An Exception to the Risk Preference Anomaly, *Journal of Business* 61, 337–346.
- Hong Kong Census and Statistics Department, Hong Kong SAR Government. 2008. Consumer Price Index (A). Available at http://www.censtatd.gov.hk/hong_kong_statistics/statistical_tables/?charsetID=1&subjectID=10&tableID=052. Accessed February 15, 2008.
- Hong Kong Census and Statistics Department, Hong Kong SAR Government. 2008. Unemployment Rate. Available at http://www.censtatd.gov.hk/hong_kong_statistics/statistical_tables/?charsetID=1&subjectID=2&tableID=006. Accessed February 15, 2008.
- Hong Kong Observatory. 2008. Summary of Meteorological and Tidal Observations in Hong Kong. Available at <http://www.weather.gov.hk/publica/pubsmo.htm>. Accessed February 15, 2008.
- Lo, V. S. Y., J. Bacon-Shone, and K. Busche. 1995. The Application of Ranking Probability Models to Racetrack Betting, *Management Science* 41, 1048–1059.
- Ray, M. 2002. Determinants of Simulcast Wagering: The Demand for Harness and Thoroughbred Horse Races. Available at <http://gaming.univ.edu/reading/MargaretRay.pdf>. Accessed February 1, 2008.
- Thalheimer, R., and M. Ali. 1992. Demand for Parimutuel Horse Race Wagering with Special Reference to Telephone Betting, *Applied Economics* 24, 137–142.
- Thalheimer, R., and M. Ali. 1995a. Intertrack Wagering and the Demand for Parimutuel Horse Racing, *Journal of Economics and Business* 47, 369–383.
- Thalheimer, R., and M. Ali. 1995b. Exotic Betting Opportunities, Pricing Policies and the Demand for Parimutuel Horse Wagering, *Applied Economics* 27, 689–703.
- Thalheimer, R., and M. Ali. 1997. Transportation Costs and Product Demand: Wagering on Parimutuel Horse Racing, *Applied Economics* 29, 529–542.
- Thalheimer, R., and M. Ali. 1998. Parimutuel Wagering and Video Gaming: A Racetrack Portfolio, *Applied Economics* 30(4), 531–544.

APPENDIX: 31 Independent Variables Examined (Excluding Quadratic Terms)

Economy:

LUP: Log unemployment rate

LSUP: Log unemployment rate (seasonally adjusted)

Weather:

MINTEMP: Min daily temperature

MAXTEMP: Max daily temperature

MEANTEMP: Mean daily temperature

MEANPRESS: Mean atmospheric pressure

TOTALRAIN: Total daily rainfall

ANYRAIN: Indicator of more than trace rainfall

MMRAIN: Indicator of more than 1 mm rainfall

MM10RAIN: Indicator of more than 10 mm rainfall

SQRAIN: Square root of rainfall

MEANCLOUD: Average daily cloud cover

MEANRH: Mean relative humidity

MEANDP: Mean DP

Pool Availability:

NUMREB: Number of active pools from the set of Win, Place, Quinella, and Quinella-Place (3–4)

NUMNREB: Number of active pools from the set of Tierce, Trio, Double, Treble, and First 4 (2–5)

BETCHANGE: Whether rebate introduced yet (0–1)

Race Information:

SEASON: Season (2000–2007)

SEMONTH: Month of season (0–11)

TRACK: Track (Shatin or Happy Valley)

DIST: Distance (1000–2400)

SURFACE: Surface (1–2)

GOING: Going (1–9)

CLASS: Class (1–9)

DAYWK: Day of week (1–7)

TIMEID: Day/night race

STARTTIME: Time of day (as proportion of day)

RACENO: Race number (1–10)

NUMHSE: Number of horses starting race (5–14)

Money Distribution Among Horses:

LWR1: Log ratio of money bet on first favorite to money bet on all other horses bet less in win pool

LWR2: Log ratio of money bet on second favorite to money bet on all other horses bet less in win pool

DOWNLOAD FULL VERSION HERE: <https://bettingbooks.org/12>

PART II: Utility, Probability, and Pace Estimation

Download More Sports Betting Books at: www.bettingbooks.org

DOWNLOAD FULL VERSION HERE: <https://bettingbooks.org/12>

This page intentionally left blank

Download More Sports Betting Books at: www.bettingbooks.org

CHAPTER 3

Empirical Evidence on the Preferences of Racetrack Bettors

Bruno Jullien

Toulouse School of Economics (GREMAQ and IDEI), Toulouse, France.

Bernard Salanié

Columbia University, New York, NY, USA.

1. Introduction	28
2. Some Stylized Facts	30
3. Expected Utility	33
4. Distortions of Probabilities	36
5. Reference Points and Asymmetric Probability Weights	39
6. Heterogeneous Preferences	42
7. Exotic Bets	45
8. Concluding Remarks	46
<i>References</i>	47

1. INTRODUCTION

This chapter is devoted to the empirical estimation of the preferences for risk of gamblers on real market data. While there have been several experimental studies trying to elicit preferences of gamblers in the laboratory,¹ the observation of real markets remains a necessary step in assessing the properties of gamblers' preferences.² This is particularly true for gambling; it is indeed often asserted that gambling on racetracks (or in a casino) involves some type of utility that is hardly replicable in experiments.³

We concentrate in this survey on the empirical work that has been conducted for horse races,⁴ in the pari-mutuel system, or in a bookmaker system.⁵ Horse races (or other types of betting markets, e.g., sports events) are very good candidates to test theories of preferences under risk: they allow researchers to collect large datasets, and the average amount of money at stake is significant.⁶ Financial markets would be a natural area where the empirical relevance of the implications of the various non-expected utility models could be tested.⁷ However, portfolio choices have a very marked dynamic character, and non-expected utility theories are difficult to handle in dynamic settings.

Racetrack studies may provide key insights for the analysis of risk-taking behavior in financial investment, as well as in other contexts where risk is a main issue (e.g., environmental risk). Betting markets have the advantage of being short-run, lasting for one period only. This allows an exact evaluation of the ex-post return on each bet. As such, they provide an archetype of a simple contingent security market as defined by Arrow (1964). For horse races, a winning bet of \$1 on a particular horse is simply a contingent security that yields a revenue $(R + 1)$ dollars in the event the horse wins the race and 0 otherwise. Note that such a security cannot be retraded. The odds R of the horse in this context is defined as the net return in the winning case.⁸ In a bookmaker system, odds are commitments of payment by bookmakers who quote the prices. In a pari-mutuel system, they are endogenous, resulting from the distribution of the wagers over the horses: the odds of horse i is the ratio between the total money B wagered on the race net of the track revenue⁹ and the total money wagered on the horse B_i ,

¹See the survey by Camerer (1995).

²An alternative is to use household surveys (see, e.g., Donkers et al., 2001).

³See for instance Thaler and Ziemba (1988).

⁴There has been some work on Lotto games, sports events, and TV shows (see the conclusion).

⁵Hausch et al. (1994) present contributions covering most aspects of the economics of racetrack betting. The book edited by Vaughan Williams (2003) discusses the economics of gambling more generally.

⁶Weitzman (1965) estimates an average \$5 win bet on individual horses in the 1960s, while Metzger (1985) evaluates that \$150 was the average amount bet by an individual during the day in 1980.

⁷For a recent overview the theory and empirical evidence of portfolio choices, see the contributions in Guiso et al. (2002).

⁸Note that 3 to 1 odds correspond to $R = 3$ and thus a revenue of \$4 for a bet of \$1 in the event that the horse wins the race. Following the empirical literature, we focus on win bets, and ignore combinatorial bets.

⁹It includes the take and the breakage. The take corresponds to the percentage of bets collected by the race-track organizers, and the taxes. The breakage corresponds to the part of the return lost due to the fact that it is rounded to the nearest monetary unit.

minus 1:

$$R_i = \frac{(1 - \tau)B}{B_i} - 1, \quad (1)$$

where τ is the take.

At any point in time, odds reflect the market information on winning probabilities and evolve over time, until the race starts. In particular, data may include odds quoted before the racetrack opens, and odds quoted on the track. The most common practice is to use starting prices, that is, odds measured at the last minute of betting.¹⁰ The empirical studies discussed below then start with odds data and winners data, and use them to derive econometric estimates of bettors' preferences.

Note that there is clearly a selection bias in focusing on bettors and starting prices. All individuals do not bet, and the population of individuals betting at the track (and thus going to the race field) is hardly representative of the whole population. It may not even be representative of the whole population of bettors, as bettors off-track are not the same as bettors on-track. So the only information that can be derived is information on the preferences exhibited by individuals betting on the fields. Still, this is indicative of the type of risk that individuals may engage in, and, given the simple nature of the market, provides a very good test for various theories of preference under risk. Moreover, as the selection bias is in the direction of selecting individuals within the most risk loving part of the population, this provides an overestimate of (and thus a bound on) the level of risk that an average individual may be willing to accept, which is clearly very useful.

Using econometric methods on racetrack data has the advantage of exploiting the large size of the samples available. Datasets usually include thousands of races, and thus allow precise estimates. Moreover, researchers can rely on fairly standard econometric models and procedures, ranging from the simple regression methods used in early work, to more sophisticated estimations of structural models. The main drawback is that individual data on bets and on bettor characteristics are typically not available. This implies several restrictions. First, the size of the wager can usually not be identified. Second, going to the racetrack and betting involves some type of entertainment value, and it is not possible to disentangle what is due to the specific utility derived from the attendance at the race, and more fundamental properties of preferences. It is also clear that racetrack bettors have heterogeneous preferences and information. To the best of our knowledge, no one has found a general approach to modeling heterogeneity in beliefs. The lack of individual data has led most researchers to focus on some form of average behavior, or more to the point, on the behavior of a representative bettor capturing the average risk attitude embedded in the dataset. In the bookmaker system, it seems to be the best one can do. Recent advances have shown that pari-mutuel data allows researchers to go beyond the representative agent model as it implicitly contains information about the total amount bet on each horse.

¹⁰The studies discussed below could be done with any odds, under a rational expectation assumption. The informational content of prices is the highest at starting prices, so that they should provide a more accurate predictor of winning probabilities than earlier odds. See for instance Asch et al. (1982).

In what follows we first discuss (Section 2) the main stylized facts of horse races that have shaped the research agenda. We then present in Section 3 the work based on the expected utility model, which put in place the foundation for subsequent work. Sections 4 and 5 review the work departing from the expected utility paradigm. Section 4 focuses on the perception on winning probabilities by bettors, while Section 5 discusses the role of the reference point and the asymmetric treatment of wins and losses.

All of this work assumes a representative bettor; but two very recent contributions (Gandhi, 2006; Chiappori et al., 2007) show that the distribution of risk preferences among bettors can be elicited from pari-mutuel data, at least if the heterogeneity of these preferences is low-dimensional. We briefly discuss their method and their preliminary results in Section 6. Then Section 7 turns to recent work by Snowberg and Wolfers (2007) that pools information across simple bets and exotic bets. We conclude by offering some ideas for future research in Section 8.

2. SOME STYLIZED FACTS

Any empirical study of the preferences of racetrack bettors must account for the most salient stylized fact of racetrack betting data: the favorite-longshot bias. The favorite-longshot bias refers to the observation that bettors tend to underbet on favorites and to overbet on outsiders (called longshots). As it is presented in more detail in the chapters by Ottaviani and Sorensen and by Tompkins, Ziemba, and Hodges in this *Handbook*, we only recall here the points that matter for our discussion.¹¹ Thus we focus on the implications of the favorite-longshot bias on how we view bettors' preferences.

The favorite-longshot bias seems to have been documented first by Griffith (1949) and McGlothlin (1959). Griffith studied 1,386 races run under the pari-mutuel system in the United States in 1947. For each odds class R , he computed both the number of entries E_R (the total number of horses with odds in odd class R entered in all races) and the product of the number of winners in this class and the odds N_R . A plot of E_R and N_R against R showed that while the two curves are very similar, N_R lies above (below) E_R when R is small (large). Since small R corresponds to short odds (favorites) and large R to long odds (longshots), this is evidence that in Griffith's words, there is "a systematic undervaluation of the chances of short-odded horses and overvaluation of those of long-odded horses." A risk-neutral bettor with rational expectations should bet all his other money on favorites and none on longshots.

A number of papers have corroborated Griffith's evidence on the favorite-longshot bias.¹² To give just one example, let us look at the dataset used by Jullien and Salanié (2000). This dataset is composed of each flat horse race run in Britain between 1986 and 1995, or 34,443 in total. British racetrack betting runs on the bookmaker system, so odds R are contractual. This dataset makes it easy to compute the expected return of betting on a horse with given odds, as plotted in Figure 1. For any given R , we compute

¹¹See also Häusch et al. (1994) for a survey of the evidence.

¹²Exceptions have been found for Hong Kong races by Bushe and Hall (1988).

Bruno Jullien and Bernard Salanié

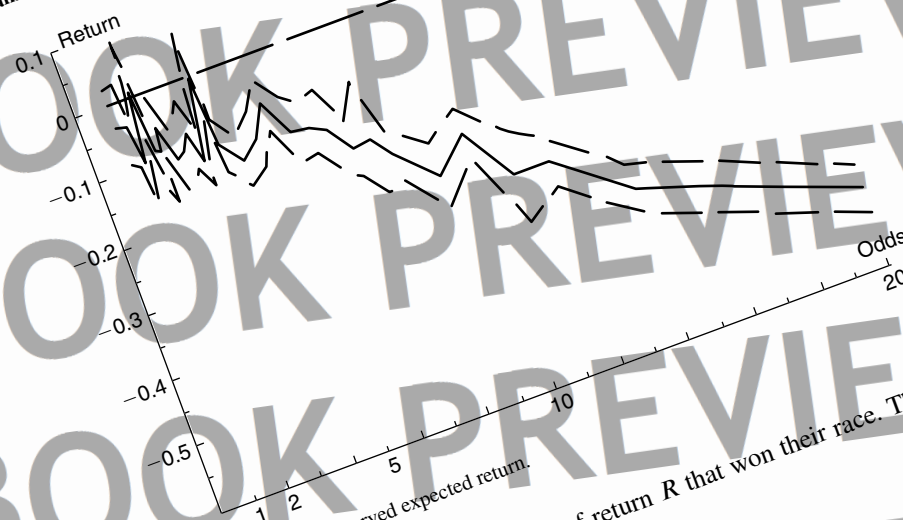


FIGURE 1 Observed expected return. The expected

$\hat{p}(R)$, the proportion of horses with rate of return R that won their race. The expected return then is

$$\widehat{ER}(R) = \hat{p}(R)R - (1 - \hat{p}(R))$$

for a bet of £1, since such a bet brings a net return of R with probability $\hat{p}(R)$ and a net return of -1 with probability $(1 - \hat{p}(R))$. Figure 1 plots $\widehat{ER}(R)$, along with a 95% confidence interval. The expected return is always negative (the occasional spikes on the left of the figure are for odds that correspond to relatively few horses): it does not pay for a risk-neutral bettor to gamble at the racetrack. More interestingly, the expected return decreases monotonically with the odds R , so that it is much less profitable for such a bettor to bet on longshots than to bet on favorites: even for very common odds of 10 to 1, the expected loss is 25 pence on the pound, as compared to less than 10 pence for horses with even odds (of 1 to 1).

The favorite-longshot bias has been much discussed and four main types of explanations have emerged in the literature:¹³

1. The original explanation of the favorite-longshot bias was given by Griffith (1949) and referred to misperceptions of probabilities by bettors. Griffith argued that as in some psychological experiments, subjects tend to undervalue large probabilities and to overevaluate small probabilities. Thus, they scale down

¹³ Ali (1977) also points out that the favorite-longshot bias can be explained by heterogeneous beliefs, reflecting different subjective probabilities of bettors and a lack of common knowledge. Modeling this would require introducing some heterogeneity in nonexpected utility models with probability distortions, and so far, the data does not allow us to do this.

the probability of a favorite winning a race and they scale up the probability that a longshot wins a race, which indeed generates the favorite-longshot bias. Henery (1985) suggests a somewhat similar explanation. He argues that bettors tend to discount losses: if the true probability that a horse loses the race is q , they take it to be $Q = fq$, where $0 < f < 1$ is some constant number. This theory can be tested by measuring $Q(R)$ to be the value that makes the expected return of betting on a horse with odds R exactly zero; from the formula above, this $Q(R)$ equals $R/(R+1)$. Now the value of $q(R)$ is given as $q(R) = 1 - \hat{p}(R)$. By regressing $Q(R)$ on $q(R)$ without a constant, Henery found an estimated f of about 0.975 and a rather good fit.¹⁴

2. Quandt (1986) showed how risk-loving attitudes generate the favorite-longshot bias at the equilibrium of betting markets. To see this, take two horses i and j in the same race, with odds R_i and R_j and true probabilities of winning p_i and p_j . The expected return of betting \$1 on horse $h = i, j$ is

$$\mu_h = p_h R_h - (1 - p_h),$$

and the variance of this bet is

$$v_h = p_h R_h^2 + (1 - p_h) - \mu_h^2,$$

which is easily seen to be

$$v_h = p_h(1 - p_h)(R_h + 1)^2.$$

Now if bettors are risk-loving, the mean-variance frontier must be decreasing in equilibrium: if $\mu_i < \mu_j$, then it must be that $v_i > v_j$. Then a fortiori equilibrium requires that

$$\frac{v_i}{(\mu_i + 1)^2} > \frac{v_j}{(\mu_j + 1)^2}.$$

But easy computations show that

$$\frac{v_h}{(\mu_h + 1)^2} = \frac{1 - p_h}{p_h},$$

so that if $\mu_i < \mu_j$, then $p_i < p_j$. The contrapositive implies that horses with a larger probability of winning (favorites) yield a higher expected return, which is exactly the favorite-longshot bias.

3. Following the evidence of informed trading (Asch et al., 1982; Craft, 1985), Shin showed in a series of papers (1991, 1992, 1993—see also Jullien and Salanié, 1994) that in a bookmakers' market, the presence of insider traders generates the favorite-longshot bias as bookmakers set odds so as to protect themselves against such well-informed bettors.

¹⁴Our own estimates on our dataset suggest that the constant term in $Q = a + bq$ is highly significant.

4. Finally, it may be that the utility of gambling is higher for bets on longshots, perhaps because they provide more excitement; this explanation is advanced by Thaler and Ziemba (1988). Then, if risk-neutral bettors equalize the sum of the expected return and the utility of gambling across horses in each race, clearly the expected return will be higher for favorites.

These four explanations are not mutually incompatible. In modern terms that were not available to Griffith (1949), explanation 1 hints at a nonexpected-utility model of bettors' preferences with non-linear weighing of probabilities.¹⁵ Explanation 2 therefore can be subsumed in explanation 1, with risk-loving appropriately redefined so that it makes sense for nonexpected-utility preferences.

Because the rest of this chapter will focus on explanations 1 and 2, we should explain here why we put aside explanations 3 and 4. The literature on insider trading is covered in Sauer (1998) along with the test of efficiency of wagering markets. For our purposes, one problem with Shin's models is that they are rather specific, so that estimating the incidence of insider trading requires strong assumptions on preferences and the information structure of insider traders. Still, it might make sense to pursue this direction. However, this is in fact not necessary so far as the gambler's preference is the object of interest. It is true that the existence of a fringe of insider traders changes the behavior of bookmakers; but under rational expectations, all the information available is incorporated into prices so that one may still estimate the preferences of a gambler with no private information. Finally, explanation 4 also is intuitively appealing: betting on a huge favorite, say with a 99% chance of making a net return of 1 cent on the dollar, is clearly less fun than betting on a longshot that brings \$100 with a 1% probability. One difficulty with this explanation is that in a sense, it explains too much: since there is little evidence on the determinants and the functional form of the utility of gambling, any feature of the equilibrium relationship between p and R can be explained by an ad hoc choice of functional form for the utility of gambling. However, we will see later that models with nonexpected-utility preferences, by reweighting probabilities, may yield similar predictions to models with a utility of gambling that depends on the probability of a win.

3. EXPECTED UTILITY

The seminal contribution in the domain is the work of Weitzman (1965) who builds on the above findings and attempts to estimate the utility function of an average expected utility maximizer. Weitzman had at his disposal a dataset of 12,000 races, collected on four New York racetracks for a period covering 1954 to 1963. Following Griffith (1949), Weitzman starts by aggregating horses over all races by odds category, obtaining 257 odds classes. From the winners dataset, he then constructs the ex post estimate $\hat{p}(R)$ of the winning probability of a horse conditional on its odds category R . This allows him to estimate a functional relation between the odds category and

¹⁵ See the conclusion for other types of cognitive biases.

the winning probability.¹⁶ Then he attempts to build the utility function of an average bettor, referred to as Mr. Avmart (average man at the racetrack), as follows. Mr. Avmart is an expected utility maximizer with a utility function $u(\cdot)$ and he bets a fixed amount on each race,¹⁷ normalized to 1 for the exposition (the actual unit Weitzman uses is \$5). Mr. Avmart is representative in the following sense: the data observed could be generated by Mr. Avmart (or a population of identical Mr. Avmarts) betting. As every odds category receives some bet, Mr. Avmart must be indifferent between all the odds categories, which implies that

$$\hat{p}(R)u(R) + (1 - \hat{p}(R))u(-1) = K \text{ for all } R,$$

where K is the constant expected utility. This yields the relation

$$u(R) = u(-1) + \frac{K - u(-1)}{\hat{p}(R)},$$

which allows him to estimate a utility function for all money levels R . Using this procedure, Weitzman found a convex utility function on the range of money value covered (\$5 to \$500), consistent with the assumption of a risk-loving attitude.

Ali (1977) conducted a similar study with a 20,247 race dataset, grouping the horses according to their ranking as opposed to their odds. For each ranking, an average odds and an empirical winning probability are computed. He then estimates the utility function of an agent indifferent to betting on any horse category. Ali confirms the Weitzman finding, with a risk-loving utility function. He estimates a constant relative risk-aversion utility (CRRA) with a coefficient of relative risk aversion of -0.1784 . Applying the methodology to different data, Kanto et al. (1992) and Golec and Tamarkin (1998) estimate somewhat similar CRRA utility functions.¹⁸

By construction, the preferences of the representative agent are based only on the information contained in the odds category (or the ranking in the case of Ali). The behavior of the agent is representative on average, in the sense that he is indifferent between betting on the horse in a given category on all races, and betting on the horse in another category on all races.¹⁹ Thus, the construction of Mr. Avmart's preferences involves two types of aggregation: of the information over odds and winning probabilities, and of the preferences. One of the drawbacks of the categorization of odds is that the number of points used to fit the utility function is usually small (except for Weitzman, 1965, who builds 257 categories). Another important aspect is that the only information used is the category of the horse, so some information on the races included in the dataset is not used by Mr. Avmart. This is the case, for instance, for the number of runners in each race. Given the nature of the pari-mutuel system, one may think

¹⁶He estimates a hyperbola $\hat{p}(R) = \frac{0.845}{R}$ and a "corrected hyperbola" $\hat{p}(R) = \frac{1.01 + 0.09 \log(1 + R)}{R}$.

¹⁷Recall that data on individual bets are not available, so the amount bet must be postulated.

¹⁸Golec and Tamarkin's estimates for relative risk aversion, based on odds category, are -0.14 for the whole dataset, and -0.2 for a data conditional on having a large favorite. Values differ but they all confirm a risk-loving attitude.

¹⁹Or in a race chosen at random in the sample of races.

that the number of runners may affect the relationship between the winning probability and the odds. More generally, this relationship may vary with the race characteristics.²⁰ A second remark is that it may also vary with the take, or more generally with the mark-up over winning probabilities that corresponds to the revenue of the betting institution. In the case of a pari-mutuel market, this is not so problematic, as the take is fixed at the race track level. However, when applied to a betting market organized by bookmakers, the procedure may create serious biases as the mark-up is chosen by the bookmakers and may vary from one race to the another.²¹

Jullien and Salanié (2000) propose a method to estimate the representative agent's preferences that accounts for heterogeneity among races. To understand the procedure, let us consider a given race r with E_r horses. Let p_{ir} denote the objective winning probability of horse i in race r and R_{ir} be the odds. Now assume that the representative agent is indifferent between betting \$1 on any horse in race r . Then there must be some constant K_r such that:

$$p_{ir}u(R_{ir}) + (1 - p_{ir})u(-1) = K_r.$$

Using the fact that probabilities add up to one, one can then recover for each race and each horse, a unique predicted probability of winning $\hat{p}(i, r, u)$ and a constant K_r consistent with this relation. The procedure then consists of using the winners' dataset to find the utility function $u(\cdot)$ that provides the best fit to the empirical dataset using a maximum likelihood method. Note that the method has the advantage of using all the data information, and getting rid of the categorization of odds. The nature of the representative agent is slightly different as he or she is indifferent between betting on any horse on any given race, as opposed to placing a systematic bet on a given odds category on all races. Thus the agent too uses all the information in the dataset, and may even use more information, in order to adjust to the specificities of races.

Applying this procedure to the estimate of a utility function, Jullien and Salanié confirmed the finding of a risk-loving utility function. It appears, however that the CRRA utility representation is outperformed by a utility function with a constant absolute risk aversion (CARA). Among the class of utility functions with a hyperbolic risk aversion, the best fit was obtained for a CARA utility function, with a fairly moderate degree of risk-loving behavior.

Expected utility estimates provide results that are consistent with explanation 2 of the favorite-longshot bias, that is, a risk-loving attitude. However, as documented by Golec and Tamarkin (1998), and Jullien and Salanié (2000), these estimates tend to perform poorly for large favorites. Indeed, the probabilities of winning implied by the estimated utility and the underlying structural model of the representative agent tend to be too small for large favorites. Arguably, this can be due to the parametric forms chosen for the utility function estimated, which restrict its curvature. Arguing that CRRA utility

²⁰McGlothlin (1956) and Metzger (1985) provide evidence that the position of the race in the day matters, as well as the amount of publicity for the race.

²¹If the average mark-up varies with the race, as is the case with bookmakers, the constant K above should depend on the race. The same issue arises when using data from racetracks with different values of the take.

functions perform poorly for large favorites, Golec and Tamarkin (1998) estimate a cubic utility function:

$$u(R) = -0.071 + 0.076R - 0.004R^2 + 0.0002R^3.$$

The utility function exhibits risk aversion for low odds (favorites) and a risk-loving attitude for larger levels of odds. As the coefficient for the variance is negative, they conclude that the risk-loving attitude is related to the skewness of the distribution (the third moment). While the risk-averse attitude for small probabilities is an interesting result, this is probably as far as one can go with the expected utility model on this type of dataset. In particular, given the specific economic context, the non-representativeness of the population studied and the lack of data on individual bets and characteristics, very detailed estimates of the curvature of the utility function at various levels of odds may not be of much relevance for other applications. We now follow a different route. As argued before, although the precise preference of racetrack bettors may not be of special interest to the economist in a different domain, they provide a simple and clear real-life experiment. The next step is thus to use the data to test various departures from the expected utility paradigm on a real-life situation. Among these, the most popular in modern theory are the so-called nonexpected utility models, which provide mathematical representations of preferences under risk that are non-linear in the space of probability distributions.

Before we proceed, let us point out that there is no inherent contradiction between the expected utility representation and a nonexpected utility model of the agent behavior. Indeed, as we already noted, the data contains no information on the individual characteristics, and in particular on wealth. This means that all the utility functions are estimated only for the revenue derived from the betting activity. One may then consider that the distribution of this revenue represents a relatively small fraction of the risk supported by individuals on their total wealth, at least for the average bettor. As shown by Machina (1982), even when an agent evaluates his or her total wealth in a nonexpected utility manner, one may still evaluate small risks with an expected utility. The utility function is then local as it represents the differential of the global functional, and it depends on the underlying total wealth. Thus one may see expected utility estimates as a first order linear approximation of preferences. The question is then whether alternative utility representations provide a better representation of preferences than this approximation.

4. DISTORTIONS OF PROBABILITIES

The empirical evidence collected in the previous section suggests that the best expected utility rationalization of the equilibrium relationship between probabilities and odds exhibits a significant but not very large degree of risk loving. Still, a very copious body of experimental literature, starting with the work of Allais (1953), has accumulated to

shed doubt on the value of the expected utility model as a representation of behavior under risk. The recent survey by Camerer (1995), for instance, strongly suggests that the independence axiom is violated in most experiments.

On the other hand, there is no consensus about which nonexpected utility model to choose, even among authors who hold that we should drop the expected utility representation. Moreover, most of the evidence is experimental; there seems to be little evidence based on real-life economic situations. As argued in the introduction, bets on horses are very simple Arrow-Debreu assets that cannot be retraded and thus offer an exciting way of testing these theories. The next three sections are dedicated to this task. The first two mostly describe our own work (Jullien and Salanié, 2000). Then we move to very recent developments that start to account for the unobserved heterogeneity in risk attitudes among bettors. We can only hope that in 10 years, there will be many more papers to present in this field.

Recall that the expected utility of betting on horse i with odds R_i and probability of winning p_i is

$$p_i u(R_i) + (1 - p_i) u(-1).$$

This is a special case of the standard formula

$$\int u(x) dF(x),$$

where the risky outcome X has a cumulative distribution function F . There are many ways of altering this formula in order to obtain a nonexpected utility representation of preferences. One of the most natural, suggested by Quiggin (1982), consists of re-weighting probabilities, so that the value of X now takes the form

$$-\int u(x) d(G \circ (1 - F))(x)$$

where G is a continuous and increasing function that maps $[0, 1]$ into itself. While this may seem opaque, the application of this formula to the bet on horse i yields

$$G(p_i) u(R_i) + [1 - G(p_i)] u(-1).$$

While Quiggin (1982) called this specification “anticipated utility,” it now goes under the name of “rank dependent expected utility” (RDEU for short). Because G is a priori non-linear, RDEU breaks the independence axiom of expected utility. It does so in ways that may allow it to account for violations such as the Allais paradox: when G is convex, RDEU preferences indeed solve what is called in the literature the “generalized Allais paradox.”

Remember that Griffith (1949) explained the favorite-longshot bias by appealing to an overestimation of small probabilities and an underestimation of large probabilities. This points to a G function that is concave and then convex. On the other hand,

the weighting function postulated by Henery (1985) does not fit within RDEU, strictly speaking. It can indeed be written as

$$G(p) = 1 - f(1 - p)$$

which gives $G(0) = 1 - f > 0$ and thus is inconsistent with the axioms of RDEU (and indeed of any reasonable theory of choice under risk). This could of course be fixed by smoothly connecting $G(0) = 0$ with the segment represented by Henery's specification. Note that neither of these specifications yields a convex weighting function $G(p)$, as required to solve the generalized Allais paradox.

Jullien and Salanié (2000) fitted various RDEU functionals to their dataset of British flat races. All of these functionals assumed that the utility of wealth function u was a CARA function; on the other hand, they allowed for much more flexibility on the shape of the weighting function $G(p)$, which allowed them to nest the shapes suggested by Henery and Griffith, among others. Figure 2 offers a summary of their results. The most striking feature of these curves is that they are very close to the diagonal for each specification. Thus the estimated RDEU functionals hardly depart from the expected utility model. This is confirmed by formal tests, since the null hypothesis of expected utility is only rejected for one specification of the weighting function, that proposed by Prelec (1998) on the basis of an axiomatic derivation. According to this study at least, rank-dependent expected utility does not appear to provide a better fit of bettors' preferences than expected utility. Note that if anything, the estimated weighting functions are slightly convex on the whole $[0, 1]$ interval and thus do not go in the direction suggested by Griffith or Henery.

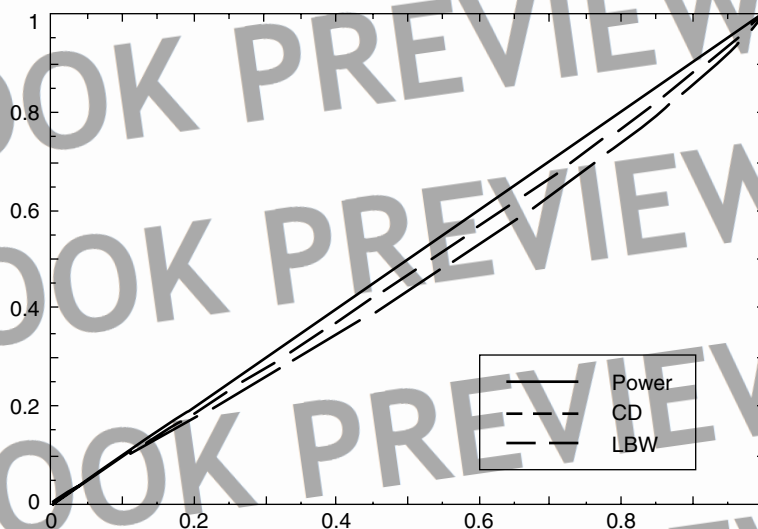


FIGURE 2 Estimated weighting functions for RDEU.

5. REFERENCE POINTS AND ASYMMETRIC PROBABILITY WEIGHTS

While the results in the previous section are not very encouraging for unexpected utility, there are many alternative specifications of these models. In particular, since Markowitz (1952), the notion of reference point has received some attention. This refers to the idea that individuals evaluate risk by comparison to some reference wealth, and treat losses and gains in an asymmetric way. This is particularly attractive in the case of betting, as there is a natural reference point (no bet) and a clear distinction between losses and gains.

In a recent work, Bradley (2003) proposes such a representation where the agent maximizes an expected utility with a reference point and a differential treatment of losses and gains.²² His representation assumes a different constant relative risk-aversion utility function for losses and for gains. This allows the representation to endogeneize the size of the bet, which is not done in other approaches. Although his investigation is still preliminary, it suggests that a representation with risk aversion on losses and a risk-loving attitude on gains may fit the data, in particular the favorite-longshot bias.

Among various theories involving a reference point, the cumulative prospect theory (CPT) has become very popular in recent years. Prospect theory was introduced by Kahneman and Tversky (1979) and developed into cumulative prospect theory in Tversky and Kahneman (1992). Most theories of choice under risk evaluate lotteries as probability distributions over final wealth. CPT diverges from this literature in that it evaluates *changes* in wealth with respect to a reference point that may for instance be current wealth. This matters in that in CPT, losses and gains are evaluated in different ways. Kahneman and Tversky first appeal to the principle of diminishing sensitivity, which states that the psychological impact of a marginal change decreases when moving away from the reference point. Applied to the utility of (changes in) wealth function, it suggests that this function is concave for gains but *convex* for losses. When applied to the probability weighting function, and given that the endpoints of the $[0, 1]$ interval are natural reference points, it suggests that this function should have the inverted-S shape implicit in Griffith (1949).

Cumulative prospect theory also adds two elements of asymmetry in the treatment of gains and losses. First, it allows for different probability weighting functions for gains and losses. Second, it assumes loss aversion, that is, that the utility of changes in wealth is steeper for losses than for gains, so that the function $u(x)$ has a kink at zero.

For a general prospect X with cumulative distribution function F , the value according to CPT is

$$\int_{x < 0} u(x) d(H \circ F)(x) - \int_{x > 0} u(x) d[G \circ (1 - F)](x)$$

²²As pointed out in Section 2, if we see the utility function as a local utility function, the notion of reference point becomes natural.

where G and H are two continuous and increasing functions that map $[0, 1]$ into itself where \circ denotes composition of functions. Given a bet on horse i with odds R_i and a probability of a win p_i , the CPT value simplifies to

$$G(p_i)u(R_i) + H(1 - p_i)u(-1).$$

Note the differences with RDEU. The most obvious one is that in general $H(1 - p) \neq 1 - G(p)$. The other one is hidden in the formula, since the function u should be convex for losses ($x < 0$), have a concave kink at zero [with $u(0) = 0$], and be concave for gains ($x > 0$). Clearly, only some of these properties can be tested from the data, since the only values of u on which we can recover information are those in -1 and on $[\underline{R}, +\infty)$, where \underline{R} is the smallest odds observed in the data.

In their paper, Jullien and Salanié (2000) chose to circumvent these difficulties by assuming that u was a CARA function. This is clearly not satisfactory, as it assumes away by construction any form of loss aversion and it violates the principle of diminishing sensitivity by forcing the concavity of u to have the same sign for losses and for gains. Jullien and Salanié normalize u by setting $u(0) = 0$ and $u'(0^+) = 1$; then the parameter of the CARA function is estimated from the relationship between probabilities and odds, and it implies a value for $u(-1)$, say A . Then they run a test of (and do not reject) the null hypothesis that $u(-1) = A$. This may be construed as a test of loss aversion by a sympathetic reader, but we admit that it is not very convincing. The best justification for their assuming a CARA utility function probably is that they want to focus on the probability weighting functions G and H and there is just not enough information in the data to probe more deeply into the function u .

Given this restriction, Jullien and Salanié tried three specifications for functions G and H . Figure 3 plots their estimation results for the function G . As in the RDEU case, the function appears to be slightly convex but very close to the diagonal: there is little evidence of a distortion of the probabilities of gains. The estimated H function, however, has a markedly concave shape for all specifications as shown in Figure 4. These results are somewhat at variance with the theory, which led us to expect inverted-S shapes for the probability weighting functions.

There are several ways to interpret these results, and Jullien and Salanié illustrate some of them. First, it can be shown that G convex and H concave explain the generalized Allais paradox. Second, local utility functions à la Machina (1982) can be derived from these estimates; they have a shape similar to that hypothesized by Friedman and Savage (1948). Let us focus here on how these preferences explain the favorite-longshot bias. To see this, remember that the function u exhibits a moderate degree of risk-loving behavior, and the function G is very close to the diagonal. Thus, to simplify things, assume $u(x) = x$ and $G(p) = p$. Then horse i is valued at

$$p_i R_i - H(1 - p_i)$$

which can be rewritten as

$$p_i(R_i + 1) - 1 - [H(1 - p_i) - (1 - p_i)].$$

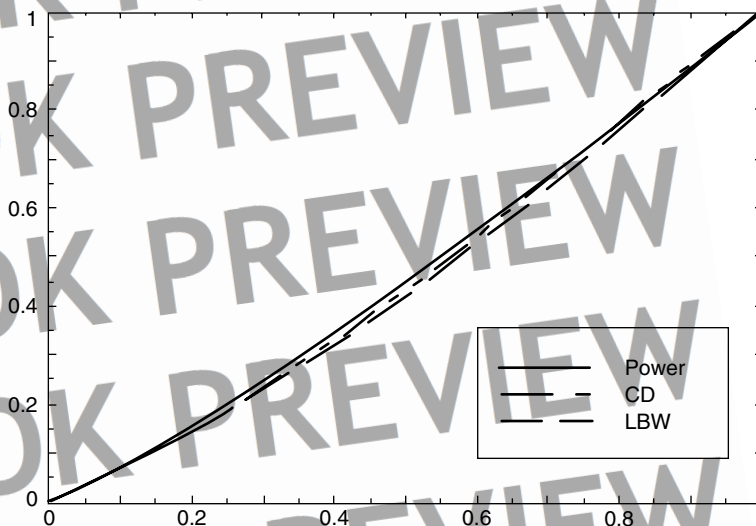


FIGURE 3 CPT probability weighting for gains.

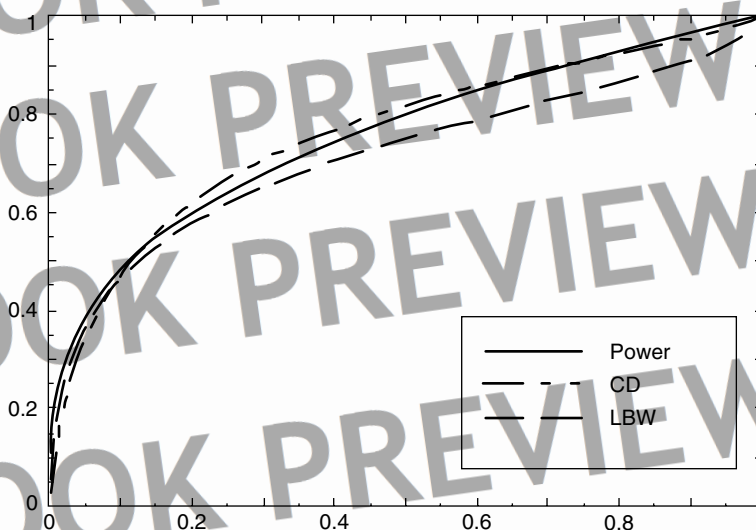


FIGURE 4 CPT probability weighting for losses.

Now given the estimates of Jullien and Salanié, the function $q \rightarrow H(q) - q$ is zero in 0 and 1 and has a unique maximum close to $q^* = 0.2$. Since most horses have a probability of winning much lower than $1 - q^* = 0.8$, it follows that $H(1 - p) - (1 - p)$ is an increasing function of p and therefore in equilibrium, the expected return $p_i(R_i + 1) - 1$ is an increasing function of the probability of a win. Thus bigger favorites are more

profitable bets for risk-neutral bettors, which is the definition of the favorite-longshot bias. The data suggest that the bias may be due not only to risk-loving behavior, as suggested by Quandt (1986), but also to the shape of the probability weighting functions. This is an intriguing alternative, since it can be shown that the concavity of H pulls toward risk-averse behavior for losses. Thus the favorite-longshot bias is compatible with risk-averse behavior, contrary to the standard interpretation.

Finally, let us return to explanation 4 of the favorite-longshot bias, based on the utility of gambling. First, note that the method used by Jullien and Salanié is robust to a utility of gambling that may differ across races but is the same for all horses in a given race. Now assume that for horse i , there is a horse-specific utility of gambling $f(p_i, R_i)$, so that the value of this bet for a risk-neutral bettor is

$$p_i(R_i + 1) - 1 + f(p_i, R_i).$$

By identifying this formula and the one above, we can see that our CPT estimates can be interpreted as representing the preferences of a risk-neutral bettor with a horse-specific utility of gambling given by

$$f(p_i, R_i) = 1 - p_i - H(1 - p_i)$$

which only depends on the probability of a win. Moreover, we know that it is a decreasing function of p_i for most horses. Thus this reinterpretation of Jullien and Salanié's CPT estimates brings us back to explanation 4 of Thaler and Ziemba (1988). There is in fact nothing in the data that allows the econometrician to distinguish between these two interpretations.

6. HETEROGENEOUS PREFERENCES

Recent work by Gandhi (2006) and Chiappori, Gandhi, Salanié, and Salanié—hereafter CGSS (Chiappori et al., 2007)—adapts the paradigm of empirical industrial organization to explore the heterogeneity of preferences among bettors on horse races. First, assume that all bettors have the same beliefs on the probability that any horse wins any given race. Then horses in a race are vertically differentiated goods: at any given price (or odds), a horse is preferred to another if and only if it is considered more likely to win.

Of course, prices are not equal; and in a bookmaker system they depend in a fairly opaque way on the bookmakers' own beliefs, the market structure, and other factors that are hard to control for. But in a pari-mutuel system, by construction, a simple formula relates prices and market shares. Denote by $S_i(R)$ the amount of money that is bet on horse i in a race with n horses and odds $R = (R_1, \dots, R_n)$, and t the track take. Then the odds on horse i are given by

$$(1 - t) \sum_{j=1}^n S_j(R) = (R_i + 1)S_i(R).$$

It is easy to get data on odds in a large number of races; then these yield directly the relative market shares (the percentage of money that was bet on each horse in a given race), since from the above

$$\frac{m_i(R)}{m_j(R)} = \frac{R_j + 1}{R_i + 1}.$$

Thus, in addition to the information on probabilities of winning that Jullien and Salanié (2000) used, pari-mutuel data contains information on market shares. For simplicity, assume here that the population of bettors is the same in all races, and let $W(p_i, R_i)$ denote the value that a bettor indexed with W attributes to a \$1 bet on a horse i which is expected to win a net return of R_i with probability p_i . Then if all bettors place \$1 bets, the market share is simply

$$m_i(R) = \Pr(W(p_i, R_i) \geq W(p_j, R_j)) \text{ for all } j = 1, \dots, n,$$

where the probability is taken over the distribution of W in the population.

Gandhi (2006) shows that even if bettors have different beliefs about the chances that any horse will win, under mild conditions there exists a rational expectations equilibrium where the odds reveal all information available to bettors. Thus, in equilibrium, all bettors effectively have the same beliefs, so that focusing on identical beliefs is at least internally consistent.

Going back to econometrics, we can recover the p_j 's by observing the odds of the winning horse in the many races in the data; and the equation above shows that observing the odds (and thus the market shares in a pari-mutuel system) conveys information about the distribution of the preferences of bettors. Gandhi estimates such a model by maximum likelihood, and tests various formulations of the EU model against RDEU. Unlike Jullien and Salanié, Gandhi uses a non-nested representation of EU and RDEU. He finds that his EU representation outperforms the RDEU model, the latter collapsing in a representative agent model. The EU estimates confirm the presence of heterogeneity; the favorite-longshot bias can then be interpreted as the outcome of the interaction between risk-averse and risk-loving bettors.

More generally, CGSS consider the case when all preferences W belong to an unknown family $V(\cdot, \cdot, \theta)$ indexed by a one-dimensional parameter θ , and satisfy a single-crossing property. CGSS show that the distribution of the $V(\cdot, \cdot, \theta)$ in the population of bettors then is non-parametrically identified, and they give a simple constructive procedure to estimate this distribution from the data. To review, assume that expected utility describes every bettor. Then bettor θ has a utility

$$p_i u(R_i, \theta) + (1 - p_i) u(-1, \theta)$$

from betting \$1 on horse i in a given race. The single-crossing condition in CGSS implies that in any race ($R = R_1, \dots, R_n$), the interval $[0, 1]$ partitions neatly: bettors with a θ in $[\theta_{i-1}(R), \theta_i(R)]$ will bet on horse i .

The bounds of these intervals are easy to recover from the market shares. For instance, normalizing the distribution of θ to be uniform in $[0, 1]$ yields:

$$m_i(R) = \theta_i(R) - \theta_{i-1}(R).$$

Now by definition, a bettor in $\theta_i(R)$ is indifferent between betting on horses i and $(i + 1)$. Therefore

$$p_i \{u[R_i, \theta_i(R)] - u[-1, \theta_i(R)]\} = p_{i+1} \{u[R_{i+1}, \theta_i(R)] - u[-1, \theta_i(R)]\}.$$

Since we know $\theta_i(R)$ and we can estimate probabilities p_i from the data, we can use this as an estimating equation for the function $u(R, \theta) - u(-1, \theta)$. The unknown utility function can be expanded into series of R and θ , so that it is very easy to estimate, and tests that bettors have identical preferences follow quite naturally.²³

With expected utility, the linearity in p allows us to circumvent the preliminary stage of estimating the probabilities. Defining $f_i = 1$ if horse i finished first, $f_i = 0$ otherwise, and noting that estimating probabilities from the data amounts in replacing p_i with $p_i(R) = E(f_i|R)$, we can use the properties of conditional expectation and obtain the following moment condition:

$$E(f_i(u(R_i, \theta_i(R)) - u(-1, \theta_i(R)))|R) = E(f_{i+1}(u(R_{i+1}, \theta_i(R)) - u(-1, \theta_i(R)))|R),$$

which can be estimated by linear regression methods from data on market shares and on which horse finished first.

CGSS apply this approach to a sample of about 50,000 races run in the U.S. in 2001.²⁴ Figure 5 gives a taste of their results, on expected utility preferences. It shows the estimated absolute risk-aversion indexes

$$\sigma(R, \theta) = -\frac{u''_{RR}(R, \theta)}{u'_R(R, \theta)}$$

for the four quantiles of the distribution of risk preferences $\theta = 0.3, 0.5, 0.7, 0.9$, along with the absolute risk-aversion index estimated under the assumption that all bettors have the same preferences ("EU Homo").

There is clearly much heterogeneity among preferences toward risk in this data,²⁵ and assuming homogeneity would be quite misleading. Beyond that, the nature of this heterogeneity is not easy to describe. All curves are U-shaped, suggesting risk-averse behavior on low- and high-probability bets, and risk-loving behavior in between.

²³This can be generalized to nonexpected utility directly, adding of course new terms that are non-linear in probability p .

²⁴Ideally, races should be differentiated by adding some covariates, and changes in the population of bettors should be modeled too.

²⁵Given the size of the sample, the estimates are very precise, and therefore we do not show confidence bands.

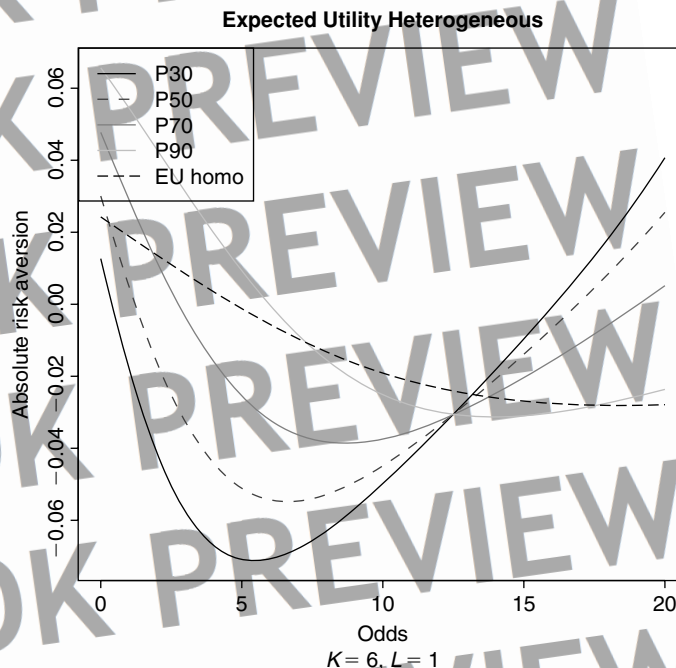


FIGURE 5 Estimated absolute risk aversions of heterogeneous bettors.

7. EXOTIC BETS

Snowberg and Wolfers (2007) exploit the existence of exotic bets²⁶ to build a non-parametric test of EU. They focus on so-called exactas, which are bets on the ordered first two horses. An exacta (i, j) wins if horse i wins and horse j is second. The idea of the test is that EU and NEU preferences that generate the same behavior on simple bets have different implications for exotic bets. Let p_{ij} be the conditional probability that horse j ends second when horse i wins. Let R_{ij} be the odds of the exacta (i, j) . Snowberg and Wolfers extend the standard model by imposing the condition that the representative bettor is also indifferent to the choice between simple bets and exactas. In the case of expected utility, we thus have $p_i p_{ij} [u(R_{ij}) - u(-1)] = p_i [u(R_i) - u(-1)]$ which reduces to

$$p_{ij} [u(R_{ij}) - u(-1)] = u(R_i) - u(-1).$$

They contrast this with models having distortions of probabilities and a linear utility function $[u(R) = R]$. They also assume that individuals do not reduce compound lotteries: defining $G(\cdot)$ as in Section 4, the distorted probability of winning an

²⁶Exotic bets are combinatorial bets allowing a bet on multiple horses in various orders.

exacta is $G(p_i)G(p_{ij})$ as opposed to $G(p_i p_{ij})$. Here the indifference condition becomes $G(p_i)G(p_{ij})(R_{ij} + 1) = G(p_i)(R_i + 1)$ or

$$G(p_{ij})(R_{ij} + 1) = R_i + 1.$$

Using a U.S. dataset, they then estimate non-parametrically the functions $u(\cdot)$ and $G(\cdot)$ on simple bets along the line of Weitzman (1965), and use the exactas to compare the EU model and the NEU model. The procedure has the advantage of being non-parametric, at the cost of ignoring race specific effects due, for instance, to the number of horses (in the above notation K_r is the same for all races). They conclude that the NEU model better fits the data than the EU model.

8. CONCLUDING REMARKS

This survey has attempted to describe the literature estimating and testing various utility representations on racetrack betting data. Clearly, much more work is required before some definite conclusion emerges. We hope to have convinced the reader that this type of study provides useful insight and is worth pursuing. In particular, the pattern that emerges is that the nature of the risk attitude that is embedded in odds and winners data is more complex than predicted by a simple risk-loving utility function, and may involve some elements of risk-aversion behavior as well. Assessing precisely which type of preference representation best fits the data would require more extensive studies. The methodology described here can apply to other existing theories, such as, for instance, regret theory (Loomes and Sugden, 1982) or disappointment theory (Gul, 1991).

As exposed in Kahneman et al. (1982), departures from expected utility involve more heuristics and biases than the static discrepancy between psychological probabilities and objective probabilities that can be captured by a non-linear preference function. The richness of the data available on horse races could help to test some of these other departures. This would require the researcher to collect more data than the odds and the winners, but there is, for instance, a potential to exploit information on the races, or the dynamics of odds. An attempt in this direction is Metzger (1985) who uses the ranking of the race during the day to provide some empirical supports for the gambler's fallacy (among others), here the effect of the outcome of the previous races on the perception of the respective winning probabilities of favorites and longshots. Ayton (1997) uses data on UK football gambling and horse races to study the support theory developed by Tversky and Koehler (1994), with mitigated conclusions.

We have focused on horse races studies. Other gambling modalities provide documented natural experiments. Because each type of gambling involves a different entertainment value and motivation of gamblers, it is difficult to compare the results obtained in different gambling contexts. Studies are still relatively scarce, and we will have to wait for more work before drawing any conclusion from the comparison of the patterns of behavior observed for various games. Still, let us mention that work has

been conducted for lotteries that sheds some light on the nature of cognitive biases.²⁷ For instance, it is well documented that the return varies with the numbers chosen (see Thaler and Ziemba, 1988; Simon, 1999). Simon (1999) and Papachristou (2004) also examine whether lotto data exhibit a gambler's fallacy pattern, with mixed conclusions. Televised gambling has also been examined. References can be found in Beetsma and Schotman (2001) estimating risk-aversion for participants to the Dutch television show *LINGO*, or in Février and Linnemer (2002) who conclude from a study of the French edition of the television show *The Weakest Link* that some pieces of information are not used by participants.²⁸ Finally, we should mention the recent work by Levitt (2004) using micro-data on gambling on the National Football League to analyze biases and skills in individual betting.

References

- Ali, M. 1977. Probability and Utility Estimate for Racetrack Bettors, *Journal of Political Economy* 85, 803–815.
- Allais, M. 1953. Le Comportement de l'Homme Rationnel Devant le Risque: Critique des Postulats et Axiomes de l'École Américaine, *Econometrica*, 21, 503–546; English translation in *The Expected Utility Hypothesis and the Allais Paradox*, M. Allais and O. Hagen (eds.), Reidel, Dordrecht, The Netherlands.
- Arrow, K. 1964. The Role of Securities in the Optimal Allocation of Risk Bearing, *Review of Economic Studies* 31, 91–96.
- Asch, P., B. Malkiel, and R. Quandt. 1982. Racetrack Betting and Informed Behavior, *Journal of Financial Economics* 10, 187–194.
- Ayton, P. 1997. How to be Incoherent and Seductive: Bookmakers' Odds and Support Theory, *Organizational Behavior and Human Decision Processes* 72(1), 99–115.
- Beetsma, R., and P. Schotman. 2001. Measuring Risk Attitude in a Natural Experiment: Data from the Television Game Show *LINGO*, *The Economic Journal* 111, 821–848.
- Bradley, I. 2003. Preferences of a Representative Bettor, *Economic Letters*, pp. 409–413.
- Bushe, K., and C. D. Hall. 1988. An Exception to the Risk Preference Anomaly, *Journal of Business* 61, 337–346.
- Camerer, C. 1995. Individual Decision Making in J. Kagel and A. Roth (eds.), *The Handbook of Experimental Economics*. Princeton University Press, Princeton, NJ.
- Chiappori, P.-A., A. Gandhi, B. Salanié, and F. Salanié. 2007. Heterogeneous Preferences Towards Risk: A Study on Horse Races, work in progress.
- Craft, N. F. R. 1985. Some Evidence of Insider Trading in Horse Race Betting in Britain, *Economica* 52, 295–304.
- Donkers, B., B. Melenberg, and A. Van Soest. 2001. Estimating Risk Attitude Using Lotteries: A Large Sample Approach, *The Journal of Risk and Uncertainty* 22, 165–165.
- Eadington, W. R. 1999. The Economics of Casino Gambling, *Journal of Economic Perspectives* 13(3), 173–192.
- Février, P., and L. Linnemer. 2006. Equilibrium Selection: Payoff or Risk Dominance? The case of the Weakest Link, *Journal of Economic Behavior and Organization* 60, 164–181.
- Friedman, M., and L. Savage. 1948. The Utility Analysis of Choices Involving Risk, *Journal of Political Economy* 56, 279–304.
- Gandhi, A. 2007. Rational Expectations at the Racetrack: Testing Expected Utility Using Prediction Market Prices. Mimeo, University of Wisconsin-Madison.

²⁷For a discussion of casino gambling see Eadington (1999).

²⁸Levitt (2003) also uses *The Weakest Link* to test for discriminatory behavior.

- Golec, J., and M. Tamarkin. 1998. Bettors Love Skewness, Not Risk, at the Horse Track, *Journal of Political Economy* 106(1), 205–225.
- Griffith, R. 1949. Odds Adjustment by American Horses Race Bettors, *American Journal of Psychology* 62, 290–294. Reprinted in Hausch, Lo, and Ziemba (1994).
- Guiso, L., Haliassos, M., and T. Jappelli. 2002. *Household Portfolios*. MIT Press, Cambridge, MA.
- Gul, F. 1991. A Theory of Disappointment Aversion, *Econometrica* 59, 667–686.
- Hausch, D., Lo, V. and W. Ziemba. 1994. Introduction to the Efficiency of Win Markets and the Favorite-Longshot Bias, in D. Hausch, V. Lo, and W. Ziemba (eds), *Efficiency of Racetrack Betting Markets*. Academic Press, San Diego, CA, pp. 251–253.
- Hausch, D. B., V. S. Lo, and W. T. Ziemba. 1994. *Efficiency of Racetrack Betting Markets*. Academic Press, San Diego, CA.
- Henery, R. 1985. On the Average Probability of Losing Bets on Horses with Given Starting Price Odds, *Journal of the Royal Statistical Society, Series A* 148, 342–349.
- Jullien, B., and B. Salanié. 1994. Measuring the Incidence of Insider Trading: A Comment on Shin, *Economic Journal* 104, 1418–1419.
- Jullien, B., and B. Salanié. 2000. Estimating Preferences under Risk: The Case of Racetrack Bettors, *Journal of Political Economy* 108, 503–530.
- Kahneman, D., and A. Tversky. 1979. Prospect Theory: An Analysis of Decision under Risk, *Econometrica* 47, 263–291.
- Kahneman, D., P. Slovic, and A. Tversky. 1982. *Judgement Under Uncertainty: Heuristics and Biases*. Cambridge University Press, Cambridge, UK.
- Kanto, A. J., G. Rosenqvist, and Suvas, A. 1992. On Utility Function Estimation of Racetrack Bettors, *Journal of Economic Psychology* 13, 491–498.
- Levitt, S. 2003. Testing Theories of Discrimination: Evidence from the “Weakest Link,” *Journal of Law and Economics* 47, 431–452.
- Levitt, S. 2004. How Do Markets Function? An Empirical Analysis of Gambling on the National Football League, *Economic Journal* 114, 2043–2066.
- Loomes, G., and R. Sugden. 1982. Regret Theory: An Alternative Theory of Rational Choice under Uncertainty, *Economic Journal* 92, 805–824.
- Machina, M. 1982. Expected Utility Analysis Without the Independence Axiom, *Econometrica* 50, 277–323.
- Markowitz, H. 1952. The Utility of Wealth, *Journal of Political Economy* 60, 151–158.
- McGlothlin, W. 1956. Stability of Choices Among Uncertain Alternatives, *American Economic Review* 69, 604–615.
- Metzger, M. A. 1985. Biases in Betting: An Application of Laboratory Findings, 2nd ed., in D. B. Hausch, V. S. Lo, and W. T. Ziemba (eds.), *Efficiency of Racetrack Betting Markets*. Academic Press, San Diego, CA.
- Ottaviani, M., and P. Sorensen. 2008. The Favorite Longshot Bias: An Overview of the Main Explanations, in D. B. Hausch and W. T. Ziemba (eds.), *Handbook of Sports and Lottery Markets*, North-Holland, Amsterdam, pp. 83–101.
- Papachristou, G. 2004. The British Gambler’s Fallacy, *Applied Economics* 36, 2073–2077.
- Prelec, D. 1998. The Probability Weighting Function, *Econometrica* 66, 497–527.
- Quandt, R. 1986. Betting and Equilibrium, *Quarterly Journal of Economics* 101, 201–207.
- Sauer, R. D. 1998. The Economics of Wagering Markets, *Journal of Economic Literature* 36, 2021–2064.
- Shin, H. S. 1991. Optimal Odds Against Insider Traders, *Economic Journal* 101, 1179–1185.
- Shin, H. S. 1992. Prices of State-Contingent Claims with Insider Traders, and the Favorite-Longshot Bias, *Economic Journal* 102, 426–435.
- Shin, H. S. 1993. Measuring the Incidence of Insider Trading in a Market for State-Contingent Claims, *Economic Journal* 103, 1141–1153.
- Simon, J. 1999. An Analysis of the Distribution of Combinations Chosen by UK National Lottery Players, *Journal of Risk and Uncertainty* 17, 243–276.
- Snowberg, E., and J. Wolfers. 2008. Examining Explanations of a Market Anomaly: Preferences or Perceptions? in D. B. Hausch and W. T. Ziemba (eds.), *Handbook of Sports and Lottery Markets*, North-Holland, Amsterdam, pp. 103–136.

- Thaler, R., and W. T. Ziemba. 1988. Anomalies—Parimutuel Betting Markets: Racetracks and Lotteries, *Journal of Economic Perspectives* 2, 161–174.
- Tompkins, R., W. T. Ziemba, and S. Hodges. 2008. The Favorite-Longshot Bias in S&P 500 and FTSE 100 Index Futures Options: The Return to Bets and the Cost of Insurance, in D. B. Hausch and W. T. Ziemba, *Handbook of Sports and Lottery Markets*, North-Holland, Amsterdam, pp. 161–180.
- Tversky, A., and D. Kahneman. 1992. Advances in Prospect Theory: Cumulative Representation of Uncertainty, *Journal of Risk and Uncertainty* 5, 297–323.
- Tversky, A., and D. Koehler. 1994. Support Theory: A Non-Extensional Representation of Subjective Probabilities, *Psychological Review* 101, 547–567.
- Vaughan Williams, L., ed. 2003, *The Economics of Gambling*, Routledge, London.
- Weitzman, M. 1965. Utility Analysis and Group Behavior: An Empirical Study, *Journal of Political Economy* 73, 18–26.

DOWNLOAD FULL VERSION HERE: <https://bettingbooks.org/12>

This page intentionally left blank

Download More Sports Betting Books at: www.bettingbooks.org

CHAPTER 4

Approximating the Ordering Probabilities of Multi-Entry Competitions by a Simple Method

Victor S. Y. Lo

Fidelity Investments, Boston, MA, USA.*

John Bacon-Shone

Social Sciences Research Centre, The University of Hong Kong, Hong Kong.

1. Introduction	52
2. Theoretical Results of the Limiting Cases	53
3. A Simple Approximation	56
3.1. <i>Empirical Analysis for the Approximated Henery Model</i>	56
3.2. <i>Empirical Analysis for the Stern Model</i>	57
4. Conclusion	59
<i>References</i>	60
Appendix: Proof of Theorem 1	61

Acknowledgments

We acknowledge Kelly Busche and Richard Quandt for the Hong Kong and the Meadowlands data, respectively.

*This work was completed while Victor Lo was with The University of Hong Kong and the University of British Columbia.

Abstract

To predict the ordering probabilities of multi-entry competitions (e.g., horse races), Harville (1973) proposed a simple way of computing the ordering probabilities based on the simple winning probabilities. This simple model is implied by assuming that the underlying model (e.g., running times in horse racing) is the independent exponential or extreme-value distribution. Henery (1981) and Stern (1990) proposed to use normal and gamma distributions, respectively, for the running time. However, both the Henery and Stern models are too complicated to use in practice. Bacon-Shone et al. (1992b) have shown that the Henery and Stern models fit better than the Harville model for particular horse racing datasets. In this chapter, we first give a theoretical result for the limiting case that all the horses have the same abilities. This theoretical result motivates an approximation of ordering probabilities for the Henery and Stern models. We then show empirically that this approximation works well in practice.

Keywords: ordering probabilities, horse races, running-time distributions.

1. INTRODUCTION

In multi-entry competitions, Harville (1973) proposed to use the following formula to compute the ordering probabilities:

$$\pi_{ij} = \frac{\pi_i \pi_j}{1 - \pi_i} \quad (1)$$

where $\pi_{ij} = P(i \text{ wins and } j \text{ finishes second})$, and

$$\pi_i = P(i \text{ wins}).$$

In horse racing, i and j are two horses, and the value of π_i can be estimated by the win bet fraction (see Ali, 1977; Synder, 1978; Busche and Hall, 1988; and Bacon-Shone et al., 1992a for details of using the win bet fractions). The case is similar for more complicated ordering probabilities.

Although Harville (1973) did not relate his model to any probability distribution, the simple formula in Equation (1) can be implied from the assumption of independent exponential distributions for running times with different scale parameters (Dansie, 1983) for each horse in each race or independent extreme-value distributions with different location parameters. Henery (1981) proposed assuming independent normal distributions for the running times (hereafter called the Henery model), that is, the

running time of the i th horse, $T_i \sim N(\theta_i, 1)$. Then

$$\begin{aligned}\pi_{ij} &= P(T_i < T_j < \min_{r \neq i,j} \{T_r\}) \\ &= \int_{-\infty}^{\infty} \Phi(u + \theta_j - \theta_i) \prod_{r \neq i,j} [1 - \Phi(u + \theta_j - \theta_r)] \phi(u) du,\end{aligned}$$

where the θ_i 's are found by solving

$$\pi_i = \int_{-\infty}^{\infty} \prod_{r \neq i} \Phi(u + \theta_i - \theta_r) \phi(u) du.$$

Hence, numerical integration or an approximation method is required. Similarly, Stern (1990) proposed to use independent gamma distributions with fixed integral shape parameter r , that is, $T_i \sim G(r, \theta_i)$. Similar to the Henery model, no closed form has been found for computing the ordering probabilities. For descriptions of the three models, see Bacon-Shone et al. (1992b). Thus, to apply these models in practice, a good approximation is essential; otherwise the Harville model has to be used (e.g., Hausch et al., 1981).

Bacon-Shone et al. (1992b) reported many empirical analyses of different complicated bets. Their conclusion is that using the information from win bet fractions alone, for the analyses of an exacta bet (in the Meadowlands), trifecta bet (in the Meadowlands and Hong Kong), and quinella bet (in Hong Kong), the Henery model was found to be better than the others in predicting the relevant ordering probabilities for those bets according to a likelihood approach. For details, see Bacon-Shone et al. (1992b). However, an exceptional case has been found in Japan where the Stern model with a fixed shape parameter is better than the Henery or Harville models (see Lo and Bacon-Shone, 1992).

In this chapter, we first give a limiting result for some functions of the ordering probabilities when all horses have equal abilities in Section 2. This motivates the method of approximation for the ordering probabilities. Empirical results which illustrate the accuracy of this approximation are included in Section 3. The conclusion is given in Section 4.

2. THEORETICAL RESULTS OF THE LIMITING CASES

In this section, we will show that some functions of the ordering probabilities under the Henery and Stern models will become constants when all the horses have equal abilities (i.e., equal mean running times).

Define

$$\lambda_{ijl}^{\text{Hen}} = \frac{\ln(\pi_{ij}/\pi_{il})}{\ln(\pi_j/\pi_l)} \quad (2)$$

and

$$\tau_{ijkl}^{\text{Hen}} = \frac{\ln(\pi_{ijk}/\pi_{ijl})}{\ln(\pi_k/\pi_l)} \quad (3)$$

where π_i = the winning probability of horse i which can be estimated by the win bet fraction of horse i ; $\pi_{ij} = P(\text{horse } i \text{ wins and } j \text{ finishes second})$ under the Henery model and $\pi_{ijk} = P(\text{horse } i \text{ wins, } j \text{ finishes second, and } k \text{ finishes third})$ under the Henery model. Then, the following theorem holds:

Theorem 1. When the running times follow independent normal distributions with means $\theta_1, \dots, \theta_n$, and let $\mu_{i:n}$ be the i th expected normal order statistics, we have:

$$\begin{aligned} \text{(a)} \quad \lim_{\theta_l \rightarrow \theta_j} \lambda_{ijl}^{\text{Hen}} &= \lim_{\theta_l \rightarrow \theta_j} \frac{\ln(\pi_{ij}/\pi_{il})}{\ln(\pi_j/\pi_l)} \quad \text{where } \theta_s = \theta_j \text{ for all } s \neq l \\ &= \left(1 - \frac{1}{n}\right) \frac{\mu_{2:n}}{\mu_{1:n}} + \frac{n-1}{n(n-2)} \left(\frac{\mu_{1:n} + \mu_{2:n}}{\mu_{1:n}}\right) \quad (n \geq 3) \\ &\simeq \frac{\mu_{2:n}}{\mu_{1:n}} \quad \text{when } n \text{ is large,} \end{aligned} \quad (4)$$

$$\begin{aligned} \text{(b)} \quad \lim_{\theta_l \rightarrow \theta_k} \tau_{ijkl}^{\text{Hen}} &= \lim_{\theta_l \rightarrow \theta_k} \frac{\ln(\pi_{ijk}/\pi_{ijl})}{\ln(\pi_k/\pi_l)} \quad \text{where } \theta_s = \theta_k \text{ for all } s \neq l \\ &= \left(\frac{n-1}{n}\right) \left(\frac{1}{n-3}\right) \left(1 + \frac{\mu_{2:n} + \mu_{3:n}}{\mu_{1:n}}\right) \\ &\quad + \left(\frac{n-1}{n}\right) \frac{\mu_{3:n}}{\mu_{1:n}} \quad (n \geq 4) \\ &\simeq \frac{\mu_{3:n}}{\mu_{1:n}} \quad \text{when } n \text{ is large.} \end{aligned} \quad (5)$$

The proof of Theorem 1 is given in the Appendix.

The above two limiting values, Equations (4) and (5), are reported in Table 1 for different race sizes, n .

For the Stern model, define

$$\lambda_{ijl}^{(r)} = \frac{\ln(\pi_{ij}^{(r)}/\pi_{il}^{(r)})}{\ln(\pi_j/\pi_l)} \quad (6)$$

$$\tau_{ijkl}^{(r)} = \frac{\ln(\pi_{ijk}^{(r)}/\pi_{ijl}^{(r)})}{\ln(\pi_k/\pi_l)} \quad (7)$$

where the ordering probabilities in Equations (6) and (7) are based on the Stern model with shape parameter r .

A similar theorem holds for the Stern model.

TABLE 1 Limiting Values of $\lambda_{ijl}^{\text{Hen}}$ and τ_{ijk}^{Hen}

Race size	$\lim \lambda_{ijl}^{\text{Hen}}$	$\lim \tau_{ijk}^{\text{Hen}}$
3	0.6667	—
4	0.6996	0.5336
5	0.7207	0.5703
6	0.7359	0.5952
7	0.7475	0.6138
8	0.7569	0.6285
9	0.7648	0.6406
10	0.7714	0.6508
11	0.7771	0.6596
12	0.7822	0.6672
13	0.7867	0.6741
14	0.7907	0.6802

Theorem 2. When the running times follow independent Gamma distributions with shape parameter r , let $\mu_{i;n}^{(r)}$ be the associated i th expected order statistics and $\theta_1, \dots, \theta_n$ the scale parameters of the running times, then

$$\begin{aligned}
 \text{(a)} \quad \lim_{\theta_l \rightarrow \theta_j} \lambda_{ijl}^{(r)} &= \lim_{\theta_l \rightarrow \theta_j} \frac{\ln(\pi_{ij}/\pi_{il})}{\ln(\pi_j/\pi_l)} \quad \text{where } \theta_s = \theta_j \text{ for all } s \neq l \\
 &= \frac{(n-1)^2}{n(n-2)} \left[\frac{(r - \mu_{1;n}^{(r)}) + (r - \mu_{2;n}^{(r)})}{r - \mu_{1;n}^{(r)}} \right] - \frac{n-1}{n} \quad (n \geq 3) \\
 &\simeq \frac{r - \mu_{2;n}^{(r)}}{r - \mu_{1;n}^{(r)}} \quad \text{when } n \text{ is very large,}
 \end{aligned}$$

$$\begin{aligned}
 \text{(b)} \quad \lim_{\theta_l \rightarrow \theta_k} \tau_{ijk}^{(r)} &= \lim_{\theta_l \rightarrow \theta_k} \frac{\ln(\pi_{ijk}/\pi_{ijl})}{\ln(\pi_k/\pi_l)} \quad \text{where } \theta_s = \theta_k \text{ for all } s \neq l \\
 &= \left(\frac{n-1}{n} \right) \left(\frac{n-2}{n-3} \right) \left[\frac{1}{(n-2)} \right. \\
 &\quad \left. + \frac{(r - \mu_{2;n}^{(r)})/(n-2) + (r - \mu_{3;n}^{(r)})}{r - \mu_{1;n}^{(r)}} \right] \quad (n \geq 4) \\
 &\simeq \frac{r - \mu_{3;n}^{(r)}}{r - \mu_{1;n}^{(r)}} \quad \text{when } n \text{ is large.}
 \end{aligned}$$

The proof of Theorem 2 is omitted because it is very similar to that of Theorem 1.

3. A SIMPLE APPROXIMATION

For the Henery and Stern models, computation of π_{ij} is not simple because it involves multidimensional numerical integrations. To use the two models in practice, we need good approximations. From Equations (2) and (3), if $\lambda_{ijl}^{\text{Hen}} \simeq \lambda_{ijl}^{\text{Hen}} = \lambda^{\text{Hen}}$ and $\tau_{ijkl}^{\text{Hen}} \simeq \lim \tau_{ijkl}^{\text{Hen}} = \tau^{\text{Hen}}$, where λ^{Hen} and τ^{Hen} depend on n , the race size, then

$$\pi_{ijk} \simeq \pi_i \frac{\pi_j^{\lambda^{\text{Hen}}}}{\sum_{s \neq i} \pi_s^{\lambda^{\text{Hen}}}} \frac{\pi_k^{\tau^{\text{Hen}}}}{\sum_{t \neq ij} \pi_t^{\tau^{\text{Hen}}}} \quad (8)$$

Similarly for the Stern model

$$\pi_{ijk} \simeq \pi_i \frac{\pi_j^{\lambda^{(r)}}}{\sum_{s \neq i} \pi_s^{\lambda^{(r)}}} \frac{\pi_k^{\tau^{(r)}}}{\sum_{t \neq ij} \pi_t^{\tau^{(r)}}} \quad (9)$$

where Equations (8) and (9) are hereafter called the discount model.

3.1. Empirical Analysis for the Approximated Henery Model

We now show that the approximation in Equation (8) works well in practice. First, consider some summary values of $\lambda_{ijl}^{\text{Hen}}$ and τ_{ijkl}^{Hen} for different race sizes in Hong Kong, the Meadowlands, and Japan through numerical integration as shown in Tables 2a and 2b. From Tables 2a and 2b, we can observe that the mean values are close to the limiting values shown in Table 1.

We now compare the above approximation with different models. The exact Henery models are based on numerical integrations. We use two approximations: (i) fixed values: $\lambda^{\text{Hen}} = 0.76$ and $\tau^{\text{Hen}} = 0.62$; (ii) varying values: λ^{Hen} and τ^{Hen} vary over the race size n . We also compare our approximation with Henery's approximation using first order Taylor series expansion. The empirical results are shown in Table 3. For example, the log likelihood for trifecta bets is given by $\sum_l \ln \pi_{[123],l}$ where $[1, 2, 3]$, l denotes the three horses finishing in the top three positions in race l .

From Table 3, it is clear that the accuracy (measured by the log likelihood) of the discount model is close to that of the Henery model. And the log likelihood values produced by Henery's approximation deviate a lot from those produced by the exact Henery model. For our discount model, varying values of λ^{Hen} and τ^{Hen} do not seem to be much better than a fixed value approximation since the log likelihood values of both approximations are very close to the log likelihood value of the exact Henery model. Hence, we conclude that the discount model with fixed values of λ^{Hen} and τ^{Hen} is good enough to use in practice.

TABLE 2a Summary Values of λ^{Hen} and τ^{Hen} for Different Race Sizes in Hong Kong*

Race size	No. of races	λ^{Hen}		τ^{Hen}	
		Mean	SD	Mean	SD
4	4	0.6685	0.0405	0.5186	0.0125
5	6	0.6914	0.0369	0.5580	0.0208
6	20	0.7209	0.0213	0.5877	0.0163
7	30	0.7407	0.0206	0.6127	0.0186
8	78	0.7507	0.0191	0.6251	0.0216
9	54	0.7620	0.0160	0.6399	0.0224
10	88	0.7654	0.0155	0.6519	0.0245
11	28	0.7726	0.0173	0.6569	0.0232
12	42	0.7776	0.0142	0.6648	0.0260
13	28	0.7821	0.0176	0.6725	0.0262
14	43	0.7859	0.0172	0.6786	0.0338
Overall	421	0.7669	0.0238	0.6518	0.0340

*In this table, i and j are the horses finishing first and second, respectively, with k and l varying over all the other horses for $\lambda_{ijl}^{\text{Hen}}$ and $\lambda_{ijkl}^{\text{Hen}}$.

TABLE 2b Summary Values of λ^{Hen} and τ^{Hen} for Different Race Sizes in the Meadowlands

Race size	No. of races	λ^{Hen}		τ^{Hen}	
		Mean	SD	Mean	SD
6	10	0.7089	0.0294	0.5817	0.0219
7	16	0.7287	0.0253	0.6077	0.0233
8	59	0.7455	0.0223	0.6214	0.0256
9	119	0.7522	0.0202	0.6358	0.0283
10	275	0.7605	0.0189	0.6451	0.0302
11	20	0.7588	0.0186	0.6262	0.1343
12	11	0.7573	0.0312	0.6644	0.0328
Overall	510	0.7561	0.0216	0.6393	0.0427

3.2. Empirical Analysis for the Stern Model

The limiting values of $\lambda^{(r)}$ and $\tau^{(r)}$ for the Stern model in Theorem 2 depend on both the race size (n) and the shape parameter (r). We have computed the limiting values for $n = 2, \dots, 18$ and $r = 2, \dots, 8$ and observed that the dependence on n is not very strong. Fixing $n = 11$, the limiting values are shown in Table 4. These limiting values

TABLE 3 Comparison Among Different Models in Different Bet Types

Models	Log likelihood	Models	Log likelihood
Exacta: 510 races (the Meadowlands)		Quinella: 4,153 races (Hong Kong)	
Harville	−1,875.77	Harville	−13,619.28
Henery	−1,859.63	Henery	−13,589.55
Henery's approximation	−1,872.81	Henery's approximation	−13,626.43
Discount:		Discount:	
(i) Fixed values	−1,859.25	(i) Fixed values	−13,586.95
(ii) Varying values	−1,859.40	(ii) Varying values	−13,586.99
Trifecta: 120 races (the Meadowlands)		Trifecta: 1,809 races (Hong Kong)	
Harville	−711.50	Harville	−10,747.98
Henery	−699.83	Henery	−10,667.25
Henery's approximation	−703.12	Henery's approximation	−10,689.61
Discount:		Discount:	
(i) Fixed values	−699.68	(i) Fixed values	−10,667.80
(ii) Varying values	−700.03	(ii) Varying values	−10,666.87

TABLE 4 Limiting Values of $\lambda^{(r)}$ and $\tau^{(r)}$ for $n = 11$

r	$\lim \lambda^{(r)}$	$\lim \tau^{(r)}$
2	0.93	0.89
3	0.90	0.84
4	0.88	0.81
5	0.87	0.80
6	0.86	0.78
7	0.86	0.77
8	0.85	0.76

will be used for our approximation. We also compute the empirical summary values of $\lambda^{(r)}$ and $\tau^{(r)}$ using Japanese data in Table 5. The limiting values in Table 4 and the mean values in Table 5 are close to each other.

Lo and Bacon-Shone (1992) reported that the Stern model is better than the Harville and Henery models in Japan. We compare the log likelihood values using numerical integration and the discount model in Table 6. For the discount model, again, two alternatives are tried: (i) fixed values of $\lambda^{(r)}$ and $\tau^{(r)}$ (by fixing $n = 11$ in Table 4);

TABLE 5 Summary Values of $\lambda_{ijl}^{(r)}$ and $\tau_{ijkl}^{(r)*}$

r	$\lambda^{(r)}$		$\tau^{(r)}$	
	Mean	SD	Mean	SD
2	0.9336	0.0135	0.8920	0.0269
3	0.9021	0.0168	0.8423	0.0250
4	0.8836	0.0186	0.8140	0.0266
5	0.8712	0.0198	0.7953	0.0278
6	0.8623	0.0206	0.7819	0.0286
7	0.8555	0.0214	0.7717	0.0292
8	0.8500	0.0219	0.7636	0.0297

*In this table, i, j, k represent the horses finishing in the top three positions in each race and l is varying over all the other horses for $\lambda_{ijl}^{(r)}$ and $\tau_{ijkl}^{(r)}$.

TABLE 6 Log Likelihood Values Under the Stern Model for Japanese Data

r	Numerical integrations	(i)	(ii)
		Fixed values	Varying values
2	-8,954.57	-8,955.98	-8,956.59
3	-8,950.60	-8,952.31	-8,953.11
4	-8,950.35	-8,952.11	-8,952.82
5	-8,950.94	-8,952.61	-8,953.36
6	-8,951.82	-8,953.45	-8,954.12
7	-8,952.65	-8,953.67	-8,954.90
8	-8,953.44	-8,954.89	-8,955.66

(ii) varying values of $\lambda^{(r)}$ and $\tau^{(r)}$ over race size, n . From Table 6, we observe that the difference between (i) and (ii) above is not great and thus we recommend using the values in Table 4 for approximation.

4. CONCLUSION

We have proposed using the discount model in Equations (8) and (9) with different parameter values. This model has been shown to provide a good approximation to both the Henery and Stern models. It also includes the Harville model ($r = 1$). More empirical evidence is given in Lo and Bacon-Shone (1992). To apply the model in practice (e.g., betting), we recommend collecting relevant data and choosing the most appropriate model, whether Henery or Stern (r), using a likelihood comparison and

then applying Equations (8) and (9) using appropriate parameter values. Alternatively, we can estimate λ and τ directly through logistic modeling, for example, see Lo and Bacon-Shone (1994) and Lo (1994). The effect of this improved probability estimation on betting strategy (e.g., the Dr. Z system proposed by Hausch et al. 1981) may result in better returns, see Lo et al. (1995) and Hausch et al. (1994). We assume that the win bet fraction is a good estimate of the win probability. Bacon-Shone et al. (1992a) suggest a method to remove any bias using a logit model before estimating the ordering probabilities. For future research opportunities, one can derive a similar approximation to a more complex probability model in Lo and Bacon-Shone (2008).

References

- Ali, M. M. 1977. Probability and utility estimates for racetrack bettors, *Journal of Political Economy* 84, 803–815.
- Bacon-Shone, J. H., V. S. Y. Lo, and K. Busche. (1992a). Modelling winning probability. Research report #10, Department of Statistics, University of Hong Kong.
- Bacon-Shone, J. H., V. S. Y. Lo, and K. Busche. (1992b). Logistic Analyses of Complicated Bets. Research report #11, Department of Statistics, University of Hong Kong.
- Busche, K., and C. D. Hall. 1988. An Exception to the Risk Preference Anomaly, *Journal of Business*, 61, 337–346.
- Dansie, B. R. 1983. A Note on Permutation Probabilities, *Journal of Royal Statistical Society B* 45(1), 22–24.
- Harville, D. A. 1973. Assigning Probabilities to the Outcomes of Multi-Entry Competitions, *Journal of the American Statistical Association* 68, 312–316.
- Hausch, D. B., W. T. Ziemba, and M. Rubinstein. 1981. Efficiency of the Market for Racetrack Betting, *Management Science*, 27, 1435–1452.
- Hausch, D. B., V. S. Y. Lo, and W. T. Ziemba. 1994. Pricing Exotic Racetrack Wagers, in D. B. Hausch, V. S. Y. Lo, and W. T. Ziemba (eds.), *Efficiency of Racetrack Betting Markets*. Academic Press, San Diego, CA, pp. 469–483.
- Henery, R. J. 1981. Permutation Probabilities as Models for Horse Races, *Journal of Royal Statistical Society B* 43(1), 86–91.
- Lo, V. S. Y. 1994. Application of Logit Models in Racetrack Data, in D. B. Hausch, V. S. Y. Lo, and W. T. Ziemba (eds.), *Efficiency of Racetrack Betting Markets*. Academic Press, San Diego, CA, pp. 307–314, 2nd edition. World Scientific, Singapore, 2008.
- Lo, V. S. Y., and J. Bacon-Shone. 1992. An Approximation to Ordering Probabilities of Multi-Entry Competitions. Research report #16, Department of Statistics, University of Hong Kong.
- Lo, V. S. Y., and J. Bacon-Shone. 1994. A Comparison Between Two Models for Predicting Ordering Probabilities in Multiple-Entry Competitions, *The Statistician* 43(2) 317–327.
- Lo, V. S. Y. and J. Bacon-Shone. 2008. Probability and Statistical Models for Racing, *J. of Quantitative Analysis in Sports* 4, issue 2, article 11.
- Lo, V. S. Y., J. Bacon-Shone, and K. Busche. 1995. The Application of Ranking Probability Models to Racetrack Betting, *Management Science* 41, 1048–1059.
- Snyder, W. W. 1978. Horse Racing: Testing the Efficient Markets Model, *Journal of Finance*, 33, 1109–1118.
- Stern, H. 1990. Models for Distributions on Permutations, *Journal of the American Statistical Association* 85(410), 558–564.

APPENDIX: Proof of Theorem 1

(a) Let $\theta_l/\theta_j = b$ and $\theta_s = \theta_j$ for $s \neq j, l$. Then,

$$\begin{aligned} \lim_{\theta_l \rightarrow \theta_j} \lambda_{ijl}^{\text{Hen}} &= \lim_{\theta_l \rightarrow \theta_j} \frac{\ln(\pi_{ij}/\pi_{il})}{\ln(\pi_j/\pi_l)} \\ &= \lim_{b \rightarrow 1} \frac{(\pi_{il}/\pi_{ij}) \frac{\partial(\pi_{ij}/\pi_{il})}{\partial b}}{(\pi_l/\pi_j) \frac{\partial(\pi_j/\pi_l)}{\partial b}} \quad \text{by L'Hospital's rule} \\ &= \lim_{b \rightarrow 1} \frac{(\pi_{il} \frac{\partial \pi_{ij}}{\partial b} - \pi_{ij} \frac{\partial \pi_{il}}{\partial b}) / \pi_{il}^2}{(\pi_l \frac{\partial \pi_j}{\partial b} - \pi_j \frac{\partial \pi_l}{\partial b}) / \pi_l^2} \quad \text{assuming this limit exists. (A.1)} \end{aligned}$$

Consider

$$\begin{aligned} \pi_j &= P(T_j < T_s, \forall s \neq j), \quad T_s \sim N(\theta_s, 1) \\ &= \int_{-\infty}^{\infty} \prod_{s \neq j} \Phi(u - \theta_j + \theta_s) \phi(u) du, \end{aligned}$$

where $\phi(\cdot)$ and $\Phi(\cdot)$ denote the pdf and cdf of standard normal, respectively.

$$= \int_{-\infty}^{\infty} \Phi(u)^{n-2} \Phi[u + \theta_j(b-1)] \phi(u) du,$$

then

$$\begin{aligned} \frac{\partial \pi_j}{\partial b} &= \int_{-\infty}^{\infty} \Phi(u)^{n-2} \phi[u + \theta_j(b-1)] \theta_j \phi(u) du \\ &\rightarrow \theta_j \int_{-\infty}^{\infty} \Phi(u)^{n-2} \phi(u)^2 du \quad \text{as } b \rightarrow 1 \\ &= -\frac{\mu_{1;n}}{n(n-1)} \theta_j, \quad \text{using integration by parts} \end{aligned} \quad (\text{A.2})$$

and

$$\begin{aligned} \pi_l &= \int_{-\infty}^{\infty} \prod_{s \neq l} \Phi(u - \theta_l + \theta_s) \phi(u) du \\ &= \int_{-\infty}^{\infty} \Phi[u + \theta_j(1-b)]^{n-1} \phi(u) du, \end{aligned}$$

$$\begin{aligned}\frac{\partial \pi_l}{\partial b} &= -\theta_j(n-1) \int_{-\infty}^{\infty} \Phi[u + \theta_j(1-b)]^{n-2} \phi[u + \theta_j(1-b)] \phi(u) du \\ &\rightarrow -\theta_j(n-1) \int_{-\infty}^{\infty} \phi(u)^2 \Phi(u)^{n-2} du \quad \text{as } b \rightarrow 1 \\ &= \frac{\mu_{1;n} \theta_j}{n},\end{aligned}\tag{A.3}$$

thus, from Equations (A.2) and (A.3),

$$\frac{\lim_{b \rightarrow 1} \left(\pi_l \frac{\partial \pi_j}{\partial b} - \pi_j \frac{\partial \pi_l}{\partial b} \right)}{\pi_l^2} = - \left(\frac{n}{n-1} \right) \mu_{1;n} \theta_j.\tag{A.4}$$

In addition,

$$\begin{aligned}\pi_{ij} &= P(T_i < T_j < T_s, \quad \forall s \neq i, j) \\ &= \int_{-\infty}^{\infty} [1 - \Phi(u - \theta_j + \theta_i)] \prod_{s \neq i, j} \Phi(u - \theta_j + \theta_s) \phi(u) du \\ &= \int_{-\infty}^{\infty} [1 - \Phi(u)] \Phi(u)^{n-3} \Phi[u + \theta_j(b-1)] \phi(u) du, \\ \frac{\partial \pi_{ij}}{\partial b} &= \int_{-\infty}^{\infty} [1 - \Phi(u)] \Phi(u)^{n-3} \phi[u + \theta_j(b-1)] \theta_j \phi(u) du \\ &\rightarrow \theta_j \int_{-\infty}^{\infty} \phi(u)^2 [1 - \Phi(u)] \Phi(u)^{n-3} du \quad \text{as } b \rightarrow 1 \\ &= \frac{\theta_j}{n-2} \int_{-\infty}^{\infty} \Phi(u)^{n-2} \{u \phi(u) [1 - \Phi(u)] + \phi(u)^2\} du \\ &= -\frac{\mu_{1;n} + \mu_{2;n}}{n(n-1)(n-2)} \theta_j, \text{ using integration by parts}\end{aligned}\tag{A.5}$$

and

$$\begin{aligned}\pi_{il} &= \int_{-\infty}^{\infty} [1 - \Phi(u - \theta_l + \theta_i)] \prod_{s \neq i, l} \Phi(u - \theta_l + \theta_s) \phi(u) du \\ &= \int_{-\infty}^{\infty} \{1 - \Phi[u + \theta_j(b-1)]\} \Phi[u + \theta_j(1-b)]^{n-2} \phi(u) du,\end{aligned}$$

$$\begin{aligned}
 \frac{\partial \pi_{il}}{\partial b} &= \int_{-\infty}^{\infty} \left\{ -\phi[u + \theta_j(1-b)](-\theta_j) \Phi[u + \theta_j(1-b)]^{n-2} \right. \\
 &\quad \left. + \{1 - \Phi[u + \theta_j(1-b)]\} (n-2) \Phi[u + \theta_j(1-b)]^{n-3} \right. \\
 &\quad \left. \phi[u + \theta_j(1-b)](-\theta_j) \right\} \phi(u) du \\
 &\rightarrow \theta_j \int_{-\infty}^{\infty} \phi(u)^2 \Phi(u)^{n-2} du \\
 &\quad - \theta_j(n-2) \int_{-\infty}^{\infty} \phi(u)^2 [1 - \Phi(u)] \Phi(u)^{n-3} du \quad \text{as } b \rightarrow 1 \\
 &= \frac{\mu_{2;n}}{n(n-1)} \theta_j, \text{ using integration by parts}
 \end{aligned} \tag{A.6}$$

thus from Equations (A.5) and (A.6)

$$\frac{\lim_{b \rightarrow 1} \left(\pi_{il} \frac{\partial \pi_{ij}}{\partial b} - \pi_{ij} \frac{\partial \pi_{il}}{\partial b} \right)}{\pi_{il}^2} = -\theta_j \left(\frac{\mu_{1;n} + \mu_{2;n}}{n-2} + \mu_{2;n} \right). \tag{A.7}$$

Hence, from (A.1), (A.4), and (A.7)

$$\begin{aligned}
 \lim_{\theta_l \rightarrow \theta_j} \lambda_{ijl}^{\text{Hen}} &= \frac{\mu_{2;n} + \frac{\mu_{1;n} + \mu_{2;n}}{n-2}}{\left(\frac{n}{n-1} \right) \mu_{1;n}} \\
 &= \left(1 - \frac{1}{n} \right) \frac{\mu_{2;n}}{\mu_{1;n}} + \frac{n-1}{n(n-2)} \left(\frac{\mu_{1;n} + \mu_{2;n}}{\mu_{1;n}} \right).
 \end{aligned}$$

(b) Let $\theta_l/\theta_k = a$ and $\theta_s = \theta_k$ for $s \neq k, l$. Then,

$$\begin{aligned}
 \lim_{\theta_l \rightarrow \theta_k} \tau_{ijkl}^{\text{Hen}} &= \lim_{\theta_l \rightarrow \theta_k} \frac{\ln(\pi_{ijk}/\pi_{ijl})}{\ln(\pi_k/\pi_l)} \quad \text{where } \theta_s = \theta_k \ (s \neq l, k) \\
 &= \lim_{a \rightarrow 1} \frac{(\pi_{ijl}/\pi_{ijk}) \frac{\partial(\pi_{ijk}/\pi_{ijl})}{\partial a}}{(\pi_l/\pi_k) \frac{\partial(\pi_k/\pi_l)}{\partial a}} \quad \text{by L'Hospital's rule} \\
 &= \lim_{a \rightarrow 1} \frac{(\pi_{ijl} \frac{\partial \pi_{ijk}}{\partial a} - \pi_{ijk} \frac{\partial \pi_{ijl}}{\partial a})/\pi_{ijl}^2}{(\pi_l \frac{\partial \pi_k}{\partial a} - \pi_k \frac{\partial \pi_l}{\partial a})/\pi_l^2} \quad \text{assuming this limit exists.}
 \end{aligned} \tag{A.8}$$

Using the result obtained in (a),

$$\frac{\lim_{a \rightarrow 1} \left(\pi_l \frac{\partial \pi_k}{\partial a} - \pi_k \frac{\partial \pi_l}{\partial a} \right)}{\pi_l^2} = - \left(\frac{n}{n-1} \right) \mu_{1;n} \theta_k. \quad (\text{A.9})$$

Now

$$\begin{aligned} \pi_{ijk} &= P(T_i < T_j < T_k < T_s, \quad \forall s \neq i, j, k) \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{u+\theta_k-\theta_j} \Phi(v + \theta_j - \theta_i) \phi(v) dv \prod_{s \neq ijk} [1 - \Phi(u + \theta_k - \theta_s)] \phi(u) du \\ &= \frac{1}{2} \int_{-\infty}^{\infty} \Phi(u)^2 [1 - \Phi(u)]^{n-4} \{1 - \Phi[u + \theta_k(1-a)]\} \phi(u) du, \\ \frac{\partial \pi_{ijk}}{\partial a} &= \frac{1}{2} \int_{-\infty}^{\infty} \Phi(u)^2 [1 - \Phi(u)]^{n-4} \{-\phi[u + \theta_k(1-a)]\} (-\theta_k) \phi(u) du \\ &\rightarrow \frac{\theta_k}{2} \int_{-\infty}^{\infty} \phi(u)^2 \Phi(u)^2 [1 - \Phi(u)]^{n-4} du \quad \text{as } a \rightarrow 1 \\ &= \frac{\theta_k}{2} \frac{1}{n-3} \int_{-\infty}^{\infty} [1 - \Phi(u)]^{n-3} [-u\phi(u)\Phi(u)^2 + 2\phi(u)^2\Phi(u)] du \\ &= \frac{-2}{n(n-1)(n-2)(n-3)} (\mu_{1;n} + \mu_{2;n} + \mu_{3;n}), \end{aligned}$$

using integration by parts

(A.10)

and

$$\begin{aligned} \pi_{ijl} &= \int_{-\infty}^{\infty} \int_{-\infty}^{u+\theta_l-\theta_j} \Phi(v + \theta_j - \theta_i) \phi(v) dv \prod_{s \neq ijl} [1 - \Phi(u + \theta_l - \theta_s)] \phi(u) du \\ &= \frac{1}{2} \int_{-\infty}^{\infty} \Phi[u + \theta_k(a-1)]^2 \{1 - \Phi[u + \theta_k(a-1)]\}^{n-3} \phi(u) du, \end{aligned}$$

$$\begin{aligned} \frac{\partial \pi_{ijl}}{\partial a} &= \frac{1}{2} \int_{-\infty}^{\infty} \left\{ 2\Phi[u + \theta_k(a-1)] \phi[u + \theta_k(a-1)] \theta_k \{1 - \Phi[u + \theta_k(a-1)]\}^{n-3} \right. \\ &\quad + \Phi[u + \theta_k(a-1)]^2 (n-3) \{1 - \Phi[u + \theta_k(a-1)]\}^{n-4} \\ &\quad \left. \{-\phi[u + \theta_k(a-1)]\} \theta_k \right\} \phi(u) du \\ &\rightarrow \frac{\theta_k}{2} \left\{ 2 \int_{-\infty}^{\infty} \phi(u)^2 \Phi(u) [1 - \Phi(u)]^{n-3} du \right. \\ &\quad \left. - (n-3) \int_{-\infty}^{\infty} \phi(u)^2 \Phi(u)^2 [1 - \Phi(u)]^{n-4} du \right\} \text{ as } a \rightarrow 1 \\ &= \frac{\mu_{3;n} \theta_k}{n(n-1)(n-2)}, \text{ using integration by parts} \end{aligned} \quad (\text{A.11})$$

thus, from Equations (A.10) and (A.11)

$$\frac{\lim_{a \rightarrow 1} \left(\pi_{ijl} \frac{\partial \pi_{ijk}}{\partial a} - \pi_{ijk} \frac{\partial \pi_{ijl}}{\partial a} \right)}{\pi_{ijl}^2} = -\theta_k \left[\mu_{3;n} + \frac{1}{n-3} (\mu_{1;n} + \mu_{2;n} + \mu_{3;n}) \right]. \quad (\text{A.12})$$

Hence, from (A.8), (A.9), and (A.12),

$$\begin{aligned} \lim_{\theta_l \rightarrow \theta_k} \pi_{ijkl}^{\text{Hen}} &= \frac{\mu_{3;n} + \frac{1}{n-3} (\mu_{1;n} + \mu_{2;n} + \mu_{3;n})}{\left(\frac{n}{n-2} \right) \mu_{1;n}} \\ &= \left(\frac{n-2}{n} \right) \left(\frac{1}{n-3} \right) \left(1 + \frac{\mu_{2;n} + \mu_{3;n}}{\mu_{1;n}} \right) + \left(\frac{n-1}{n} \right) \frac{\mu_{3;n}}{\mu_{1;n}}. \end{aligned}$$

Note that all the limits are essentially iterated limits for the case that all the horses have equal mean running times:

$$\lim_{\theta_l \rightarrow \theta_j} \lim_{\substack{\theta_s \rightarrow \theta_j \\ (s \neq l, j)}}.$$

It is easy to show that these iterated limits are also equal to

$$\lim_{\substack{\theta_s \rightarrow \theta_j \\ (s \neq l, j)}} \lim_{\theta_l \rightarrow \theta_j}.$$

DOWNLOAD FULL VERSION HERE: <https://bettingbooks.org/12>

This page intentionally left blank

Download More Sports Betting Books at: www.bettingbooks.org

CHAPTER 5

Modeling Distance Preference and Pace Character in Thoroughbred Turf Racing

David Edelman

University College, Dublin, Ireland.

1. Background	68
2. Case Study 1: Sha Tin (Hong Kong, SAR, PRC)	69
3. Case Study 2: Randwick (Sydney, Australia)	71
4. Qualitative Questions	77
4.1. <i>Do Distance Specialists Exist?</i>	77
4.2. <i>Pace, Class, and Time: The Central Paradox of Racing</i>	78
4.3. <i>Jockeys: Distance or Pace Preference?</i>	79
5. Discussion	80
References	80

Acknowledgment

The author would like to thank Donald Hausch for his helpful comments and suggestions.

Abstract

A number of questions are raised about the modeling of distance preference and pace character of individual horses in the context of horse racing, particularly races run on a turf surface. We investigate some of these via two case studies of racing on turf, the first in Hong Kong, and the second in Sydney, Australia. The results obtained tend to confirm certain maxims of folk wisdom in racing such as the “horses for courses” principle that different horses tend to have different favored race distances, and that higher-class races tend to be run more sensibly with regard to pace and energy expenditure. The results and methods presented are of interest in and of themselves in the understanding of horse racing, and also serve as examples of how such questions may be explored methodologically. Discussion and speculation about related questions not specifically included in the case study are included here as well, pointing to directions for further enquiry.

1. BACKGROUND

While it may be clear that the average speed a horse (or other animal, including humans) is capable of running will decrease the longer the distance traveled, it is not immediately obvious how this relationship varies from one horse (or animal, or individual) to another, nor is it immediately clear whether running should be even or varied in order to obtain the optimal result.

With regard to the latter question, among a body of work on physiology of exercise and sport, arguably the most widely accepted basic model of energy expenditure in running (human and animal) is that proposed by Hill and Keller (see Keller, 1973; see also Noble, 1986 for a wider treatment) which implies that in order to optimize the overall average speed over a distance (and thus complete it in the shortest possible time), pace should be planned so as to enable as even a proportion of energy expenditure as possible over the distance to be covered, resulting in an exponential decline in velocity over time (and distance) during the course of the race.

Surprisingly, however, in the highest classes of horse racing, particularly on the turf surface, the speed (or more technically, instantaneous velocity) over a race distance tends to deviate significantly from this, as horses tend to run more slowly early on to conserve energy, and accelerate sharply in the final 400 m (roughly 1/4 mi) or so. This would appear to constitute a paradox, one which will be discussed later in this chapter, but fortunately for modeling purposes, while such pace character appears to be suboptimal, it seems from exploratory evidence that jockeys and trainers tend to run their horses in a somewhat consistent fashion (with regard to pace) from race to race. This allows us, as a first pass, to address the first of our topics, namely distance preference, independently of the second, which is pace character. In fact, we shall really only study distance preference in detail, while raising questions, and suggesting hypotheses and further research, in the area of pace. As the most marked instances of the

phenomena described above appear to be in turf racing, as opposed to dirt racing (which has a different profile with regard to energy expenditure), we focus on the former.

2. CASE STUDY 1: SHA TIN (HONG KONG, SAR, PRC)

We begin with a study of pace in a relatively small number of races in which some sectional time information is available, taking place at Sha Tin Racecourse in Hong Kong, during the period of August 2003 through July 2004, where we have limited the study to a single track in order to control for as many external factors as possible. For the same reason, we limit our study to local or noninternational races (i.e., with group and listed races excluded), and the class range limited from one (highest) to five (second-lowest, apart from races consisting of Griffin, or previously unraced horses), and races where the track condition was not slow or heavy. During this period, there remain 405 relevant races, clearly not a large enough sample to make any definitive conclusions, but larger than a mere exploratory study.

In order to study pace under maximum exertion, we focus on the winning horses in the respective races, where the time of that particular horse to the final 400 m mark (as opposed to the leading runner, as is often reported) and the time of finish are used. While it would obviously be desirable to use as much information as possible, a study of discouragement effects (the possibility that some runners may not exert when it is clear they will not win), will be left to a future study and the dataset here is restricted as a precaution.

Next, we assume that the three main determinants of the overall average speed of which horses are capable over a race are (i) the distance traveled, (ii) the track conditions, and (iii) the class of race. We conjecture, and propose to test, that (controlling for the aforementioned three factors) races that are run too slowly in the early stages result in significantly reduced overall average speed.

Arguably, the most difficult to model of the factors mentioned above is the track condition. We avoid this entirely by centering all data by meeting, which removes any systematic effects of track condition, but avoids having to rely on the accuracy and precision of official (going) ratings (i.e., “good to yielding,” etc.). Class may be incorporated as four indicator variables (for five classes), where Class 3 has no variable, and as in previous studies (and in the study to follow) distance is seen to be well-modeled on the logarithmic scale. (As the track is large and the final 400 m entirely straight, curve effects may be neglected, as may be seen by [near-linear] graphs of velocity and log-distance for a similar track analyzed in the second case study. Also, since runners that run extremely wide rarely win, and we only analyze winners here, we can be reasonably confident that the distance run is not too far from that measured for the present study.)

Before attempting to include a pace variable, we note that average velocity of the winner over a race (in km/hr) for this sample is 60.38. The results of a regression, centered by meeting, are presented in Table 1. The going/meeting, class, and distance variables described above are seen to explain 92% of the observed variation of velocity,

TABLE 1 Regression Results, Centered by Meeting

Variable	Coefficient	S.E.	T	Correlation with model
Int	0.000	0.019	0.000	—
log d	−6.816	0.107	−63.932	−0.974
cl1?	0.439	0.078	5.667	0.099
cl2?	0.180	0.060	3.007	0.052
cl4?	−0.290	0.048	−5.985	−0.029
cl5?	−0.610	0.066	−9.203	−0.182

NOTE: R^2 : 0.915; s_{res} : 0.376.

TABLE 2 Second Regression Results, Centered by Meeting

Variable	Coefficient	S.E.	T	Correlation with model
Int	0.000	0.018	0.000	—
log d	−6.675	0.111	−60.038	−0.972
cl1?	0.470	0.077	6.134	0.099
cl2?	0.199	0.059	3.370	0.052
cl4?	−0.300	0.048	−6.297	−0.029
cl5?	−0.661	0.067	−9.932	−0.182
Pace	−0.043	0.011	−3.821	−0.309

NOTE: R^2 : 0.918; s_{res} : 0.369.

resulting in a residual error of approximately 0.38 km/h (down from 1.38, the raw standard deviation), with all effects being highly significant.

As a proxy for pace, the average velocity is subtracted to the final 400 m mark from the average velocity in the final 400 m section. When the additional pace variable is added to the model, the resulting coefficient is negative (as had been anticipated), with a value of −0.043 and a T -ratio of −3.82, highly significant on 356 degrees of freedom.

The results of the regression (with variables centered by meeting) are presented in Table 2. The pace variable used here (which is actually just a difference in average velocities—late minus early) has a mean of 0.81 km/h with a standard deviation of 2.17 km/h (significantly positive, incidentally at much higher than the 1 in 1,000 level). Here, it appears that for every km/h of difference, the overall average velocity effect is roughly 0.043 km/h overall, corresponding to the equivalent of nearly 1 m in finishing length over 1,600 m (a metric mile), enough to determine the difference between winning and losing in a significant proportion of races.

Interestingly, it appears that the pace variable is 23% negatively correlated with class (one to five), suggesting that lower-class races seem to be more likely to be run at

a relatively slow early pace (accelerating afterwards), a tendency which (given the regression results of this section) could well result in slower overall runtimes. (This begs the question as to why more lower-class runners do not merely choose a quicker early pace to improve their chances of winning.)

The dataset examined here is not extensive enough to extend the above analysis to the more general relationship between pace and distance. Instead, then next section will explore how individual horses respond to distance.

3. CASE STUDY 2: RANDWICK (SYDNEY, AUSTRALIA)

In order to investigate the variation in distance preference between horses with regard to the time required to finish a race, we study a dataset from one of the most significant venues in Australian racing, namely Royal Randwick Racecourse in the Kensington suburb of Sydney, and look at data collected during 1995–1998. This dataset has been used by the author more in a forecasting context in Edelman (2005), but as the emphasis here is on relating the mathematics to the physical system of racing, a more transparent analysis will be presented.

Just as the use of a single racecourse is used to eliminate variation across racing venues, analysis is restricted to races with track conditions corresponding to days on which there was no significant rainfall. Also, in order to eliminate the lowest classes of races, only data from Saturday race meetings are included. As a further restriction, the discouragement effect (whereby jockeys on horses not in contention in the final stages of a race do not run them to their full potential) suggests that only the first half of the runners across the finish line in a given race be included here. As a final restriction, only horses with at least three runs in the sample will be included, resulting in some 495 horses and 2,655 runs. This is done since the most general model considered here will effectively include separate subregressions on each of the horses, with three observations or more being required for each having non-zero degrees of freedom.

Having carefully selected the data, a number of important modeling choices must be made. Arguably chief among these is a suitable numeraire or measurement of the observed outcomes. While the order of finish (either raw or normalized) is used in many contexts in racing, for the present study, the controlling of track conditions facilitates the use of overall runtime, or (equivalently) the average velocity over the distance for each race, which is what shall constitute the numeraire here.

It is interesting to note that the average velocity of racehorses at distances from 1,000 m (approximately 5/8 mi) to 3,200 m (approximately 2 mi) varies around a figure of approximately 60 km/h (roughly 40 mph), with a trend which (corresponding to intuition) generally decreases, on average, over this distance range, as can be seen in Figure 1. It appears that the trend is approximately log-linear, which, on reflection, might not be too different from a reasonable first guess (i.e., that relative increase in distance is more significant than absolute increase).

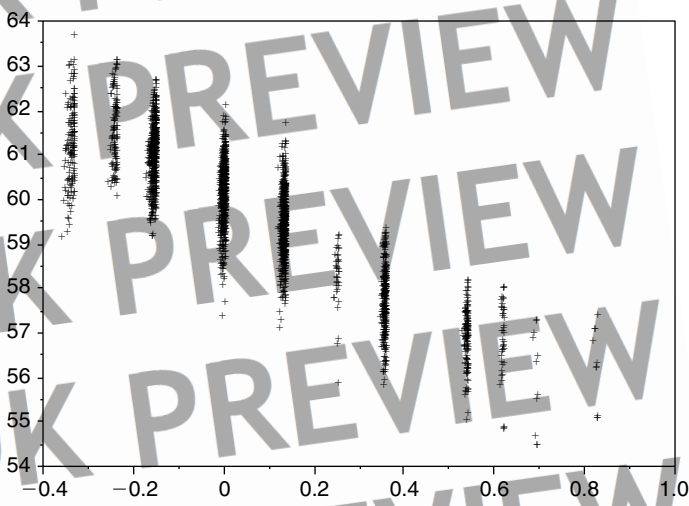


FIGURE 1 Velocity vs. log-distance.

TABLE 3 Regression of Velocity vs. Distance and Carried Weight

Coefficient	Value	S.E.	T-ratio
const	60.058	0.016	3,781.449
log-d	-5.583	0.061	-91.334
wt	0.024	0.007	3.634

NOTE: d.f. = 2,652; $R^2 = 0.759$; $s = 0.754$; $RSS = 1,505.895$.

One other control must be considered, which is the weight carried by each horse. Generally speaking, if all else is held constant, an increase in carried weight will result in a slower average velocity. Thus, a regression is run with velocity (in km/h) as a function of log-distance (in relation to 1,400 m) and weight (differenced from 55 kg). The results are summarized in Table 3.

The resulting fit is highly significant, with a prediction of approximately 60 km/h, on average, for a horse at 1,400 m carrying a weight of 55 kg. For fixed carried weight, this speed degrades by approximately 5 km/h for every 100% increase in distance traveled, or (more usefully) 0.5 km/h for every 10% increase in distance.

Paradoxically, the model appears to suggest that average velocity (over a given distance) increases with carried weight, a non-physical relationship that is explained by the fact that the better horses are handicapped by being required to carry higher weights, but tend to run faster anyway (suggesting that the handicappers do not appear to be penalizing them enough, if their objective is to even out the race).

This may be seen more clearly by including a different intercept or ability variable for each horse (where most of the horses are each running a number of races in the dataset).

TABLE 4 Regression of Velocity vs. Distance, Weight, and Ability

Coefficient	Value	S.E.	T-ratio
const	0.000	1.569	0.000
log-d	-5.130	0.050	-91.717
wt	-0.032	0.006	-5.189
ablty	1.000	0.026	34.538

NOTE: Matrix xsp : 2,655 rows, 495 columns. In each row, all zeros, except for a "1" in column i corresponding to horse i (i.e., a sparse matrix). d.f. = 2,158; $R^2 = 0.845$; $s = 0.671$; RSS = 969.791.

In this case, the weight effect is estimated for a given level of horse ability, resulting in a coefficient of the correct sign. The regression is carried out using sparse matrices (i.e., "xsp" in Table 4), and then with the 494 horse ability intercept values combined into a single input variable (the third variable in Table 4) for diagnostic purposes only. The results are shown in Table 4.

The results in Table 4 show an improvement in R^2 from 75% in the earlier model, to approximately 85% in the model with individual horse ability intercept included. The F -test may be used to test the significance of the ability intercept variable, which has 493 and 2,155 degrees of freedom. The resulting F -statistic is

$$F_{\text{calc}} = \frac{(1,506 - 970)/493}{970/2,155} = 2.42$$

Under the null hypothesis of no ability intercept (heterogeneity) effect, this statistic has a mean of approximately 1.0 and a standard deviation of approximately

$$\sqrt{\frac{2}{493}} = 0.064$$

and (for large denominator degrees of freedom) is well-approximated by the normal distribution. Since 2.42 represents many (more than 20) standard deviations above the mean, the individual ability effect is highly significant.

Thus far, the inclusion of individual ability intercept variable has only captured an overall effect (i.e., not varying with distance), whereas it would be a widely held truth in racing that horses respond differently to changes in distance. It is therefore useful to introduce individual distance gradient parameters, and to see if they significantly improve prediction accuracy. Thus, slope and intercept for each horse are fit (combined into the third variable in Table 5, where the second variable contains the combined individual ability intercept variables) resulting in 494 additional parameters.

The F -ratio for the test of whether the addition of individual distance gradient parameters significantly added to the fit may be carried out by comparing the sum of

TABLE 5 Weight, Ability, and Gradient Model

Coefficient	Value	S.E.	T-ratio
wt	−0.036	0.005	−6.168
abity	1.000	0.000	4428.895
dgrad	1.000	0.004	195.229

NOTE: d.f. = 1,666; $R^2 = 0.889$; $s = 0.646$; $RSS = 694,013$.

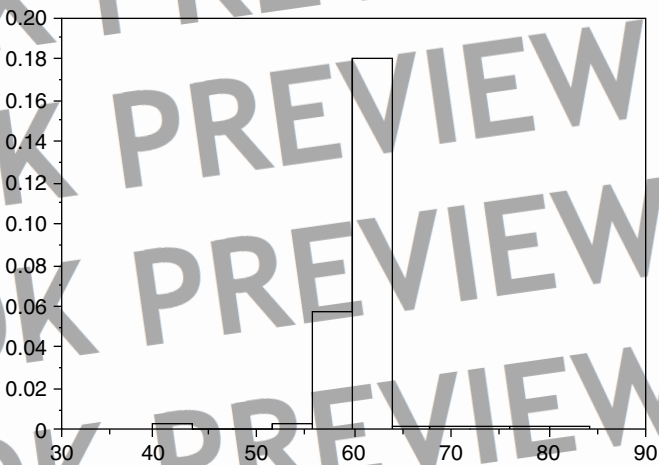


FIGURE 2 Histogram of fitted individual ability parameters.

squares of this fit to the previous:

$$F = \frac{(970 - 694)/493}{694/1,664} \doteq 1.34$$

Under the null hypothesis, this should represent a typical value sampled from a normal distribution with mean of 1 and standard deviation of approximately

$$\sqrt{\frac{2}{493}} \doteq 0.064$$

Since the observed value of 1.34 is more than five standard deviations above the mean, one would be forced to reject the null hypothesis at any reasonable level of significance.

Figure 2 contains a histogram of fitted ability intercept parameters for the various horses in the sample, showing a variation that, because of the relatively low number of races per horse, is arguably higher than any underlying inherent heterogeneity that might be expected, keeping in mind that the units are km/hr.

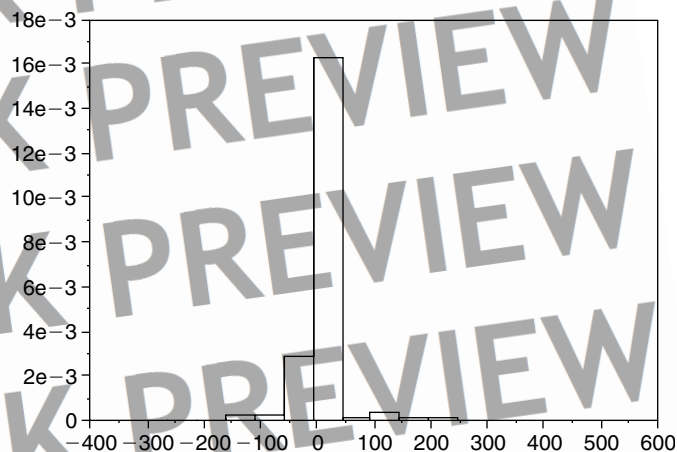


FIGURE 3 Histogram of fitted individual gradient parameters.

TABLE 6 Toward Reasonable Values

Coefficient	Value	S.E.	T-ratio
const	59.966	0.012	3,857.936
log-d	-5.039	0.047	-85.595
wt	-0.034	0.005	-5.278
ablt	1.067	0.026	32.842
dgrad	1.714	0.069	19.664

NOTE: d.f. = 1,664; $R^2 = 0.869$; $s = 0.700$; RSS = 817.001.

Similarly, in Figure 3 where the fitted distance gradient parameters for the various horses are shown, the observed variation is much more than would be expected inherently.

While these distributions of the aforementioned heterogeneity parameters would clearly not suggest strong predictive performance of the estimates, as a tool for inference (to determine significance) they have proved helpful. In Edelman (2005) a more delicate analysis similarly tests whether adding individual distance convexity parameters to the ability intercept and distance gradient parameters, as suggested by Benter (Benter et al., 1996) yields significant improvement in fit, with a negative result, suggesting the number of variables in the model presented here to be adequate.

In order to address the issue of excess heterogeneity, the technique of ridge regression (a version of empirical Bayes modeling) may be employed. This merely involves a small penalty for the sum of squares of fitted heterogeneity parameters, resulting in little deterioration in in-sample fit, and greatly improved forecasting performance.

The resulting fitting and diagnostic regression are shown in Table 6, and the parameter values are summarized in Figures 4 and 5.

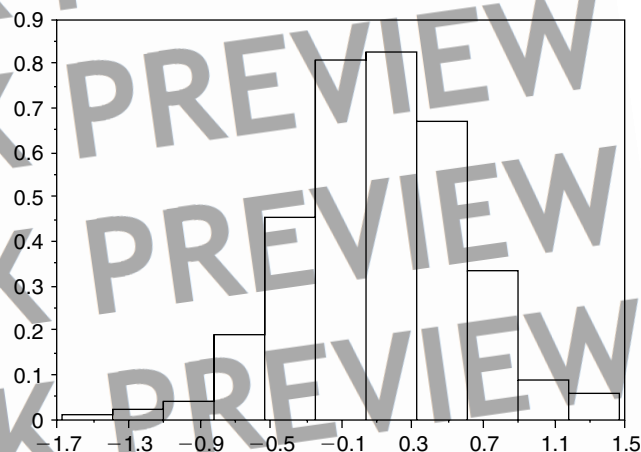


FIGURE 4 Histogram of ridge-fitted individual ability parameters.

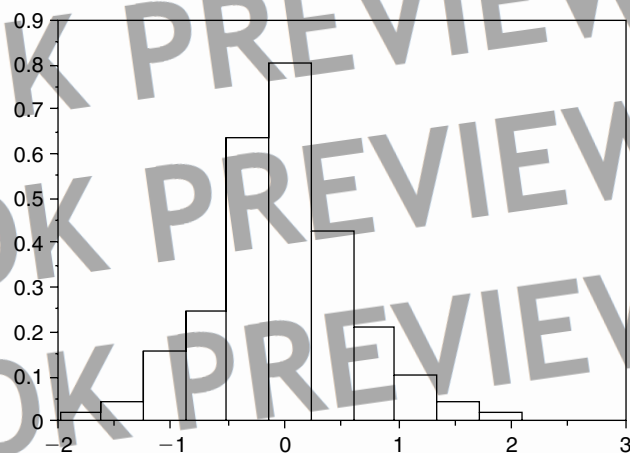


FIGURE 5 Histogram of ridge-fitted individual gradient parameters.

As can be seen, the resulting parameter distributions for the various horses are much more reasonable (i.e., the range of 10 km/hr in the former does not match intuition). Further, Figure 6 shows that the quality of the fitted values do not appear to be adversely affected.

Thus, to summarize the qualitative and practical lessons learned from this case study, it appears that individual horses not only have significantly different overall abilities, but overall degradation of average speed as the distances over which they run are increased as well. It is possible to model this variation reasonably using ridge regression, which (as mentioned earlier) yields not only distributions of heterogeneity, which seem intuitively reasonable, but resulting fitted values of velocity which appear to be effective and seem appropriate.

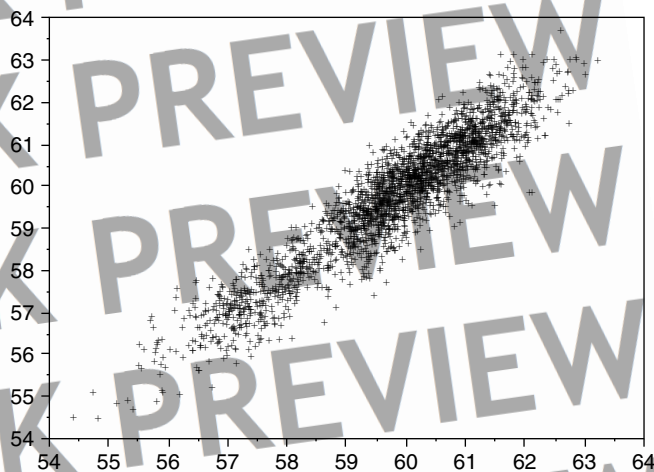


FIGURE 6 Scatterplot of fitted vs. actual velocities.

4. QUALITATIVE QUESTIONS

It is hoped that the case studies presented in the previous sections may have shed some light on the general modeling of pace character and distance preference in racing, but it is of interest to explore any qualitative ramifications of the model developed there, and to raise several other related issues that may require further work of a similar nature to unravel.

4.1. Do Distance Specialists Exist?

First, the identification of the distance preference/gradient model in the case study of the previous section might on first reflection appear to suggest that a given horse's chances relative to the others will either increase or decrease monotonically with distance, and this will certainly be the case when comparing two horses. However, as will be argued below, if a horse's distance suitability is compared to two or more other horses, a different picture emerges. To see this, a graphical representation is helpful. Figure 7 graphs the distance preference residual models of three horses, where the slopes are equal to the difference between the distance gradients of each individual horse and a particular average or median distance gradient for the entire population of horses. In Figure 7 there is one horse whose distance gradient is lower than average in magnitude, and that therefore is relatively advantaged with respect to the overall average (log-linearly) as distance increases. Then there is a horse with an average distance gradient that is neither advantaged nor disadvantaged relative to the average as distance increases, and a third horse with a distance gradient of greater magnitude than the average that is most advantaged the shorter the distance and disadvantaged (log-linearly) the longer.



FIGURE 7 Distance preference—hypothetical example.

As can be seen from the graph in Figure 7, the more average distance gradient horse will be advantaged relative to the best of the other two at moderate ranges. Thus a distance specialist may result even though the distance preference model derived in the previous section is monotonic for each horse.

4.2. Pace, Class, and Time: The Central Paradox of Racing

It was mentioned earlier that the most widely accepted physiological models imply that humans and animals wishing to run a distance in the shortest possible time should attempt to run so that the proportion of energy expended in each instant relative to the total amount expended previously is roughly constant, which will imply an instantaneous velocity that decreases exponentially over the entire distance traveled. This does not always match the manner in which horses are run in races. In fact, it is a not uncommon occurrence that the total runtime for a top-class race is slower than that of a lower-class race run by much lower-class horses over the same distance and conditions. It is an observed fact that many races (even some higher-class races) are won by horses accelerating markedly in the final stages (say 400 m or 1/4 mi) in a manner which (from the analysis) appears to be suboptimal with regard to overall runtime. This raises the question of why (theoretically) a low-class horse couldn't run an evenly paced race and win a high-class (or for that manner, any) event where the early speed was slow, of which there appear to be frequent examples. In the opinion of the author of this article, this question constitutes a paradox so basic that it might be deserving of the status of the *Central Paradox of Racing*.

One potential explanation might be demonstrated by the special case of a two-horse race. Suppose that each of the jockeys riding the horses believes that, provided that energy expenditure has not been excessive to the last call position (final 400 m mark), his or her horse is capable of outrunning the other horse by at least 2 lengths (6 m) during this final stage. Then there is no incentive for either horse to run quickly during

the early stages, provided (from the point of view of each horse) neither horse overexerts or falls farther behind than an amount slightly less than 2 lengths.

This argument highlights the fact that the objective sought by jockeys when deciding tactics is *winning*, rather than achieving the quickest time. The above example of the two-horse race might fruitfully be generalized to a full field of horses, and with potentially more game-theoretic issues involved, such as each jockey's uncertainty about the capabilities of his or her horse. A particular challenge in such an analysis would be the modeling of a wide range of endowments (i.e., horse characteristics and capabilities).

It is not immediately clear how such game-theoretic models might be posited and validated empirically, nor indeed whether jockeys could be assumed to be employing optimal competitive riding strategies, but if such modeling were possible, it would constitute a very significant contribution indeed.

Another potentially relevant factor in pace and time is wind resistance. While there appear to be no empirical studies on the effect of wind resistance on speed in horse racing, it has been proposed that just as the V flight formation of bird flocks has evolved in such a way as to minimize collective wind resistance (and therefore energy expenditure), so might the herding of galloping horses be expected to have an analogous effect. If true, this might allow an alternative resolution of the *Central Paradox* identified above, for if an evenly paced horse were to get too far ahead of the herd, it could well end up expending much more energy than if it shared wind resistance with the rest of the horses by racing closer to them. Likewise, horses traveling slightly behind the herd might perhaps be expected to save energy from slip-streaming which is the effect of wind eddies from the other horses in front, and which might help to pull a hind horse along. Such an effect appears to be widely acknowledged in automobile racing (with regard to fuel consumption), but does not appear to have been studied systematically in the horse racing context.

It is not clear how these hypotheses might be tackled, but perhaps as more data becomes available on the positions of horses at various times in running, quantitative exploration of such questions might become feasible.

4.3. Jockeys: Distance or Pace Preference?

Often in racing, it is difficult to unravel the net effect of jockey ability on racing outcomes, because certain popular jockeys tend to win more races just because they are given better horses to ride.

Of course, simultaneous League Table analyses of horses and jockeys via Analysis of Variance may be used to determine the overall marginal effect of jockeys on race outcomes, but it is interesting additionally to test whether some jockeys tend to run better races on forward-running horses, as opposed to late-finishing horses. Exploratory analyses on Hong Kong data appear to suggest a rather strong effect in this regard, though (as horses tend to run with similar pace style from race to race) this is somewhat confounded with the possibility that jockeys might have different marginal effects on different individual horses. In this, as with some other topics raised here, further study along these lines is needed.

5. DISCUSSION

The two case studies presented, while limited in their scope, appear to shed light on several aspects of pace and overall runtime in turf racing. In the first study, a highly significant statistical pace proxy variable was found for predicting overall runtime of the winner of a race, namely the difference between its late (last 400 m) average velocity and its earlier (pre-400 m-mark) average velocity. The coefficients suggest that races that are run too slowly in the early stages result in slower overall times. This appears consistent with physiologists' view that, other factors being equal, an even pace should result in the best runtimes. Assuming this is true, it appears that lower-class races are more frequently run too slowly in the early stages, as these tend to have a larger than average late-minus-early velocity differential.

While the first case study did not analyze characteristics of individual horses due to limited data, the second, more extensive, dataset enabled examination of certain aspects of overall runtime characteristics for individual horses (albeit not with regard to pace). In particular, the question of whether different horses appear to respond differently to different distances was asked, and answered in the affirmative. It is hoped that data may become available that would enable the study of both pace character and distance preference for individual horses in the same study, the results of which could help greatly not only in the understanding and forecasting of runtimes and race outcomes, but potentially in informing racing connections with regard to optimal riding tactics for their horses as well.

In general terms, the intent here has been to stimulate thought on and to demonstrate methodology in the modeling of distance and pace aspects in horse racing.

In the case studies presented, some aspects of the modeling of horses that are racing at different paces and distances have been addressed, and some tentative conclusions reached. With regard to pace, or the variation in velocity of runners during the course of a race, results from physiology can motivate important questions to be asked, but more extensive data are required if studies are to be carried out to draw any clear conclusions. The lack of existing quantitative studies on such topics (apart from that presented here) does not prevent speculation on certain aspects, which can only serve to stimulate the type of curiosity that will lead to further endeavor.

References

- Benter, W. F., Miel, G. J., and Turnbough, P. D. 1996. Modeling Distance Preference in Thoroughbred Racehorses, *Proceedings of the Third Mathematics and Computers in Sport Conference (3MCS)*, Bond University.
- Edelman, D. 2005. Modeling Distance Preference in Horserace Handicapping, in L. Vaughan-Williams (ed.), *Information Efficiency in Financial and Betting*, chap. 16, pp. 339–345. Cambridge University Press, Cambridge, UK.
- Keller, J. B. 1973. A Theory of Competitive Running, *Physics Today* 26, 42–47.
- Noble, B. J. 1986. *Physiology of Exercise and Sport*. Times Mirror/Mosby College Publishing, St. Louis, MO.

DOWNLOAD FULL VERSION HERE: <https://bettingbooks.org/12>

PART III: Favorite-Longshot Bias in the Win Market

Download More Sports Betting Books at: www.bettingbooks.org

DOWNLOAD FULL VERSION HERE: <https://bettingbooks.org/12>

This page intentionally left blank

Download More Sports Betting Books at: www.bettingbooks.org

CHAPTER 6

The Favorite-Longshot Bias: An Overview of the Main Explanations

Marco Ottaviani

Kellogg School of Management, Northwestern University, Evanston, IL, USA.

Peter Norman Sørensen

Department of Economics, University of Copenhagen, Copenhagen, Denmark.

1. Introduction	84
2. Notation	87
3. Misestimation of Probabilities	87
4. Market Power by Informed Bettors	89
5. Preference for Risk	90
6. Heterogeneous Beliefs	93
7. Market Power by Uninformed Bookmaker	95
8. Limited Arbitrage	96
9. Simultaneous Betting by Insiders	97
10. Timing of Bets	99
10.1. Early Betting	99
10.2. Late Betting	99
References	100

Abstract

In betting markets, the expected return on longshot bets tends to be systematically lower than on favorite bets. This favorite-longshot bias is a widely documented empirical fact, often perceived to be an important deviation from the market efficiency hypothesis. This chapter presents an overview of the main theoretical explanations for this bias proposed in the literature.

1. INTRODUCTION

A central theme of the literature on betting is the occurrence of the favorite-longshot bias. The first documentation of this bias is attributed to Griffith (1949), who observed that horses with short odds (i.e., favorites) yield on average higher returns than horses with long odds (i.e., longshots). This means that market probabilities of longshots (obtained from market prices) overpredict on average their empirical probabilities (computed from race outcomes). At the other end of the spectrum, the market probabilities of favorites tend to underpredict their empirical probabilities. This chapter presents an overview of the wide range of theories proposed in the literature to explain this bias.

Betting markets have attracted a lot of attention by economists because they provide a particularly appealing environment for testing theories of market efficiency.¹ First, the outcomes in these markets are publicly observed at a prespecified time. In regular financial markets, the uncertainty about asset values is typically resolved only in the long run, if ever. Second, in the case of betting, the realized outcomes are exogenous to the trading process and the resulting prices. In comparison, the intrinsic value can easily be affected by market prices in more traditional financial settings. Third, pari-mutuel betting markets are particularly suited to testing for market efficiency because prices there are not set (and so potentially misaligned) by individual market makers. By placing a pari-mutuel bet on an outcome, a participant demands a share of all the funds supplied by the other participants conditional on the realization of that outcome and, contemporaneously, supplies funds to all the other participants if that outcome is not realized.

Indeed, most of the empirical literature has focused on pari-mutuel markets, in which the money bet on all outcomes is pooled and then shared proportionally among those who picked the winning outcome, after a fractional sum is deducted for taxes and expenses. If the market were efficient, all the bettors were risk neutral, and they shared the same belief about the outcome, the final distribution of pari-mutuel bets should be directly proportional to the market's assessment of the horses' chances of winning. This is because the gross expected payoff of a bet on an outcome is equal to the ratio of the outcome's probability to the proportion of bets placed on that outcome. If the fraction

¹A clear drawback of betting markets is that traders may be motivated by recreational objectives that we expect to play less of a role in regular financial markets.

of money bet on each outcome is equal to its probability, the expected payoffs would then be equalized across all outcomes.

The bias is often perceived to be an important deviation from the market efficiency hypothesis. A voluminous empirical literature (surveyed, among others, by Thaler and Ziemba, 1988; Hausch and Ziemba, 1995; Sauer, 1998; and Jullien and Salanié, 2008) has documented the extent of this bias across different events, countries, and market structures. Despite the presence of a fair amount of variation in the extent and sometimes the direction of the bias, the favorite-longshot bias has emerged as an empirical regularity. While the initial literature focused on pari-mutuel markets, the favorite-longshot bias is also observed (often to a greater extent) in markets dominated by bookmakers.

In a short paper that appears to have gone completely unnoticed, Borel (1938) presented the first theoretical analysis of optimal betting behavior and equilibrium in the pari-mutuel game.² This remarkable paper foretells a number of the essential elements of the theories developed later in the literature.

- First, Borel introduced the problem and defined the equilibrium in the context of betting on the sum obtained rolling two dice, for which players naturally share common (or objective) probability assessments: “This game will be equitable, if the total amount bet on each point is proportional to the probability of obtaining that point; but there appears to be no a priori reason for this condition to be realized on its own.”³ He then discussed informally the forces that bring the system to equilibrium in the context of this game.
- Second, Borel described how pari-mutuel odds adjust against the bettor and then determined the optimal amount a bettor should place on one of two outcomes to maximize the expected return, given an initial distribution of bets.⁴ He then considered the case of a sequence of players who make optimal bets, after observing the amounts placed in the past. He noted that the amounts bet (in the subgame-perfect equilibrium) make the game asymptotically equitable.
- Third, Borel modeled a pari-mutuel market with two classes of strategic bettors with heterogeneous subjective probability beliefs about the outcome of a race between two horses. In each class, there are two bettors who share the same probability belief about the race outcome, but beliefs are different for bettors belonging to the two classes. Borel characterized the (Nash) equilibrium of this game and concluded with an informal discussion of the timing incentives.⁵

Over the last seven decades, a number of theories have been advanced to explain the favorite-longshot bias. In this chapter we review the main theoretical explanations for the favorite-longshot bias proposed thus far in the literature, in order of their chronological development:

²We searched extensively for references to this article, but did not find any.

³This equilibrium notion is the main benchmark with respect to which the favorite-longshot bias is defined in the literature.

⁴This problem is further explored by Borel (1950) and later generalized by Isaacs (1953). See Section 4.

⁵As explained below in Section 10.1, pari-mutuel betting is a version of Cournot’s quantity competition game.

1. *Misestimation of Probabilities (Section 3)* The bias can be due to the tendency of individual decision makers to overestimate small probability events. This explanation was initially advanced by Griffith (1949), who suggested that there is a psychological bias that leads individuals to subjectively ascribe excessively large probabilities to rare events.
2. *Market Power by Informed Bettors (Section 4)* A monopolist bettor who bets large amounts should not equate the expected return on the marginal bet to zero, since this would destroy the return on the inframarginal bets. If this large bettor has unbiased beliefs and bets optimally on the favorite, the favorite-longshot bias results. This explanation follows from the analysis of Isaacs (1953).
3. *Preference for Risk (Section 5)* If individual bettors love risk or skewness, they are willing to accept a lower expected payoff when betting on longshots. This explanation was articulated by Weitzman (1965), for the case of a representative bettor who loves risk, and so is willing to give up a larger expected payoff when assuming a greater risk on a longshot with longer odds.
4. *Heterogeneous Beliefs (Section 6)* If bettors have heterogeneous beliefs, the market probabilities resulting in the pari-mutuel system tend to be less extreme than the bettors' median belief. This theory, formulated by Ali (1977), can explain the favorite-longshot bias if one is prepared to assume that the bettors' median belief is equal to the empirical probability.
5. *Market Power by Uninformed Bookmakers (Section 7)* For fixed odds betting markets, Shin (1991 and 1992) explained the favorite-longshot bias as the response of an uninformed bookmaker to the private information possessed by insiders.
6. *Limited Arbitrage by Informed Bettors (Section 8)* The favorite-longshot bias results when price-taking (and risk-neutral) bettors possess superior information, since the amount of arbitrage is limited by the presence of the track take and the inability to place negative bets. This explanation was proposed by Hurley and McDonough (1995).
7. *Simultaneous Betting by Partially Informed Insiders (Section 9)* In pari-mutuel markets, the bias arises if privately informed bettors place last-minute bets without knowing the final distribution of other bettors' bets. This explanation is due to Ottaviani and Sørensen (2006), who derived the bias in a pari-mutuel market as the result of bets placed simultaneously.

We begin by introducing the notation in Section 2 and then proceed to develop the explanations. Given that these explanations have similar qualitative implications about the favorite-longshot bias, it has proven difficult for empiricists to distinguish between the alternative theories. To reach some tentative conclusions about the relative merits of the different explanations, it is promising to test for the ability of theories to simultaneously explain the favorite-longshot bias as well as other regularities regarding the dynamic adjustment in market prices and the timing of bets. While a comparison of the performance of the different explanations is well beyond the scope of this chapter, we

conclude in Section 10 by giving an indication of some additional theoretical predictions regarding the timing of bets.

2. NOTATION

Given our aim at presenting the explanations in the simplest possible setting, we focus on the case with two outcomes. We denote the outcome (corresponding to the winning horse) by $x \in \mathcal{X} = \{1, 2\}$. We denote bettors by $n \in \mathcal{N} = \{0, 1, \dots, N\}$, where $n = 0$ represents all outsiders and bettors with index $n \geq 1$ are strategic insiders. Bettors may have different prior beliefs. For convenience, we denote bettor n 's prior belief for outcome 1 by $q_n = \Pr(x = 1)$. When allowing for private information, bettor n 's signal is $s_n \in \mathcal{S}$, resulting in posterior belief $r_n = \Pr(x = 1 | s = s_n)$.

The amount placed by bettor n on horse x is $b_n(x)$, and the total amount placed on x is $b(x) = \sum_{n \in \mathcal{N}} b_n(x)$. The total bets placed on x by the opponents of bettor n are $b_{-n}(x) = b(x) - b_n(x)$. The total overall bets are $B = \sum_{x \in \mathcal{X}} b(x)$ (also known as the pool), while total bets placed by the opponents of bettor n are $B_{-n} = \sum_{x \in \mathcal{X}} b_{-n}(x)$.⁶ The track take, τ , is a percentage subtracted from the pool for taxes and expenses.⁷

The pari-mutuel odds are then $\rho(x) = [(1 - \tau) B - b(x)] / b(x)$, so that every bet on horse x wins $1 + \rho(x) = [(1 - \tau) B] / b(x)$ if x is realized. The pari-mutuel system results in the market probability $m(x) = (1 - \tau) / [1 + \rho(x)] = b(x) / B$, with $m = m(1)$ for convenience. We denote the objective (or empirical) probability by $p(x)$, with $p = p(1)$.

3. MISESTIMATION OF PROBABILITIES

When Griffith (1949) uncovered the favorite-longshot bias, he referred informally to a simple psychological explanation based on biases in the market participants' assessment of the probability attached to the different outcomes. For the purpose of our illustration with two outcomes, assume that all bettors attribute a perceived probability equal to $\pi(p)$ to an outcome with objective probability p . The key assumption of the theory is that bettors overestimate the chance of unlikely outcomes but underestimate the chance of likely outcomes: $1/2 > \pi(p) > p$ for $p < 1/2$ and $1/2 < \pi(p) < p$ for $p > 1/2$.

Let p be the objective probability of horse 1 and $\pi(p)$ the perceived probability. The perceived expected net monetary payment (or payoff) of a bet on horse 1 is

$$\pi(p) \left(-1 + \frac{1}{m} \right) + [1 - \pi(p)] (-1) = \frac{\pi(p)}{m} - 1. \quad (1)$$

⁶Note that b_{-0} is different from b_0 , and similarly B_{-0} is different from B_0 .

⁷The presence of a positive track take plays a key role in the explanation proposed in Section 8. To simplify our derivations, we set $\tau = 0$ when presenting most of the other explanations.

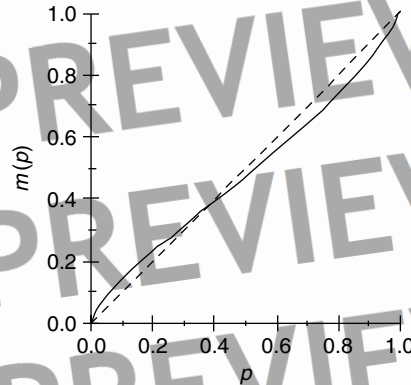


FIGURE 1 Market probability against objective probability for $a = 0.8$. The dashed line is the diagonal.

Similarly, the perceived expected payoff of a bet on horse 2 is

$$\pi(1-p) \left(-1 + \frac{1}{1-m} \right) + [1 - \pi(1-p)](-1) = \frac{\pi(1-p)}{1-m} - 1. \quad (2)$$

In equilibrium, the perceived expected payoffs of the two bets must be equal,

$$\frac{\pi(p)}{m} = \frac{\pi(1-p)}{1-m}.$$

Suppose that $p > 1/2 > 1-p$. Then $p > \pi(p) > 1/2 > \pi(1-p) > 1-p$ by the key assumption of the theory. Hence,

$$\frac{p}{1-p} > \frac{m}{1-m} = \frac{\pi(p)}{\pi(1-p)} > 1.$$

We can conclude that the favorite-longshot bias arises as stated in following proposition.

Proposition 1 *If $\pi(p) < p$ for $p > 1/2$ and $\pi(p) > p$ for $p < 1/2$, the market probability of the favorite (respectively longshot) is lower (respectively higher) than its objective probability: if $p > 1/2$, then $1/2 < m < p$ (respectively if $p < 1/2$, then $p < m < 1/2$).*

We illustrate this explanation by using Prelec's (1998) weighting function with $a < 1$,

$$m = \pi(p) = \exp \left[-(-\ln p)^a \right],$$

where m is the market probability and p is the objective probability.⁸ Figure 1 plots $\pi(p)$ for $a = 0.8$.

⁸As is evident from Figure 1, Prelec's function has $\pi(\hat{p}) = \hat{p}$ for a threshold $\hat{p} = 1/e$ somewhat below $1/2$. The spirit of the favorite-longshot bias is preserved, nevertheless.

4. MARKET POWER BY INFORMED BETTORS

Borel (1938 and 1950) and Isaacs (1953) proposed a theory compatible with an explanation for the favorite-longshot bias based on optimal betting behavior by a large privately informed bettor. This explanation is based on the fact that the more money this bettor places on a horse, the lower the odds that result in the pari-mutuel system. Then, an informed bettor would want to limit the amount bet in order to maximize the profits made.

To understand why additional bets on a horse depress the horse's odds, imagine a single strategic bettor (the insider, $N = 1$) who estimates that horse 1 is more likely to win than according to the prevailing market odds (set by outsiders). Suppose that the outsiders are not placing any money on 1, while they are placing some money on the competing horse 2. By betting just one dollar, when horse 1 wins our bettor can be sure to obtain all the money bet by the outsiders on the other horses. Hence, the first dollar has a higher expected return than the second dollar, which has a zero marginal return. Since our bettor loses the dollar when horse 1 does not win the race, in this extreme example our bettor does not want to bet more than one dollar on horse 1.

More generally, given the pari-mutuel payoff structure, the payout per dollar bet on horse 1 is decreasing in the fraction of money that is bet on horse 1. Because the insider takes into account the payoff on all of his or her bets, it is optimal to cease betting before the payout on the marginal bet equates the marginal cost. This implies a bias, that our bettor does not bet until the market probability equals his or her posterior belief.

From the mathematical point of view, the bettor's problem is the same as the problem of a monopolist who decides how much quantity of a product to sell in a market with a downward sloping demand curve. Essentially, the pari-mutuel market structure induces such a downward sloping demand, because the average payout decreases in the amount wagered.

We now illustrate this explanation in the simplest possible setting, keeping $N = 1$. Clearly, it is never optimal for our insider to bet on both horses.⁹ Suppose that the insider believes sufficiently more than the outsiders that horse 1 will win so that there is a positive gain from betting a little on it. Precisely, assume $q_1 > b_0(1)/[B_0(1 - \tau)]$. If bettor 1 bets the amount $b_1(1) = b$ on this horse, the price of each bet is determined on the basis of the inverse demand curve

$$P(b) = \frac{q_1}{m} = (1 - \tau)q_1 \frac{B_0 + b}{b_0(1) + b}. \quad (3)$$

The insider's objective is to maximize the expected revenue $[P(b) - 1]b$. The marginal revenue is $P(b) - 1 + P'(b)b$. Our assumption that $q_1 > b_0(1)/[B_0(1 - \tau)]$ means precisely that the marginal revenue evaluated at $b = 0$ is positive, $P(0) > 1$. The optimal positive bet size is then determined by the first order condition

$$P(b) - 1 = -P'(b)b. \quad (4)$$

⁹Isaacs (1953) shows more generally that a bettor should bet at most on all but one horse.

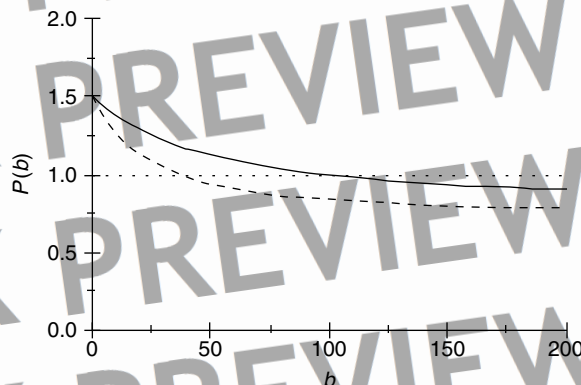


FIGURE 2 The optimal bet equalizes the marginal revenue (the dashed curve) to the unit marginal cost (the dotted line). The solid curve represents the demand.

Since the demand curve is downward sloping, the right-hand side is strictly positive, so we can conclude that the insider's optimal bet satisfies $P(b) > 1$. This is clearly equivalent to $q_1 > m$. If we additionally suppose that the insider is betting on the favorite¹⁰ and suppose that the insider's belief q_1 is correctly equal to p , we obtain the main result in the following proposition

Proposition 2 *Suppose that the insider bets on the favorite with correct beliefs. Then the favorite's market win chance is lower than the empirical chance: $p > m > 1/2$.*

Figure 2 displays the demand curve, the marginal revenue, and the marginal cost for an example with $B_0 = 100$, $b_0(1) = 50$, $\tau = 0$, $p = 3/4$. It is possible to solve the insider's optimization problem for the optimal bet $b_1(1) = 50(\sqrt{3} - 1) = 36.603$. The market probability is then $m(1) = b(1)/B = (3 - \sqrt{3})/2 = 0.633.97 < 3/4 = p$.

Chadha and Quandt (1996) extended this explanation by considering the case with multiple bettors who play a Nash equilibrium.¹¹

5. PREFERENCE FOR RISK

The third explanation for the favorite-longshot bias is based on the different variability in the payout of a bet on a longshot compared to one on a favorite. Longshots tend to pay out more, but with smaller probability. If bettors prefer riskier bets, the relative

¹⁰A sufficient condition for outcome 1 being the favorite is that the outsiders have it as their favorite, $b_0(1)/B_0 \geq 1/2$. Outcome 1 is also likely to become the favorite because the insider bets on it, so the condition is not necessary.

¹¹On the way to deriving the competitive limit with infinitely many bettors, Hurley and McDonough (1995) also characterize the Nash equilibrium resulting with a finite number of bettors.

price of longshots should be relatively higher. This explanation was first spelled out by Weitzman (1965), followed by many others.¹²

Keeping with Weitzman (1965), we assume that there are many identical bettors, all with the same beliefs and risk preferences. Without loss of generality, normalize the mass of bettors to $N = 1$. For simplicity, we further assume with Quandt (1986) that bettors have mean-variance preferences, $u = E - \rho V$, with coefficient of risk aversion equal to ρ . Each bettor is small relative to the size of the market, and so takes the market prices as given when deciding which of the two horses to back. Set $\tau = 0$.

If horse 1 is the market favorite and attracts a fraction $m = m_F = 1 - m_L > 1/2$ of the pool of pari-mutuel bets, it yields the expected net monetary payment

$$E_F = \frac{p}{m_F} - 1. \quad (5)$$

Similarly, the expected net monetary payment on the longshot is

$$E_L = \frac{1-p}{m_L} - 1. \quad (6)$$

The variance of the net monetary payment on the favorite is

$$V_F = p \left(-1 + \frac{1}{m_F} \right)^2 + (1-p)(-1)^2 - \left(\frac{p}{m_F} - 1 \right)^2 = \frac{(1-p)p}{m_F^2}. \quad (7)$$

Similarly, the variance of a bet on the longshot is

$$V_L = p(-1)^2 + (1-p) \left(-1 + \frac{1}{m_L} \right)^2 - \left(\frac{1-p}{m_L} - 1 \right)^2 = \frac{(1-p)p}{m_L^2}. \quad (8)$$

Note that $V_L > V_F$ because $m_F = 1 - m_L > 1/2$ because by definition the favorite has a higher market probability.

In equilibrium, the following indifference condition must hold $E_F - \rho V_F = E_L - \rho V_L$, or equivalently

$$-\rho = \frac{E_F - E_L}{V_L - V_F}. \quad (9)$$

Now, if the representative bettor has a preference for risk ($\rho < 0$), at equilibrium we must have

$$\frac{E_F - E_L}{V_L - V_F} > 0.$$

Given that the denominator of the left-hand side is positive ($V_L > V_F$, as shown above), this inequality implies that $E_F - E_L > 0$, that is, that $m_F < p$ or, equivalently by

¹²Golec and Tamarkin (1998) advanced the related hypothesis that bettors love skewness, while being risk averse.

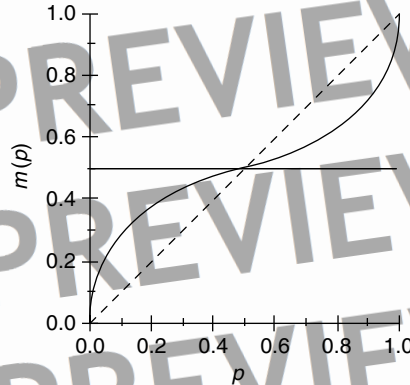


FIGURE 3 Market probability against objective probability for $\rho = -1$. The market probability is above the objective probability (represented by the dashed diagonal) for the favorite ($p > 1/2$). The pattern is reversed for the longshot ($p < 1/2$).

$m_F = 1 - m_L$, that $1 - p < m_L$. Combining these inequalities with $m_F = 1 - m_L > 1/2$, we obtain the favorite-longshot bias: $1 - p < m_L < 1/2 < m_F < p$.

Proposition 3 Suppose that there is zero track take, $\tau = 0$, and a representative bettor with mean-variance preferences and negative coefficient of risk aversion, $\rho < 0$. The market probability of the favorite (respectively longshot) is lower (respectively higher) than its objective probability: if $p > 1/2$, then $1/2 < m < p$ (respectively if $p < 1/2$, then $p < m < 1/2$).

It is easy to further characterize the equilibrium in this setting. In equilibrium, each bettor must be indifferent between betting on the favorite and betting on the longshot,

$$u_F = \frac{p}{m_F} - 1 - \rho \frac{(1-p)p}{m_F^2} = \frac{1-p}{1-m_F} - 1 - \rho \frac{(1-p)p}{(1-m_F)^2} = u_L,$$

where we have substituted Equations (5), (6), (7), and (8) and the identity $m_L = 1 - m_F$ into the indifference condition (9). This equation can be solved to obtain the equilibrium market probability as a function of the objective probability, $m(p)$. The inverse function $p(m)$ has the simpler analytic expression

$$p(m) = \frac{1}{2} + \frac{\sqrt{[\rho(2m-1) + m(1-m)]^2 - 4\rho(2m-1)m^2(1-m) - m(1-m)}}{-2\rho(2m-1)}$$

Figure 3 displays the market probability as a function of the objective probability for an example with $\rho = -1$. Quandt (1986) further generalized this example by allowing for heterogeneous risk attitudes.

6. HETEROGENEOUS BELIEFS

Following Ali (1977), suppose that bettors have heterogeneous prior beliefs. Suppose that bettors do not observe any private signal, so that they do not have superior information. Suppose that the track take is zero, $\tau = 0$, bettors are risk neutral, $u(w) = w$, have identical wealth available for betting, and have beliefs drawn from the same distribution, $q_n \sim F(\cdot)$.

It follows that each bettor bets all the available wealth on either of the two horses. The competitive equilibrium is characterized by the indifference threshold belief \bar{q} , at which the expected payoff from betting on either horse is equalized. Bettors with belief above threshold \bar{q} bet on horse 1 and bettors with belief below \bar{q} bet on horse 2.

Proposition 4 (i) *The market probability on the favorite is below the belief of the median bettor: If $m > 1/2$, then $m < p$ where $p = F^{-1}(1/2)$. (ii) If in addition the belief of the median bettor is equal to the objective probability, then the market probability that the favorite wins is lower than the objective probability.*

We focus on a horse whose market probability in the pari-mutuel system is

$$m > 1/2. \quad (10)$$

By definition, the fraction of all bets that are placed on this horse is equal to the market probability m . Note that a risk-neutral bettor optimally bets on the horse when subjectively believing that this horse is more likely to win than indicated by the market probability. So, it must be that the fraction of bettors who have subjective beliefs above m is equal to the market probability,

$$1 - F(m) = m. \quad (11)$$

By definition of the median p , half of the bettors have beliefs above it,

$$1 - F(p) = 1/2. \quad (12)$$

Combining Equations (10), (11), and (12), and the property that the belief distribution F is increasing, the market probability is below the median, $m < p$, as stated in part (i) of the proposition. Part (ii) follows immediately.

This result is a general implication of competitive equilibrium behavior, and holds independently of the pari-mutuel market structure. What drives this result is the fact that bettors are allowed to put at risk a limited amount of money. To see this, we now reinterpret the pari-mutuel equilibrium as the Walrasian equilibrium in a complete Arrow-Debreu market. Traders are allowed to buy long positions or sell short positions on the asset that pays 1 conditional on horse 1 winning. Each trader is allowed (or desires) to lose at most an amount equal to 1. This means that, when wishing to buy long the asset that is traded at price m , a trader purchases at most $1/m$ asset units—by risk neutrality, this is the exact amount purchased. At price m , the demand curve for the

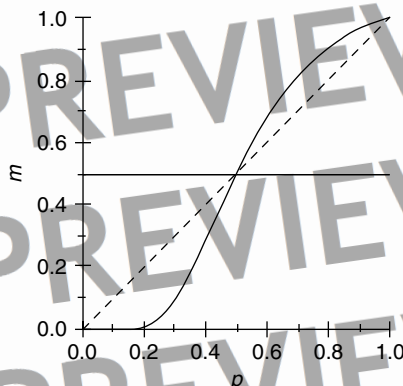


FIGURE 4 Plot of the median belief against the market probability. The dashed line is the diagonal.

asset is then $[1 - F(m)]/m$. Similarly, a trader who wishes to sell short actually sells at most $1/(1 - m)$ asset units when the asset price is m .¹³ The asset's supply curve is then $F(m)/(1 - m)$. The equilibrium price m prevailing in the market equates demand and supply

$$\frac{1 - F(m)}{m} = \frac{F(m)}{1 - m},$$

which is equivalent to (11).¹⁴

Intuitively, $m < p$ means that the median bettor (with belief p) strictly prefers to risk all his or her money on horse 1, by taking long positions. By continuity, a bettor with a belief slightly more pessimistic is also long on horse 1. If all traders were to invest the same amount on either horse, the market could not equilibrate because the demand for long positions would outstrip the supply at price m , given that $1 - F(m) > F(m)$ for a price below the median, $m < p$. In equilibrium, the number of assets bought is equal to the number of securities sold, so each trader on the long side must be allowed to buy less than each trader on the short side can sell, that is, $m > 1 - m$.

Ali's result may be illustrated by this class of belief distributions $F_\alpha(q) = q^\alpha$, parameterized by $\alpha > 0$. There is a one-to-one relationship between the parameter $\alpha > 0$ and the equilibrium market probability $0 < m < 1$ solving Equation (11), which can be expressed as $\alpha = [\log(1 - m)]/[\log(m)]$. The median of F_α solving Equation (12) is $p = 2^{-1/\alpha}$. Figure 4 plots the median belief against the market probability. Note that the favorite-longshot bias results: If $m > 1/2$ we have $1/2 < m < p$, while if $m < 1/2$ we have $1/2 < p < m$.

Blough (1994) extended this result to an arbitrary number of horses under a natural symmetry assumption.

¹³The reason for this is that this trader's income on the contracts sold is $-m/(1 - m)$, while the trader's outlays are $1/(1 - m)$ when horse 1 wins. Overall, this trader risks $1/(1 - m) - m/(1 - m) = 1$.

¹⁴Eisenberg and Gale (1959) considered the properties of the pari-mutuel equilibrium price as an aggregation device for the heterogeneous beliefs.

7. MARKET POWER BY UNINFORMED BOOKMAKER

We now turn to Shin's (1991 and 1992) explanation, based on the response of an uninformed bookmaker to the private information possessed by insiders. Shin mostly focused on the case of a monopolist bookmaker who sets odds in order to maximize profits.¹⁵

Shin (1991) considered the case of a bookmaker who faces a heterogeneous population of bettors. Some bettors are informed insiders, while others are uninformed outsiders who have heterogeneous beliefs. The N insiders are perfectly informed, so that they always pick the winning horse. In the absence of outsiders, the bookmaker would then be sure to make a loss at any finite price. In his model, the bookmaker is active thanks to the presence of outsiders, who play the role of noise traders. The outsiders are assumed to have beliefs distributed uniformly, with $q_0 \sim U[0, 1]$, and to have aggregate wealth b_0 .

Shin first established that the bookmaker sets prices such that some outsiders find it most attractive to abstain from betting. This is a natural precondition for the bookmaker's ability to make any positive profit at all. For the purpose of our analysis, this partial abstention implies that we can consider odds-setting on one horse, say horse 1, in isolation.

A unit bet on the horse under consideration (i.e., a bet that pays 1 if horse 1 wins) will have price m set by the bookmaker.¹⁶ Outsiders with beliefs above a unit bet's price place the bet. At price m , the outsiders' demand is then equal to $b_0(1 - m)$.

The bookmaker believes that horse 1 wins with chance q . He or she chooses m to maximize the profit

$$-q[b_0(1 - m) + N]\left(\frac{1 - m}{m}\right) + (1 - q)b_0(1 - m). \quad (13)$$

The bookmaker believes that horse 1 wins with probability q , in which case the bookmaker makes a net payment equal to $(1 - m)/m$ to $b_0(1 - m)$ outsiders and to the insiders. If instead horse -1 wins, which happens with probability $1 - q$, the bookmaker gains $b_0(1 - m)$, the amount placed by the outsiders on horse 1.

The monopolist bookmaker's first-order condition for choosing m to maximize Equation (13) is

$$q(b_0(1 - m) + N)\frac{1}{m^2} + b_0q\left(\frac{1 - m}{m}\right) - b_0(1 - q) = 0,$$

¹⁵Bookmakers play the role of market makers, as in Copeland and Galai (1983) and Glosten and Milgrom (1985). Shin, however, makes different assumptions about the relative elasticities of the demands of informed and uninformed traders.

¹⁶In this setting, these prices do not sum to one. Market probabilities can nevertheless be obtained by dividing these prices by their sum.

solved by

$$m = \sqrt{\frac{q}{1 - \frac{N}{b_0 + N}}}.$$

The fraction of the insiders' wealth over the total, $z = \frac{N}{b_0 + N}$ is a measure of the amount of insider information in the market. While the market price m is an increasing function of the bookmaker's belief q , now the average return to a bet on horse 1 is $q/m = \sqrt{(1 - z)q}$ which is increasing in q . This is the favorite-longshot bias.

Proposition 5 *If the bookmaker's belief is correct, $q = p$, market probabilities underreact to changes in p as m/p is decreasing in p (and m).*

Intuitively, the lower is the market price m , the fewer outsiders participate, and the greater the bookmaker chooses the bias ratio q/m to protect against the adverse selection of bettors. In this setting, the effect vanishes if bookmaking is perfectly competitive.¹⁷ The market price m then makes the profit of Equation (13) equal to zero, a condition solved by $m = q/(1 - z)$. In that case m/q is constant so there is no favorite-longshot bias.¹⁸

8. LIMITED ARBITRAGE

Hurley and McDonough (1995) explained the favorite-longshot bias on the basis of limited arbitrage by informed bettors.¹⁹ Our illustration of this logic is similar to that in Section 4, except that there is now perfect competition among the insiders.

Suppose that very (or infinitely) many insiders know that the probability that horse 1 wins is $q = p > 1/2$. Given that the number of insiders is large, it is reasonable to assume that they are price takers.

In the absence of transaction costs these insiders will place their bets such that the expected payoffs on both horses are equal. Equating these two expected payoffs gives

$$p \frac{B_0 + B_{-0}}{b_0(1) + B_{-0}} = \frac{p}{m} = \frac{1 - p}{1 - m} = (1 - p) \frac{B_0 + B_{-0}}{b_0(2)},$$

which determines the amount B_{-0} bet by the insiders on horse 1. In this case with zero track take, the equation implies $m = p$, so that there is no favorite-longshot bias.

Suppose now that the track take is positive, $\tau > 0$. Now the insiders will keep betting on the favorite, horse 1, only if the net expected payoff is non-negative. Equating to zero

¹⁷However, Ottaviani and Sørensen (2005) show that the favorite-longshot bias results in a natural model with competitive fix-odds bookmakers, when insiders are partially (rather than perfectly) informed.

¹⁸In Shin (1992), the bookmaker solves a related constrained maximization problem. In an initial stage of bidding for the monopoly rights, the bookmaker has committed to a cap $\beta > 0$ such that the implied market probabilities satisfy $\sum_{x \in \mathcal{X}} m(x) \leq \beta$. The favorite-longshot bias is derived along similar lines.

¹⁹See also Terrell and Farmer (1996).

the net expected payoff of betting on horse 1,

$$(1 - \tau)p \frac{B_0 + B_{-0}}{b_0(1) + B_{-0}} = (1 - \tau) \frac{p}{m} = 1,$$

we determine the amount B_{-0} the insiders bet on horse 1. If the insiders bet a positive amount, $m = p(1 - \tau) < p$.

Proposition 6 *Suppose that there is a positive track take, and an infinite number of insiders with correct common belief on horse 1, $p > 1/2$. If they bet on horse 1, its market probability is lower than the insider's probability, $m = p(1 - \tau) < p$, and the expected payoff on the favorite is greater than on the longshot, $1 > (1 - \tau)(1 - p)/[(1 - p) + p\tau]$.*

It can be immediately verified that the longshot's market probability is $1 - m = 1 - p + p\tau$. The expected payoff is $(1 - \tau)(1 - p)/[(1 - p) + p\tau] < 1$. Because arbitrage is limited, relatively too many bets are placed on the longshot, and the bets placed on the favorite are not sufficient to bring up the expected return on the longshot to the same level as on the favorite. The track take thus induces an asymmetry in the rational bets, resulting in the favorite-longshot bias.

9. SIMULTANEOUS BETTING BY INSIDERS

Ottaviani and Sørensen (2006) proposed a purely informational explanation for the favorite-longshot bias in the context of pari-mutuel betting. To illustrate this explanation, we consider the simplest case in which the two horses are ex ante equally likely to win ($q = 1/2$), the outsiders bets equal amounts on the two horses ($[b_0(1) = b_0(2) = b_0]$), the track take is zero ($\tau = 0$), and the number of privately informed insiders is large (with mass N).

For the purpose of the starkest illustration of this explanation, focus on the case with a continuum of insiders. The conditional distributions of the insiders' initial beliefs are such that $G(r|x = 2) > G(r|x = 1)$ for all $0 < r < 1$, given that these beliefs contain information about the outcome of the race. Conditional on the outcome, the insiders' beliefs are independent. In addition, we make the natural assumption that these beliefs have symmetric distributions: $G(r|x = 1) = 1 - G(1 - r|x = 2)$.

Since higher private beliefs are more frequent when horse 1 wins, in equilibrium each individual bets more frequently on horse x in state x . For simplicity of exposition, suppose that all insiders choose to bet (even if they make negative profits). The equilibrium has a simple form, with bettors placing their bet on horse 2 with beliefs below a cutoff level and on horse 1 above the cutoff. Given that the outsiders' bets are balanced ($[b_0(1) = b_0(2) \equiv b_0]$), the cutoff of the posterior belief at which the expected payoff of a bet on horse 1 is equalized to the expected payoff of a bet on horse 2 is $\bar{r} = 1/2$.

Conditional on horse x winning, the market probability for horse 1 is

$$m = \frac{b_0 + N [1 - G(1/2|x)]}{2b_0 + N}.$$

Horse 1 has a higher market probability (i.e., is more favored) when state 1 is true,

$$\frac{b_0 + N [1 - G(1/2|1)]}{2b_0 + N} > \frac{b_0 + N [1 - G(1/2|2)]}{2b_0 + N},$$

given that $G(1/2|2) > G(1/2|1)$. This means that the identity of the winning horse, x , is fully revealed upon observation of the market probability. In this symmetric setting with a continuum of bettors, the horse with higher market probability, $m > 1/2$, is revealed to be the sure winner, $p = 1$, and the horse with lower market probability, $m < 1/2$, is revealed to be the sure loser, $p = 0$. The market probabilities are always less extreme than the objective probabilities, hence the favorite-longshot bias.

Proposition 7 *When there is a large number of privately informed bettors, equilibrium betting with pari-mutuel payoffs results in the favorite-longshot bias.*

The bias would be reduced if bettors could instead adjust their positions in response to the final market distribution of bets (or, equivalently, the odds), as in a rational expectations equilibrium.²⁰ However, the assumption that bettors observe the final odds is not realistic. Given that a large amount of bets are placed at the end of the betting period, the information on the final market odds is typically not available to bettors. The aggregate amounts bet are observable only after all bets have been placed. The explanation that we have exposed here is based on the fact that in a Bayes-Nash equilibrium the bettors do not observe the final distribution of bets.

For example, suppose that each bettor observes a signal with conditional distributions $F(s|1) = s^2$ and $F(s|2) = 1 - (1 - s)^2$. This signal structure can be derived from a binary signal with uniformly distributed precision. With fair prior $q = 1/2$, we have $r = s$ so that $G(r|1) = r^2$ and $G(r|2) = 2r - r^2$. Hence, conditional on horse 1 winning, the market probability for horse 1 is

$$m = \frac{b_0 + N [1 - G(1/2|x)]}{2b_0 + N} = \frac{b_0 + (3/4)N}{2b_0 + N},$$

while the market probability for horse 2 is

$$1 - m = \frac{b_0 + (1/4)N}{2b_0 + N} < m.$$

Instead, conditional on the information revealed by the bets, the objective probability is 1 for horse 1 and 0 for horse 2.

²⁰If the insiders have unlimited wealth, the favorite-longshot bias would be fully eliminated in a rational expectations equilibrium.

10. TIMING OF BETS

Ottaviani and Sørensen (2004) identified two countervailing incentives for timing bets in pari-mutuel markets.

- On the one hand, bettors have an incentive to place their bets early, in order to capture a good market share of profitable bets. This effect is best isolated when there is a small number of large bettors who share the same information. These bettors have the power to affect the odds, given that they have a sizeable amount of money.
- On the other hand, if bettors have private information, they have an incentive to delay their bets. As in open auction with fixed deadline, waiting allows the bettors to conceal their private information and maybe gain the information possessed by the other bettors. To abstract from the first effect, this second effect is best isolated when bettors are small and so have no market power.

10.1. Early Betting

If bettors are not concerned about revealing publicly their private information (for example because they have no private information), they have an incentive to bet early. As also observed by Isaacs (1953) and discussed in Section 4, with pari-mutuel betting the expected return on each additional dollar bet on horse 1 is decreasing in the amount the insider bets on this horse.

Next, consider what happens when there are two insiders. Suppose that there are two large strategic bettors, insider 1 and 2, who both think that horse 1 is more likely to win than horse 2. These bettors need to decide *how much* to bet and *when* to bet. For simplicity, suppose that there are just two periods, $t = 1$ (early) and $t = 2$ (late). Each bettor prefers to bet early rather than late because by being early a bettor can secure a higher payoff and steal profitable bets from the other bettor. Both bettors do so, and in equilibrium they end up both betting early.²¹

Pari-mutuel betting among the two insiders is thus a special case of the classic Cournot (1838) quantity competition game. The result that betting takes place early is a corollary of Stackelberg's (1934) result that under quantity competition the first mover (or leader) derives higher payoff than the second mover (or follower).²²

10.2. Late Betting

In addition to the incentive to bet early discussed above, Ottaviani and Sørensen (2004) also analyzed the incentive to bet late in order to conceal private information and maybe observe others, as in an open auction with a fixed deadline. This second effect is best

²¹Pennock's (2004) dynamic modification of pari-mutuel payoffs would further increase the incentive to bet early.

²²Ottaviani and Sørensen (2004) extend this bet timing result to the case with more than two bettors.

understood by considering the case with small bettors without market power, but with private information.

It would be interesting to characterize the interplay of the two contrasting timing incentives that are present when market power and private information coexist.

References

- Ali, M. M. 1977. Probability and Utility Estimates for Racetrack Bettors, *Journal of Political Economy* 85(4), 803–815.
- Blough, S. R. 1994. Differences of Opinions at the Racetrack, in D. B. Hausch, V. S. Y. Lo, and W. T. Ziemba (eds.), *Efficiency of Racetrack Betting Markets*. Academic Press, San Diego, CA, pp. 323–341.
- Borel, É. 1938. Sur le Pari Mutuel, *Comptes Rendus Hebdomadaires des Séances de l'Académie des Sciences* 207(3), 197–200.
- Borel, É. 1950. *Éléments de la Théorie des Probabilités*. Albin Michel, Paris.
- Chadha, S., and R. E. Quandt. 1996. Betting Bias and Market Equilibrium in Racetrack Betting, *Applied Financial Economics* 6(3), 287–292.
- Copeland, T. E., and D. Galai. 1983. Information Effects on the Bid-Ask Spread, *Journal of Finance* 38(5), 1457–1469.
- Cournot, A. A. 1838. *Recherches sur les Principes Mathématiques de la Théorie des Richesses*. Hachette, Paris.
- Eisenberg, E., and D. Gale. 1959. Consensus of Subjective Probabilities: The Pari-Mutuel Method, *Annals of Mathematical Statistics* 30(1), 165–168.
- Glosten, L., and P. Milgrom. 1985. Bid, Ask and Transaction Prices in a Specialist Market with Heterogeneously Informed Traders, *Journal of Financial Economics* 14(1), 71–100.
- Golec, J., and M. Tamarkin. 1998. Bettors Love Skewness, Not Risk at the Horse Track, *Journal of Political Economy* 106(1), 205–225.
- Griffith, R. M. 1949. Odds Adjustment by American Horse Race Bettors, *American Journal of Psychology* 62(2), 290–294.
- Hausch, D. B., and W. T. Ziemba. 1995. Efficiency of Sports and Lottery Betting Markets, in Robert A. Jarrow, Vojislav Maksimovic, and William T. Ziemba (eds.), *Handbook of Finance*. North-Holland, Amsterdam, pp. 545–580.
- Hurley, W., and L. McDonough. 1995. Note on the Hayek Hypothesis and the Favorite-Longshot Bias in Parimutuel Betting, *American Economic Review* 85(4), 949–955.
- Isaacs, R. 1953. Optimal Horse Race Bets, *American Mathematical Monthly* 60(5), 310–315.
- Jullien, B., and B. Salanié. 2008. Empirical Evidence on the Preferences of Racetrack Bettors, in D. B. Hausch and W. T. Ziemba (eds.), *Handbook of Sports and Lottery Markets*. North-Holland, Amsterdam, pp. 27–49.
- Ottaviani, M., and P. N. Sørensen. 2004. The Timing of Pari-Mutuel Bets. Mimeo, London Business School and University of Copenhagen.
- Ottaviani, M., and P. N. Sørensen. 2005. Parimutuel Versus Fixed-Odds Markets. Mimeo, London Business School and University of Copenhagen.
- Ottaviani, M., and P. N. Sørensen. 2006. Noise, Information, and the Favorite-Longshot Bias. Mimeo, London Business School and University of Copenhagen.
- Pennock, D. M. 2004. A Dynamic Pari-Mutuel Betting Market for Hedging, Wagering, and Information Aggregation, *ACM Conference on Electronic Commerce*.
- Prelec, D. 1998. The Probability Weighting Function, *Econometrica* 66(3), 497–527.
- Quandt, R. E. 1986. Betting and Equilibrium, *Quarterly Journal of Economics* 101(1), 201–207.
- Sauer, R. D. 1998. The Economics of Wagering Markets, *Journal of Economic Literature*, 36(4), 2021–2064.
- Shin, H. S. 1991. Optimal Betting Odds Against Insider Traders, *Economic Journal* 101(408), 1179–1185.
- Shin, H. S. 1992. Prices of State Contingent Claims with Insider Traders, and the Favourite-Longshot Bias, *Economic Journal* 102(411), 426–435.

Stackelberg, H. 1934. *Marktform und Gleichgewicht*. Springer, Vienna.

Terrell, D., and A. Farmer. 1996. Optimal Betting and Efficiency in Pari-Mutuel Betting Markets with Information Costs, *Economic Journal* 106(437), 846–868.

Thaler, R. H., and W. T. Ziemba. 1988. Anomalies: Pari-Mutuel Betting Markets: Racetracks and Lotteries, *Journal of Economic Perspectives* 2(2), 161–174.

Weitzman, M. 1965. Utility Analysis and Group Behavior: An Empirical Analysis, *Journal of Political Economy* 73(1), 18–26.

DOWNLOAD FULL VERSION HERE: <https://bettingbooks.org/12>

This page intentionally left blank

Download More Sports Betting Books at: www.bettingbooks.org

CHAPTER 7

**Examining Explanations of a Market
Anomaly: Preferences or Perceptions?**

Erik Snowberg

Division of Humanities and Social Sciences, California Institute of Technology, Pasadena, CA, USA.

Justin Wolfers

The Wharton School, University of Pennsylvania, Philadelphia, PA, USA.

1. Introduction	104
2. Preferences—Expected Utility Models with Linear Probabilities	108
3. Perceptions—The Weighting of True Probabilities	111
4. Perceptions—Informational Effects	112
5. Definition of Models and Implications for Combinatoric Bets	114
6. Using Combinatoric Markets to Test the Models	119
6.1. <i>Testing Conditional Independence</i>	123
6.2. <i>Relaxing Conditional Independence Further</i>	125
7. Conclusion	129
<i>References</i>	130
Appendix A: Pricing of Combinatoric Bets Using Conditional Independence	134
Appendix B: Data	135

Acknowledgments

We would like to thank David Siegel of Trackmaster for supplying the data, and Scott Hereld and Ravi Pillai for their valuable assistance in managing the data. Betsey Stevenson, Matthew White, and William Ziemba provided valuable insights. Later sections of this article draw heavily from our paper “The Favorite-Longshot Bias: Is it Risk-Love or Misperceptions?”

Abstract

This paper compiles and summarizes the theoretical literature on the favorite-longshot bias, an anomaly that has been found in sports betting markets for over half a century. Explanations of this anomaly can be broken down into two broad categories, those involving preferences and those involving perceptions. We propose a novel test of these two classes of models that allows us to discriminate between them without parametric assumptions. We execute these tests on a new dataset, which is an order of magnitude larger than any used in previous studies, and conclude that the perceptions model, in which bettors overestimate the chances of small probability events, provides a better fit to the data.

1. INTRODUCTION

Gambling has long been of interest to economists as it provides a natural laboratory for studying decision-making under uncertainty. In addition, the very existence of gambling is difficult to rationalize in a neo-classical framework along with the fact that gamblers often purchase insurance at actuarially unfair premiums. The first behavior is evidence of risk-love, and the latter is evidence of risk aversion. Although progress has been made, there is no single agreed upon way to understand and model such behavior.

Any model that explains the purchase of insurance and gambling must also explain other features of gambling markets. The longest standing empirical regularity of these markets is the favorite-longshot bias: bettors systematically overbet longshots and underbet favorites relative to their true probabilities of winning. This anomaly has persisted since it was discovered in horse racing markets over half a century ago.¹ Figure 1 shows the favorite-longshot bias in our data.

Figure 2 shows the same rate of return calculations for several other datasets. We present new data from 2,725,000 starts in Australia from the South Coast Database, and 380,000 starts in Great Britain from flatstats.co.uk. The favorite-longshot bias appears equally evident in these countries, despite the fact that odds are determined by a bookmaker-dominated market in the United Kingdom, and bookmakers competing with a state-run pari-mutuel market in Australia.² Figure 2 also includes historical estimates of the favorite-longshot bias, showing that it has been stable since it was first noted in Griffith (1949).³

¹Griffith (1949), McGlothlin (1956).

²This is consistent with Dowie (1976) and Crafts (1985).

³The favorite-longshot bias has been documented in the gambling markets of many other sports, while the opposite pattern—often termed the reverse favorite-longshot bias—has been observed in some others. Woodland and Woodland (1994, 2001, 2003) find a reverse favorite-longshot bias in baseball and hockey. Gray and Gray (1997) find a favorite-longshot bias in football, and Metrick (1996) finds a reverse favorite-longshot bias in office NCAA pools. Cain and Peel (2000) find a favorite-longshot bias in UK football (soccer) betting markets. Gander et al. (1998) find it in New Zealand horse racing markets. Busche and Hall (1988) document that the favorite-longshot bias does not exist in Japanese and Hong Kong horse racing markets, and in fact, a reverse favorite-longshot bias may be present.

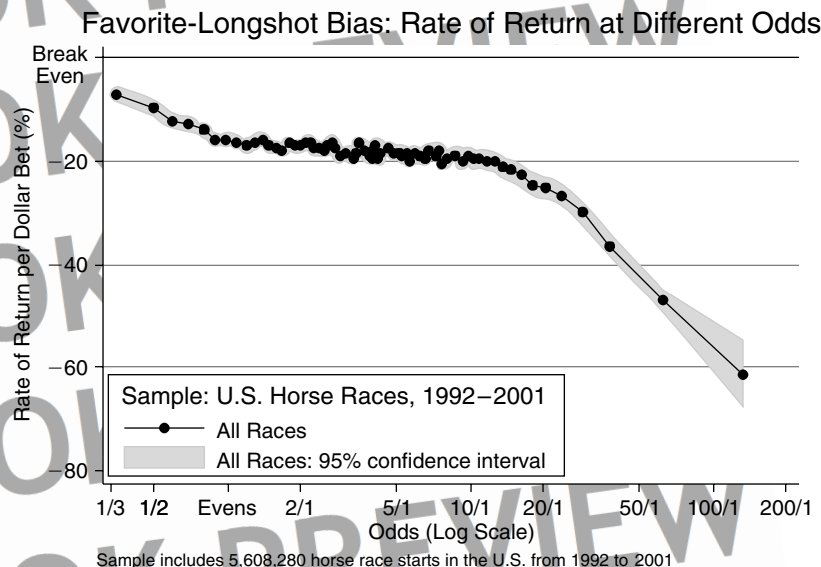


FIGURE 1 The favorite-longshot bias.

There are two related regularities of interest, although they have not been as extensively documented. The first is the tendency for gamblers to exhibit greater risk preference in the last race or two as compared with earlier races (McGlothlin, 1956; and Ali, 1977). We call this the *end of the day difference*. The second regularity is positive expected profits on horses with very short odds—that is, overwhelming favorites (Hausch and Ziemba, 1995). These features do not appear in our dataset. These two regularities may have been the result of statistical imprecision, or the market may have corrected them over time. The latter hypothesis is suspect, however, as the favorite-longshot bias has persisted while these other anomalies have faded.

Starting with Rossett (1965), gambling markets have been used to examine the sophistication and rationality of gamblers. Snyder (1978) uses a gambling market, specifically betting on horse races, to examine market efficiency.⁴ Since then, gambling markets have provided a fertile ground for financial economists testing the efficient markets hypothesis. As Thaler and Ziemba (1988) note:

The advantage of wagering markets is that each asset (bet) has a well-defined termination point at which its value becomes certain. The absence of this property is one of the factors that has made it so difficult to test for rationality in the stock market. Since a stock is infinitely lived, its value today depends both on the present value of future cash flows *and* on the

⁴There is a large body of literature focused on testing market efficiency in gambling markets. A sample includes: Figelewski (1979) and Losey and Talbott (1980) who study horse racing, Zuber et al. (1985), Sauer et al. (1988), Golec and Tamarkin (1992) and Gander et al. (1998) who study football, and Brown and Sauer (1993) who study basketball.

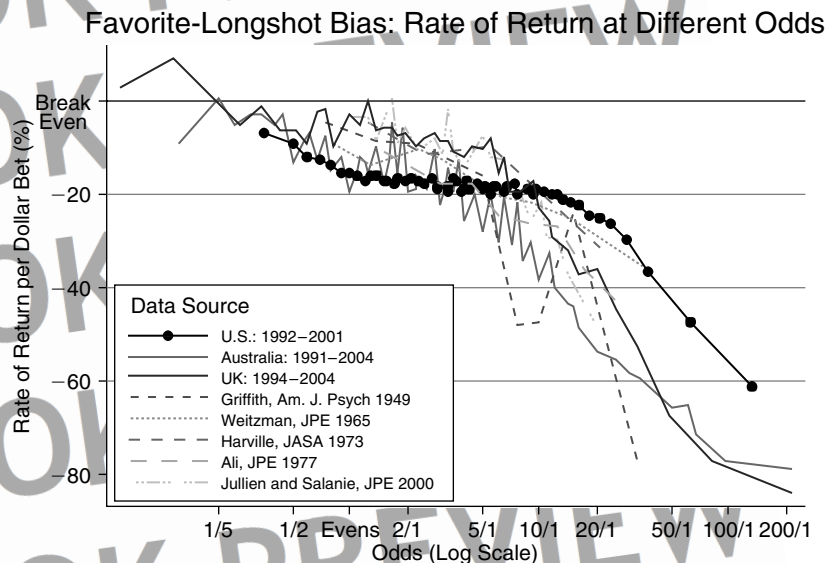


FIGURE 2 The favorite-longshot bias across the world and time.

price someone will pay for the security tomorrow. Indeed, one can argue that wagering markets have a better chance of being efficient because the conditions (quick, repeated feedback) are those which usually facilitate learning. However, empirical research has uncovered several interesting anomalies.

The favorite-longshot bias is interesting then for two reasons. The first is that any theory that explains the taking of small unfair gambles and the purchase of insurance must explain this feature. The favorite-longshot bias indicates that gamblers are willing to accept greater risk with lower expected return, which flies in the face of the intuitive risk-return trade-off. The second is that in some cases it provides evidence against weak-form market efficiency.⁵

This chapter summarizes existing theories of the favorite-longshot bias and shows that those theories stress either preferences or perceptions. We then discriminate between those classes of theories using data from *combinatoric* gambling markets.⁶

⁵Weak form tests of market efficiency check that current prices reflect all information contained in past prices. That is, it should not be possible to come up with a profitable strategy only based on past prices. Positive returns on extreme favorites would violate weak form efficiency since it would be possible to make a profit simply by betting on extreme favorites as defined by horses with odds of 0.2/1 or less. For more on the efficient markets hypothesis and a summary of tests see Fama (1970).

⁶These gambles are often called *exotics* and consist of exactas, which is a bet on a given horse to finish first and a given horse to finish second; quinellas, which is a bet on two horses to come first and second in either order; and trifectas, which is a bet on given horses to come in first, second, and third. We call these bets *combinatoric* because they are generically available (i.e., not exotic) and this taxonomy is a more accurate representation of their implementation.

Theories of the favorite-longshot bias can be split into two classes of theories. First, standard neo-classical theory suggests that the price one is willing to pay for various gambles can be used to recover a utility function. While betting at any odds is actuarially unfair, the data suggests that this is particularly acute for longshots—which are also the riskiest investment. Thus, the neo-classical approach can reconcile both gambling and the longshot bias by positing locally risk-loving utility functions (Friedman and Savage, 1948). Because this rationalization of the favorite-longshot bias relies on the shape of the utility function, we refer to it as a *preference-based* model. Alternatively, behavioral theories suggest that cognitive errors play a role in market mispricing. These theories generally point to laboratory studies by cognitive psychologists that suggest that people are systematically poor at discerning between small and tiny probabilities (and hence they will price each similarly). Further, certain events are strongly preferred to extremely likely events, leading even very likely events to be underpriced. These results form a key part of Kahneman and Tversky's (1979, 1992) prospect theory. Beyond the laboratory, these theories can rationalize the purchase of sometimes extremely unfavorable lottery tickets, and actuarially unfair insurance on items like internal telephone wiring.⁷ We term these nonexpected utility theories perception-based because they rationalize the favorite-longshot bias by referring to (*mis*)perceptions of the likelihood of each horse winning.

There is also a third class of models that focus on groups of bettors with different information available to them. These models, however, yield implications in aggregate data that are equivalent to a model with a representative agent that bases decisions on a set of weights that diverge from true probabilities. We will argue that all of the models posited to date yield observationally equivalent results to those that posit either a risk-loving or misperceiving representative agent.⁸ As such, the preferences versus perceptions distinction is not so much between two sharply defined theories, but rather a taxonomy for organizing the two sets of theories. Alternatively, using the language in Gabriel and Marsden (1990), we ask: “are we observing an inefficient market or simply one in which the tastes and preferences of the market participations lead to the observed results?”

The rest of this chapter proceeds as follows. In Sections 2 through 4, we review the favorite-longshot literature and categorize the theories into our perceptions versus preferences taxonomy, with an additional section exploring theories based on informational effects. We then lay out the implications of both preference- and perception-based theories for the pricing of exotic bets. This is followed by our empirical findings. In summary, the pricing function implied by the perception-based models better matches the observed prices of exactas, quinellas, and trifectas. The key implication is that

⁷Camerer (2001).

⁸For instance, a preference for the “bragging rights” that accrue from picking a longshot that wins (Thaler and Ziemba, 1988) yields predictions that are observationally equivalent in aggregate data to a risk-loving representative agent. This is also true of models with heterogeneity, where some naifs bet randomly (and hence overbetting the favorite), and imperfect arbitrage preventing sophisticated bettors from offsetting this. See Section 5 for further discussion.

rationalizing prices of win bets and combinatoric bets require a utility function that is not linear in probabilities. Section 7 reviews the robustness of this result, and concludes.

2. PREFERENCES—EXPECTED UTILITY MODELS WITH LINEAR PROBABILITIES

Thaler and Ziemba (1988) list five explanations that are commonly offered to rationalize the favorite-longshot bias, to which we add one more. They are:

1. Bettors are locally risk-loving.
2. Bettors may derive utility simply from holding a ticket on a longshot.
3. Bragging rights for predicting that a longshot comes through are much higher.
4. Bettors might overestimate the chances that the longshots will win.
5. Bettors might overweight the small probability of winning in calculating the utility of the bet.
6. Some bettors may choose horses for essentially irrational reasons like the horse's name.

The first three reasons specify a representative bettor's preferences, reasons four and five specify how a representative gambler views probabilities, and reason six implies that participants in gambling markets are asymmetrically informed. We have chosen to divide the theories into three categories, rather than the more familiar neo-classical/behavioral division, for three reasons. The first is that all of the above reasons, with the exception of the first, are to some extent behavioral. The second is that the classical distinction made in the literature is misleading and an artifice of the way the literature has developed. Finally, these categories are amenable to discrimination through empirical tests we propose and carry out in this chapter.

One of the earliest and best known attempts to explain a consumer's purchase of insurance and acceptance of unfair gambles simultaneously is that of Friedman and Savage (1948). This theory posits a wiggle in the consumer's utility function that can make consumers exhibit risk-loving behavior in certain circumstances even though they are generally risk-averse.⁹ That is, consumers may be locally risk-loving, but globally risk-averse agents.

The first attempts by economists to explain the favorite-longshot bias relied on the Friedman-Savage theory.¹⁰ Weitzman (1965) introduces the concept of a representative bettor. In equilibrium, each bettor must be indifferent between the odds offered on all horses. If the odds were too high on some horse, then the representative bettor would

⁹By locating this wiggle at the level of current wealth, Markowitz (1952) avoided several pitfalls of the Friedman-Savage theory.

¹⁰Although not applied specifically to the favorite-longshot bias, Brunk (1981) argues that the fact that lottery play is highly correlated with dissatisfaction about current levels of income, while gambling on horse races (which has smaller payoffs) is not as consistent with the Friedman-Savage hypothesis.

bet on that horse, driving the odds down. If the odds were too low, the opposite would happen.¹¹ Weitzman (1965) fits an indifference curve to race betting data and finds that the representative bettor is risk-loving.

Ali (1977) uses a similar technique and also finds that gamblers are risk-loving. He suggests that bettors display constant relative risk aversion. This, coupled with risk-love, implies that a bettor's coefficient of absolute risk aversion increases with wealth—that is, a bettor becomes more risk-averse as his or her wealth rises. Thus, if bettors take a position in each race and bet a constant fraction of their wealth, then, due to the track take and other costs, the aggregate wealth of the bettors will be lower in the last race than in the first.¹² As wealth declines, the representative bettor becomes more risk-loving, rationalizing the end of the day difference discussed in the introduction.

The most explicit explanation of the preferences theory is in Quandt (1986), which shows that the favorite-longshot bias is a necessary condition of equilibrium with risk-loving gamblers. A risk-lover's utility increases as the variance of a bet increases, thus, they must be willing to accept a lower expected payout for a higher variance bet (a bet on a longshot).

A subclass of preference theories suggests that gamblers are either risk-averse or risk-neutral, but have a definite preference for positive skew. That is, bettors derive greater utility from an increase in the third moment of the distribution of a gamble. Evidence of skew-love is found in the finance literature; thus, we might expect bettors to exhibit similar behavior. Evidence of skew-love is found first in Bird et al. (1987) followed by Golec and Tamarkin (1998). Golec and Tamarkin note that skew-love is consistent with taking multiple bets throughout a night, and taking multiple bets in a single race. Risk-love is not consistent with these facts. Finally, skew-love can be seen as more consistent (than risk-love) with gamblers' participation in minus pools.¹³

A final subclass of preference theories models gamblers as having an extra utility from gambling. The most persuasive model is Conlisk (1995), although Fishburn (1980) and others have proposed similar explanations. Such an explanation is generally avoided because, as Conlisk notes, it seems “to be devoid of any scientific content.” However, it can be said that this rationalization is no more ad hoc than a wiggle in a utility function. An additional utility from gambling is not always associated with the favorite-longshot bias, but the additional utility can easily be structured in such a way as to yield the bias.

¹¹More specifically, “If the price on a ticket that pays \$1 if the horse wins was lower than a fair bet based on the agent's subjective assessment of the probability of that horse winning is too high.”

¹²The track take is the amount the betting establishment withholds from the betting pool, and from which most of its profits are derived. In addition there is “breakage” due to the rounding down of odds. Track take plus breakage is usually 15–25% of the total amount bet.

¹³A minus pool occurs when a horse is an overwhelming favorite. Since in most places the minimum payout on a \$2 bet is \$2.10, if a horse is a strong enough favorite, then the track will actually lose money. Such bets have positive expected return and very small variance. A risk-lover would prefer a bet with higher variance, although the bettor would have to trade this off versus the lower expected return of higher variance bets at the racetrack. Hamid et al. (1996) find the opposite: bettors are risk-loving and have an aversion to positive skew.

As noted in Asch and Quandt (1990), it is impossible to distinguish between a representative bettor with a risk-loving utility function and one that has additional utility for holding a ticket on a longshot. Intuitively, there is no difference between loving risk and loving a risky activity. However, Piron and Smith (1994) attempt to distinguish between these two theories. They use a laboratory gambling situation to try to eliminate the consumption effects of gambling and find that the favorite-longshot bias still exists. However, if the additional utility were due to placing a risky bet on a longshot or bragging rights, it would be impossible to get rid of this consumption effect no matter how sterile and boring the laboratory setting.

The additional utility in Conlisk's model is an additive term that only takes effect when the bettor gambles. This is functionally similar to mental accounting. Thaler and Ziemba (1988) explicitly apply this concept to gambling on horse races.

The idea behind mental accounting is that agents keep different accounts in their head for different activities. Thus, losses or gains incurred in other realms of life have nothing to do with an agent's behavior in the separate—mentally and physically—realm of gambling. Although Thaler (1985) indicates that agents have the same value or utility function for all their mental accounts, the suggestion in Thaler and Ziemba (1988) is that the value function for the gambling account is risk-loving, while those for most other everyday activities are risk-averse. The resulting utility functions in this model may look much like the Friedman-Savage or other models discussed above, except that rather than being risk-loving around a certain wealth level, or “location” in the wealth scale, risk-love is local to the a physical location, specifically, the racetrack. Mental accounting escapes many of the difficulties of pure expected utility theory because it does not force the theorist to posit a single utility function that has certain apparently contradictory features.

Mental accounting can also be used to rationalize the end of the day difference (Camerer, 2001). The effect is very neatly explained by gamblers who open an account at the beginning of a day at the track with the intention of closing it at the end of the day with a profit.

Ali (1977) posits a slightly different model of the favorite-longshot bias. Two bettors bet on two horses—one on the longshot and the other on the favorite. Each must gamble \$1, thus, the odds are always 1/1 on both horses. This means the favorite is underbet and the longshot is overbet. As illustrated here, the driving assumption is somewhat unreasonable. Despite this weakness, Ali's model has been extended to prediction markets.¹⁴

Finally, Bradley (2003) relaxes the assumption that bettors make equal sized bets. This assumption is explicitly or implicitly made by all of the theories thus far. This model endogenizes the bet size while allowing the value function for losses to differ from that for gains. In order for the optimal bet size to be finite and different from zero, it must be that the value function for losses is less concave (or convex) than the one for gains—which is consistent with prospect theory.

¹⁴See Manski (2006), where a similar model is used to show that prediction market prices may be inaccurate. For a more complete description of prediction markets, see Snowberg et al. (this volume).

3. PERCEPTIONS—THE WEIGHTING OF TRUE PROBABILITIES

There are two obvious ways to extend the theories in the previous section. The first incorporates a subjective view of probabilities into the representative bettor's value function. The second is to change our model to endogenize features of the real world that we feel are not accurately represented. This section addresses theories of the first type, and the next section addresses theories of the second type.

Theories involving subjective probabilities draw their inspiration from Kahneman and Tversky's (1979) prospect theory. Prospect theory modifies expected utility theory in two important ways. The first is that all expected utility calculations are made with respect to current wealth. Expected utility is evaluated over gains and losses, not over possible ending wealth levels, through a value function. Additionally, expected utility is not linear in probabilities. Instead, objective probabilities enter expected utility through an arbitrary weighting function that multiplies the value function. In addition, both the value and weighting functions can be different for gains and losses. Note that prospect theory nests expected utility theory, that is, we can regain the expected utility framework by simply making the weighting function equal to the probabilities and making the value function equal for gains and losses.¹⁵

Jullien and Salanié (2000) use maximum likelihood estimation to fit gambling data to prospect theory as well as expected utility theory.¹⁶ They find that the additional coefficients allowed by prospect theory are economically and statistically different from the values that would return expected utility theory through the nesting described above. They find that the additional flexibility of prospect theory is appropriate in describing the favorite-longshot bias.

In Jullien and Salanié's results, the weighting function for losses is quite concave—it overweights small probabilities. This is consistent with the fifth explanation offered in the previous section and is intuitively very simple to understand. The weighting function can be interpreted as a bettor's subjective evaluation of the true probability of a horse winning. Thus, if bettors overweight low probabilities, they will believe that low odds horses are more likely to win than they actually are, and will bet more money on them.

Camerer (2001) offers another interpretation of Jullien and Salanié's results. He notes that bettors weight their high chances of losing and small chances of winning roughly linearly. They hate to bet on favorites not because they are risk-loving, but because they are disproportionately afraid of the small chance of losing when betting on a favorite.

¹⁵Although we will posit a risk-neutral representative agent for our perceptions based model, prospect theory allows for a risk-averse agent who still takes unfair gambles. This would also allow us to take advantage of theories like those of Woodland and Woodland (1991) which explains why some gambling markets are based on odds, and others on point spreads. It should be noted that there is a competing theory in this realm—that of Bassett (1981), which appeals to profit maximization of the market maker. Data from the emergence of internet gambling markets, which have significantly lower profit margins for market makers, should allow one to test these two theories against each other.

¹⁶Jullien and Salanié (this volume) update their results.

An oft-overlooked and elegant explanation of the favorite-longshot bias that centers on perceptions is that of Henery (1985). If bettors discount a constant fraction of their losses, then the average return to a bettor from a bet at given odds decreases with the odds. This is exactly the favorite-longshot bias.

Mathematically, the idea that bettors discount a constant fraction of their losses can be represented as follows. If the true probability of a horse losing a race is $(1 - p_i)$ and the bettor discounts his or her losses by the fraction $(1 - f)$, then he or she believes that the probability of winning is actually $f(1 - p_i) + p_i$. Intuitively (or behaviorally) this discounting could take the form of ignoring losses where the bettor's horse came "really close" to winning.

4. PERCEPTIONS—INFORMATIONAL EFFECTS

Most recent explanations of the favorite-longshot bias have focused on the information sets of bettors. However, since it is the marginal, not the average dollar that determines the final odds, the presence of a large mass of uninformed, or wrongly informed, bettors alone cannot explain the favorite-longshot bias.¹⁷

To develop the intuition of these models, start with a simple case. There are two horses and two groups of bettors. One is a group of uninformed bettors who bet on the longshot and whose total bets are determined exogenously. Then a group of informed, risk-neutral bettors have a chance to bet. These informed bettors know the true probabilities of a horse winning and continue to bet until the odds reflect the true probabilities. This illustrates that information asymmetries alone cannot explain the favorite-longshot bias.

Hurley and McDonough (1996) extend this simple model by adding the transaction cost of the track take. This is sufficient to generate the favorite-longshot bias. As before, the informed bettors will bet on the favorite until all expected profits have been taken off the table. In the presence of the track take it is not profitable to bet on the favorite at odds that are only slightly better than an actuarially fair bet. The odds have to be much more favorable than a fair bet in order to make up for the fact that a certain portion of the profits must go to the track. Thus, the favorite is underbet, and the longshot is overbet. Note that the longshot has negative expected profits in this model. In a normal financial market this could not persist because investors could simply short sell the longshot. However, this model also disallows short selling, which allows the favorite-longshot bias to persist.

Terrell and Farmer (1996) extend this model by endogenizing the cost to bettors of becoming informed. They find that the magnitude of the favorite-longshot bias depends on the track take and relative probabilities of the horse's winning, as well as the cost of becoming an informed bettor.

¹⁷A trip to the track is enough to convince most people that gamblers bet based on different information sets. There are also some empirical tests in the literature that find evidence of this. Gander et al. (1998) finds evidence of informed bettors in basketball betting. Crafts (1985) finds evidence of insider knowledge in British horse race betting.

Most models are significantly more complex than the one offered above. However, in each model, what drives the favorite-longshot bias is not the asymmetry of information per se, but additional marketplace restrictions that are incorporated into the model. In these models, asymmetric information is a necessary, but not sufficient, condition for the existence of the favorite-longshot bias.

Vaughan Williams and Paton (1998) create a model where the uninformed bettors bet on all horses evenly, and the informed bettors have an additional utility for betting on the favorite. They find that if the additional utility for betting on the favorite is positive, then even if the track take is zero, there will be a favorite-longshot bias. This is essentially the same as the basic model above, except the cost that prevents the equalization of odds with true probabilities comes from forgoing the extra utility due to betting on a favorite rather than the transaction costs from the track take. This model has an additional implication, which is that if the transaction costs in the betting market are less than the extra utility of betting on the favorite, then a reverse favorite-longshot bias will exist.

Shin (1991, 1992, 1993) models betting markets controlled by a bookmaker. As noted above, the favorite-longshot bias has been shown to exist in these markets. In his model outsiders bet rationally, that is, in proportion with the actual win probabilities of the horses. In addition, there are a certain proportion of the bettors who are insiders and learn ahead of time which horse will win the race. As long as the proportion of insider bets does not increase faster than the inverse of the probability that a horse will win, then the favorite-longshot bias exists as the optimal response of the bookmaker to the presence of informed bettors.¹⁸ Moreover, by offering odds that exhibit the favorite-longshot bias, the bookmaker increases his or her profits. This accords with Levitt's (2004) finding that bookmakers do not set odds to clear the market, but rather to exploit known biases in the betting public's behavior.

Laboratory settings have found that the presence or absence of a track take has no effect on the size or existence of the favorite-longshot bias, which does not accord with the above theories.¹⁹ Potters and Wit (1996) model informed bettors with asymmetric information who do not take into account the information revealed about others bettors' signals through the market odds when they place their bets. Since each bettor receives an independent signal of win probabilities from a known distribution, having more signals would allow a more precise estimate of the true win probabilities. By watching the bets of others, a bettor would be able to capture this information and make a better assessment of the true odds—and the odds would then converge to the objective probabilities. However, by ignoring the bets of others, a bettor is relying too heavily on his or her own (imprecise) signal, which introduces a bias into the betting market. This effect is very similar to the winner's curse in auctions with private information.

¹⁸That is, the proportion of insiders cannot be significantly higher when a favorite wins than when a longshot wins. A constant amount of insider bets easily satisfies this condition.

¹⁹See Hurley and McDonough (1995). Although by examining fixed odds vs. spread betting, Vaughan Williams and Paton (1998) find that transaction costs may be partly responsible for the favorite-longshot bias.

Ottaviani and Sørensen (2003) exploit this insight in their own models by only allowing informed gamblers to bet simultaneously at the end of the betting period. Thus, the informed bettors are unable to gather information contained in the market odds. Hence, the betting exhibits a favorite-longshot bias. The idea that informed bettors place their bets last has some empirical support. Asch et al. (1982) find that the bets placed in the last 5 minutes show evidence of informed betting.

In the Potters and Wit model, sequential and simultaneous betting are the same since each gambler knows the other bettors will not take into account the information revealed by their bets. However, if bettors believe that others will use their bets to deduce something about their private information, they must take this into account and act strategically in equilibrium. This is the approach taken by Koessler and Ziegelmeyer (2002). They note that in a simultaneous game with asymmetric information, a fully revealing equilibrium always exists and becomes unique as the number of bettors grows large. In contrast, in a sequential game, a fully revealing equilibrium does not exist as the number of bettors grows large. When a fully revealing equilibrium does not exist, the game is subject to extreme prices and persistent biases. Note that this is not inconsistent with Ottaviani and Sørensen, since in that model after the final odds are posted everyone can see that the longshot was overbet—there is just nothing they can do about it. Koessler et al. (2003) extend this insight by proposing a sequential game in which a favorite-longshot bias arises. However, the strategic bettors in their model have subjective attitudes toward probabilities, which, as we saw in the last section, is enough to create a favorite-longshot bias even without asymmetric information.

5. DEFINITION OF MODELS AND IMPLICATIONS FOR COMBINATORIC BETS

We start with two extremely stark models, each of which has the merit of simplicity. Both are representative agent models, but as we suggest below, can be usefully expanded to incorporate heterogeneity. Ultimately, aggregate price data will not be able to separately identify more complex models from these representative agent models.

Under the preference-based approach, we start by postulating an expected utility maximizer with unbiased beliefs. In equilibrium, bettors must be indifferent between betting on the favorite horse (at odds of $F/1$, and a probability of winning of f), and betting on a longshot (at odds of $L/1$, with probability of winning, l)

$$fU(F) = lU(L) \quad (\text{normalizing utility to zero, if the bet is lost}).^{20}$$

²⁰We also assume that each bettor chooses to bet on only one horse in a race. This simplifies the analysis, as otherwise we would need to also keep track of the fraction of wealth that each bettor chooses to gamble.

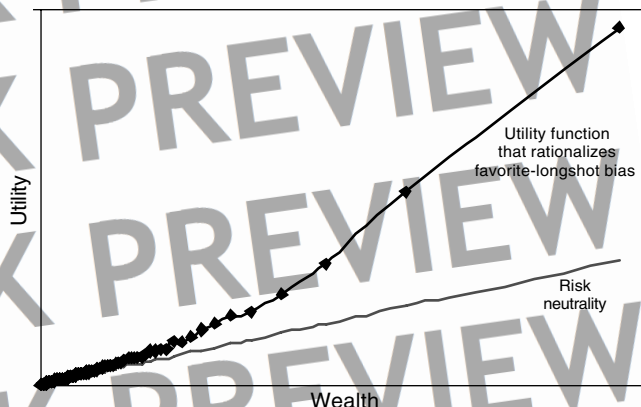


FIGURE 3 Utility function implied by betting odds.

Given that we observe in the data the probabilities of each horse winning (f, l) and their odds (F, L), these data reveal the representative bettor's utility function. That is, by comparing the betting odds and the winning probability of pairs of “nearby” bets, we can identify the utility function of the representative bettor (up to a scaling factor).²¹ In order to simplify notation, we normalize so that utility is zero if the bet loses, and utility is one if the bettor chooses not to bet. Thus, if the bettor is indifferent as to whether to accept a gamble paying odds G , with probability g , then $U(G) = g^{-1}$. Figure 3 performs precisely this analysis, backing out the utility function required to fully rationalize the choices shown in Figure 1.

The observation that bettors are willing to accept bets on horses at increasingly long odds, even as the expected payoff becomes increasingly actuarially unfair, implies that $U'' > 0$. Beyond this, the specific shape of the declining rates of return identifies the slope of the utility function at each point.

As can be seen quite clearly, a risk-loving (or locally risk-loving) utility function is required in order to rationalize these results. The utility function—by construction—fully explains all of the variation in Figure 1.²² Several other theories of the favorite-longshot bias have also been proposed that yield implications that are observationally equivalent to a simple risk-loving representative agent model. For instance, Thaler and Ziemba (1988), echoing Hausch and Ziemba (1995) argue that bragging rights accrue from winning a bet at long odds. Formally, this suggests agents maximize expected utility, where utility is the sum of the felicity of wealth, v , and the felicity of bragging rights or the thrill of winning, w , and hence the expected utility of a gamble that returns

²¹ See Weitzman (1965), Ali (1977), and Jullien and Salanié (2000) for examples.

²² Imposing a CRRA functional form on the utility function, a simple non-linear probability model yielded an estimate of the coefficient of relative risk aversion of -0.16 (standard error 0.0006; $n = 4,867,857$, which is remarkably close to Ali's (1977) estimate.

G with probability g can be expressed as

$$EU(G) = g[w(G) + v(x)] + (1 - g)w(0)$$

where the initial wealth level is normalized to zero.

As in the representative agent model, bettors will be prepared to accept lower returns on riskier wagers (betting on longshots) if $U'' > 0$. This is possible if either utility of wealth is sufficiently convex ($w'' > 0$), or bragging rights are increasing in the payoff at a sufficiently increasing rate. More to the point, decisions under uncertainty do not allow us to separately identify effects operating through w , rather than v , and this is the sense in which the model is observationally equivalent with the simple representative agent who is risk-loving. A similar argument shows that a model in which the mere purchase of a ticket on a longshot confers some utility (such as the dream of striking it rich), is also observationally equivalent.

Alternatively, under the perceptions-based approach, we postulate a subjective, risk-neutral, expected wealth maximizer. The agent's subjective beliefs, π , are systematically biased estimators of the true probabilities. In equilibrium, bettors must believe that the rates of return to betting on the favorite and the longshot are equal, and hence

$$\pi(f)F = \pi(l)L.$$

Consequently, data on the probabilities of each horse winning (f, l) and their odds (F, L), reveal the systematic component of the representative bettor's subjective expectations. Thus, we can identify the decision weights of the representative bettor, and Figure 4 (which is a simple transform of Figs. 1 and 3) can be interpreted as showing precisely this function. The low rates of return to betting longshots are thus rationalized by the assertion that bettors tend to bet as though horse's "tiny" probabilities are

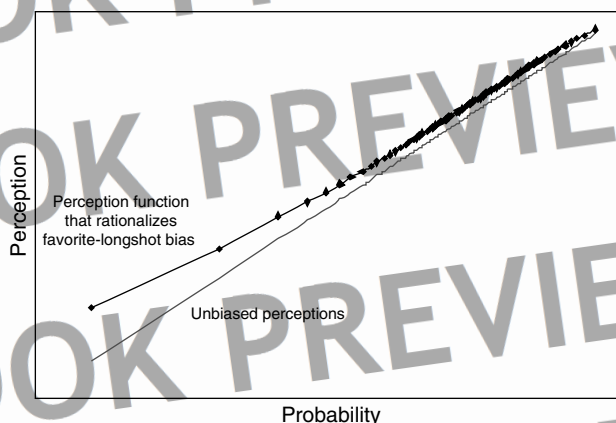


FIGURE 4 Perception function implied by betting odds.

actually “small” probabilities. Beyond this, the specific shape of the declining rates of return identifies the decision weights at each point. The overall form of the weighting function of a gamble with probability of winning g and odds G is

$$\pi(g) = (1 + G)^{-1}.$$

Note that as in Figure 3, by construction the function derived here explains all of the variation in Figure 1.

While the assumption of risk-neutrality is clearly too stark, as long as bettors gamble small proportions of their wealth, the approximation error is second-order. For instance, assuming log utility, a bettor is indifferent over betting $\varepsilon\%$ of their wealth on the favorite or the longshot if: $\pi(f) \log(w + F) = \pi(l) \log(w + L)$, which under the standard approximation implies that: $\pi(f)F \approx \pi(l)L$.

While we have presented the perceptions model as a highly stylized representative agent model, a range of somewhat richer alternatives that have been proposed in the literature yield similar implications. For instance, Hurley and McDonough consider a simple form of heterogeneity in which naifs bet randomly (for instance based on the horse’s name), and rational agents bet according to the true probabilities. Based on this, the rational agents will partly offset the longshot bias caused by the naifs, but the track take means that it is not optimal to fully offset this bias. Consequently, the market as a whole prices gambles as if it overestimated the probability of longshots winning, while only a fraction of market participants have biased beliefs. Since we do not have data on individual gamblers it is impossible to identify whether there are separate groups of bettors. Thus, maintaining the representative agent assumption, this is indistinguishable from the perceptions model.

Related models by Ali (1977) and Manski (2004) posit that the betting population has—on average—unbiased beliefs, but that there is heterogeneity around these beliefs. Thus, those with beliefs that the favorite is particularly likely to win will bet the favorite, while others will bet the longshot. The driving force in the model is that longshot bettors back their horse to win more than favorite-bettors, which causes the odds of the longshot to become compressed relative to the true probabilities. Thus position-weighted average perceptions exhibit a favorite-longshot bias. Finally, Henery (1985) and Williams and Patton (1997) argue that bettors discount a constant proportion of the gambles in which they backed a loser, possibly due to a self-serving bias in which they argue that conditions were atypical. Because longshot bettors lose more often, this discount makes betting longshots relatively more attractive. In this case, it is entirely accurate to describe these bettors as motivated by misperceptions of probabilities.

By construction, both of these models explain all of the variation in the data from win betting (betting on which horse will cross the finish line first). These data provide no traction in establishing which of the two models are better explanations, as both are tautological in this space. The only way to argue that one model is “better” than another is to make strong parametric assumptions. This is what all previous authors have done. Our innovation is to consider data from combinatoric markets as well, which allows us to make minimal assumptions and still discriminate between the two models.

We start by examining in detail how our method works for an exacta bet (picking the first two horses in order), before proceeding to other combinatoric bets: the quinella and trifecta. As before, we price these bets by considering indifference conditions. Pricing an exacta requires data on the perceived likelihood of the pick for first actually winning, and conditional on that, the likelihood of the pick for second coming second, as well as the bettor's utility function. As such, a bettor will be indifferent between betting on an exacta with horses A then B paying odds of $O_{AB}/1$ and not betting (which yields no change in wealth, and hence a utility of one), if:

Preferences Model

(Risk-lover, unbiased expectations)

$$p_A p_{B|A} U(O_{AB}) = 1$$

$$\text{Noting that } p = \frac{1}{U(O)},$$

$$O_{AB} = U^{-1}(U(O_A)U(O_{B|A}))$$

Perceptions Model

(Biased expectations, risk-neutral)

$$\pi(p_A)\pi(p_{B|A})O_{AB} - (1 - \pi(p_A)\pi(p_{B|A})) = 0$$

$$\text{Noting } \pi(p) = \frac{1}{O + 1},$$

$$O_{AB} = (O_A + 1)(O_{B|A} + 1) - 1$$

Under the preferences model, we estimate the utility function based on the pricing of win bets (in Figure 3), and we then invert this to compute unbiased win probabilities from the betting odds. While we do not have data on $O_{B|A}$ —the odds of B coming second, conditional on A coming first—we can infer this conditional probability from win odds, by assuming conditional independence. That is, $\pi(B|A) = \pi(B)/(1 - \pi(A))$ ²³ under the perceptions model. This assumption is akin to thinking about the race for second as a “race within the race.”²⁴ With this assumption in hand, we can explore how either the utility function in Figure 3 or the decision weights in Figure 4 yield different implications for pricing of exactas.²⁵

Figure 5a, b shows the pricing functions implied by these two models; the x - and y -axes show the odds on the first and second place horse, and the z -axis shows the equilibrium exacta odds implied by each model. Appendix A carries out similar calculations for quinellas and trifectas in more detail.

²³The calculation for the preferences model takes the same form, replacing p for $\pi(p)$ in the equation.

²⁴While relying on conditional independence (as per Harville, 1973) is standard in the literature, Hausch et al. (1981) find that Harville produces estimates for second and third place horses that are too high but that the biases cancel such that the place and show probabilities are not biased much. We showed in Section 6 that we can relax this assumption and our major results remain unchanged. See also Appendix A for more details on how these pricing functions are calculated.

²⁵There remains one minor issue: as Figure 4 shows, horses never win as often as suggested by their win odds. In mapping these empirical probabilities to the $\pi(p)$ function, we have two choices: (1) apply this mapping literally, even though it implies that agents systematically overestimate all probabilities, or (2) adjust the odds for the track take by dividing the odds-implied probabilities by their sum, within each race. This latter possibility can be rationalized by assuming that bettors gain a small utility from having a horse to cheer in each race as in Conlisk (1993). This small utility determines bet sizes (it equals bet size times the track take), and can both rationalize why risk-neutral or even risk-averse bettors gamble, and also results in a $\pi(p)$ function that is not systematically an overestimate for all p . We choose the latter, although our results are qualitatively similar either way.

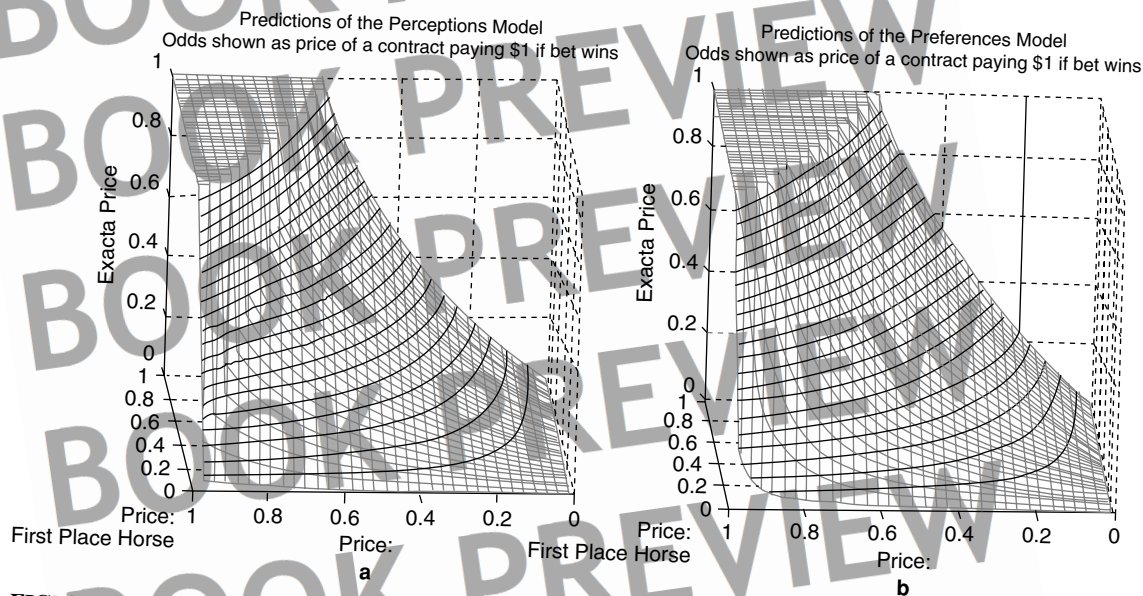


FIGURE 5 a: Pricing functions for exactas for the perceptions model. b: Pricing functions for exactas for the preferences model.

6. USING COMBINATORIC MARKETS TO TEST THE MODELS

This chapter proposes a test capable of differentiating between the perceptions and preferences models. A secondary aim is to bring a new dataset to the question, documenting the stylized facts from all 5,600,000 horse race starts in the U.S. between 1991 and 2001. These data are an order of magnitude larger than any other dataset previously examined, and allow us to be quite precise in establishing the relevant stylized facts.

Our chapter is most closely related to the papers of Julien and Salanié (2000) and Gandhi (2007) which both attempt to sort out the extent to which the favorite-longshot bias reflects risk-love versus misperceptions in win pool betting data.²⁶ Julien and

²⁶There are other papers that try to distinguish between different explanations of the favorite-longshot bias. Coleman (2004) concludes on the basis of Occam's razor that the bias is due to the interaction of informed and uninformed bettors. Golec and Tamarkin (1995) find that overconfidence (i.e., misweighting of probabilities) fits the data in the win pool better than risk-love, but cannot reject the hypothesis of risk-love. Plott et al. (2003) find that a rational expectations model is better than one centered on private information in explaining the favorite-longshot bias. Sobel and Raines (2003) test between risk-love and informational models and come down in favor of the information model. Their models, however, make functional form assumptions. See Sobel and Ryan (this volume) for an update.

Salanié (2000) find evidence consistent with the misperceptions model (which they label as nonexpected utility theories). Gandhi (2007) shows that it is possible to identify the preferences of bettors without the representative bettor assumption. He finds the preferences identified in this way are more consistent with expected utility theory (rather than risk-love, since some bettors are estimated to be risk-averse) rather than misperceptions.

Our innovation is to argue that combinatoric bets can be used to derive testable restrictions that differentiate these theories. The question then is whether the specific forms of preferences and perceptions that rationalize the favorite-longshot bias in the win pool can also explain exacta (and quinella and trifecta) pricing; and further which gives a fuller accounting of the variation in that data. By expanding the choice set under consideration (to correspond with the bettor's actual choice set!), we have the opportunity to use the relevant theory to derive testable restrictions of each theory. While authors such as Asch and Quandt (1987, 1988) and Ali (1979) have tested the efficiency of these exotic betting markets, we are the first to use these prices to distinguish between competing theories of possible market inefficiency.

Our data contains every horse race run in the U.S. between 1992 and 2002. These data are official jockey club data, and hence are the most precise data available. Data of this nature are prohibitively expensive, and so we are extremely grateful to David Siegel of TrackMaster Inc. for providing us access. While we have a vast database on every horse and every race, jockey, owner, trainer, sire and dam, we will only exploit the betting data, and whether or not a horse won the race. Appendix B further describes the data.

We summarize our data in Figures 6 and 7. We group horses into 74 separate groups; this allows for fine distinctions to be made, but also ensures that we have sufficient observations to make precise estimates of the rates of return to betting horses in each group. The x-axis of both figures inverts the odds so as to work out an implied probability, or, alternatively, how often a horse at those odds must win in order for the bettor to break even. Data are graphed on a log scale so as to better show the relevant range of the data. Figure 1 shows the actual rate of return to betting on horses in each category. The average rate of return for betting favorites is about -10% , while horses at a midrange of $1/1$ to $20/1$ yield a rate of return of -20% , and real longshots—horses at $100/1$ or more—are much more expensive to bet on, costing 60 cents per dollar bet. This is, once again, the favorite-longshot bias. Figure 6 also shows the same pattern for the 54,013 races for which the jockey club recorded payoffs to exacta, quinella, and trifecta bets. Given that much of our analysis will focus on this smaller sample, it is reassuring to see a similar pattern of returns.

The literature has suggested two other empirical regularities that we can explore. Ziemba and Hausch (1986) and Hausch and Ziemba (1995) present data suggesting that there are positive rates of return to betting extreme favorites, perhaps suggesting limits to arbitrage. However, as the confidence intervals in Figures 1, 6, and 7 show, there is substantially greater statistical uncertainty about returns on extreme favorites and longshots, and in none of these datasets is there a positive effect of betting on horses less than $0.2/1$.

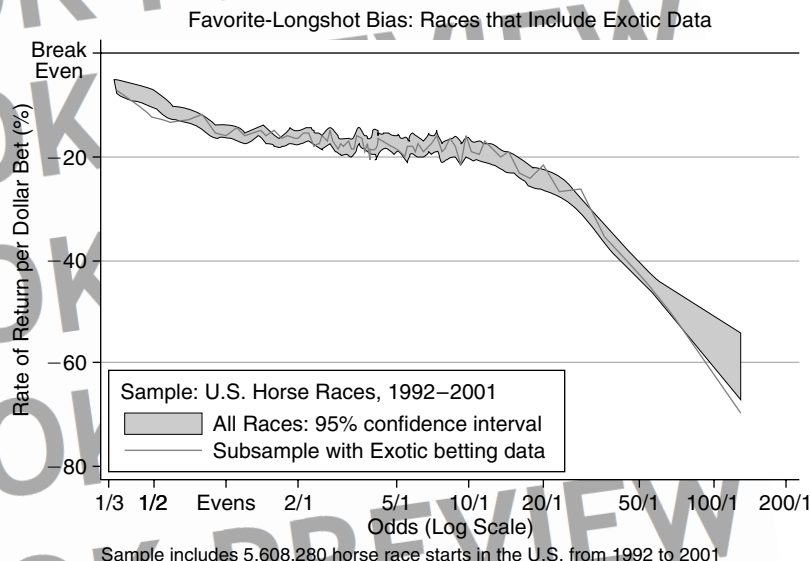


FIGURE 6 The favorite-longshot bias in different subsets of data.

Second, McGlothlin (1956) and Ali (1977) argue that the rate of return to betting moderate longshots falls in the last race of the day. While this conclusion was based on a small sample, Thaler and Ziemba (1988), and Camerer (2001) interpret this as consistent with loss aversion: most bettors are losing at the end of the day, and the “get out of jail” race provides them with a chance to leave the racetrack even for the day. Thus, bettors underbet the favorite even more than usual, and overbet horses at odds that would eliminate their losses—around 8/1. Figure 7 replicates Figure 1, but separates out the last race of the day from earlier races. As should be clear, there is no statistically discernible difference between the last race and earlier races.²⁷ If there were evidence of loss aversion in McGlothlin or Ali’s data, it no longer appears evident in more recent data, even as the favorite-longshot bias has persisted.²⁸

We now turn to the major contribution of this paper—testing the different models on combinatoric bets. Our empirical approach is simply to estimate which of the pricing functions shown in Figure 5 better fits the data. In Table 1, we convert the odds into the price of a contingent contract that pays \$1 if the chosen exacta wins (that is, $\text{Price} = (\text{Odds} + 1)^{-1}$), and then regress the price of the winning exacta against the prices implied by preference model (column 1), the perceptions model (column 2) and then put them both in the regression (column 3). Comparing columns 1 and 2,

²⁷Note that the last race sample is about one-ninth the size of the all races sample. As such, the standard errors on the estimates for the last race are approximately three times as wide as those for all races.

²⁸There are reasons other than statistical uncertainty why positive returns may have dissipated. Since these studies, there have been many changes in the market structure, such as a larger role for rebates and more off-track betting, which may have changed the incentives of bettors.

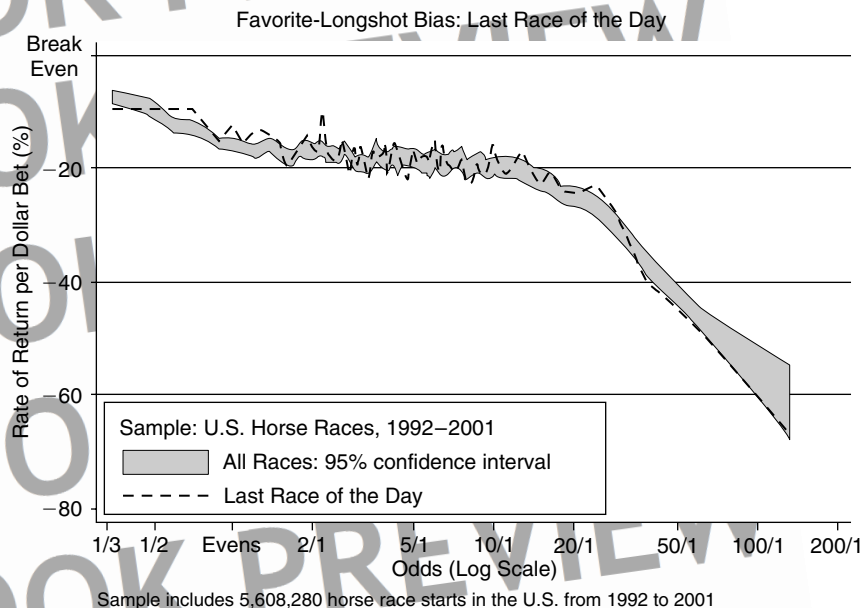


FIGURE 7 The favorite-longshot bias in the last race vs. all other races.

the explanatory power of the perceptions model is substantially greater, and the regressions in column 3 confirm this, showing that when the regression is allowed to choose optimal weights on the implications of each theory, it strongly prefers the perceptions model. When we weight by the size of the relevant betting pool, we obtain qualitatively similar results.

Panels B and C repeat this analysis, but this time extending our test to see which model can better explain the pricing of quinella and trifecta bets. While the specific formulae to generate the implications of each model differs, the intuition is precisely the same. Appendix A shows the relevant pricing functions in more detail.

Our results are clear: the perceptions model gives us more traction in explaining the pricing of combinatoric bets. First, observe that in all three betting pools, the coefficient on the perceptions model is closer to unity than the preferences model. Also, in all three pools, the perceptions model explains a significantly larger part of the variation in the data than the preferences model does. Although the differences in R^2 are small numerically, given the size of the dataset, the perceptions model is actually quite an improvement in explanatory power. Finally, the bake-off prefers the perceptions model. Given how similar the predictions of each model are, this is actually quite a striking result.²⁹

We have also re-run these regressions a number of other ways to test for robustness, and our conclusions are unaltered by whether or not we include constant terms in the

²⁹This is normally called a horse race, but to avoid confusion, we call it a bake-off.

TABLE 1 Testing the Fit of Each Model

Specification	(1) Preferences	(2) Perceptions	(3) Bake-Off
Panel A: Exacta Bets ($n = 52,168$)			
Preferences-based model predictions (risk-love)	0.7993 (0.0016)		0.1791 (0.0102)
Perception-based model predictions (misperceive probabilities)		0.9172 (.0017)	0.7159 (0.0105)
Constant	0.0076 (0.0001)	.0056 (.0001)	0.0058 (0.0001)
Adjusted R^2	0.8358	0.8482	0.8492
Panel B: Quinella Bets ($n = 52,778$)			
Preferences-based model predictions (risk-love)	0.8124 (0.0014)		-0.216 (0.0123)
Perception-based model predictions (misperceive probabilities)		0.8826 (0.0014)	1.1137 (0.0132)
Constant	0.0129 (0.0002)	0.0103 (0.0002)	0.0101 (0.0002)
Adjusted R^2	0.8602	0.8761	0.8768
Panel C: Trifecta Bets ($n = 34,313$)			
Preferences-based model predictions (risk-love)	0.6410 (0.0021)		0.1391 (0.0071)
Perception-based model predictions (misperceive probabilities)		0.8257 (0.0025)	0.6575 (0.009)
Constant	0.0027 (0.0001)	0.0030 (0.0001)	0.0027 (0.0001)
Adjusted R^2	0.7247	0.7593	0.7619

NOTE: Dependent variable: price of a contract paying \$1 if the combinatoric bet wins.

regressions, whether or not we weight by the size of the betting pool, whether we drop observations where the models imply very long odds, or whether or not we adjust the perceptions model in the manner described in footnote 37. The results are also robust different functional forms, including the natural log price of a \$1 claim, the odds, or log-odds.

6.1. Testing Conditional Independence

The assumption of conditional independence was a key assumption simplifying the analysis in the previous section. We now turn to both testing this assumption, and then deriving two further tests that can distinguish between our families of models even if conditional independence fails.

Recall that the only unobservable in the previous section was the probability of horse B finishing second, conditional on horse A winning. Conditional independence posits a specific functional form for this relationship, based on the chances of horses A and B winning the race. However, there may be good reason to doubt this assumption. For instance, if a heavily favored horse does not win a race, this may reflect the fact that it was injured during the race, which then implies that it is very unlikely to come second. That is, the odds may provide useful guidance on the probability of winning, but conditional on not winning, may be a poor guide for the race to come second.

We can directly test the assumption of conditional independence by asking whether the Harville formulae provide a sufficient statistic for whether a horse will come second. We compute the Harville statistic as $p_B(1 - p_A)^{-1}$, where p_A and p_B reflect the probability that horses at odds of $A/1$ and $B/1$, respectively, win their races. We then run a linear probability model where the dependent variable is an indicator variable for whether horse B runs second.³⁰ Beyond the Harville predictor, we both loosen up the functional form of the specified relationship, adding linear and quadratic terms in both p_A and p_B to the regression. In the third column we also append other factors to the regression, such as the probability that a horse at the odds of the favorite wins the race, and a Herfindahl index of the probabilities of each of the horses in the race. Table 2 shows our results.

Two main conclusions can be drawn from Table 2. First, the Harville formulae are extremely useful predictors of the probability of a horse finishing second. To provide a yardstick for thinking about the explanatory power, note that this is about four-fifths as high as the R^2 one gets when trying to explain which horse wins the race, using the predictions in Figure 5. Second, the Harville formula is not a sufficient statistic; our other independent variables raised the explanatory power of the regression somewhat.

An immediate concern is that the Harville approximation error might be driving our main results in Table 1. However, there is an easy solution, which is to simply to calculate $p_{B|A}$ directly from the dataset. We implement this procedure in Table 3.

In order to do this, we had to find the true probability that an exacta with odds $A/1$ and $B/1$ actually wins using the same odds ranges to organize the bins as before. This means that there are $74 \times 74 = 5,476$ combinations for which we need to calculate the probabilities, so on average there are $\sim 1,000$ observations per combination. However, certain combinations are much more likely, and others, such as having two horses at $0.1/1$ odds in the same race, are impossible. The range of the number of observations in these cells goes from 1 to 441,930. This affects the precision to which we can estimate $p_{B|A}$, so when we run regressions, we weight by the inverse of the standard error to which we estimated $p_{A|B}$. The standard error of each cell is given by the familiar formula $\text{sqrt}((p_{B|A}^*(1 - p_{B|A}))/n)$. This is just standard WLS to account for independent heteroskedastic observations.

³⁰Probits yield similar results.

TABLE 2 Conditional Independence of Second Place Finish

Specification ($n = 4,121,025$)	(1)	(2)	(3)
Prediction from conditional independence (Harville formula)	0.7881 (0.0013)	0.8794 (0.0083)	0.8265 (0.0085)
Predicted probability of first place finish from odds		0.1484 (0.0123)	0.2083 (0.0127)
Predicted probability of first place finish squared		-0.7371 (0.0116)	-0.6699 (0.0133)
Predicted probability of first place horse		0.0165 (0.0042)	0.0377 (0.0051)
Predicted probability of first place horse squared		-0.0756 (0.0075)	-0.1377 (0.0099)
Horse was favorite			-0.0173 (0.0008)
Number of horses			-0.0035 (0.0110)
Max probability of horse in race			-0.0312 (0.0110)
Max probability of horse in race squared			0.1392 (0.0170)
Herfindahl index			0.0770 (0.0103)
Constant	0.0293 (0.0003)	0.0172 (0.0006)	-0.0190 (0.0009)
Adjusted R^2	0.0767	0.0778	0.0782

NOTE: Dependent variable: indicator for whether a horse came in second.

The results in Table 3 are consistent with those in Table 1. For each of the three exotic bets, the perceptions model has greater explanatory power than the risk-love model, and, in the bake-off, is strongly preferred.

6.2. Relaxing Conditional Independence Further

Our final test of the two models is even more non-parametric, and relies only on the relative pricing of exacta and quinella bets.³¹ As before, we derive predictions from each model and test which better explains the observed data. The advantage of focusing only on comparisons between the first two horses is that these tests are—by

³¹Note that these tests are distinct from the work by authors such as Asch and Quandt (1987, 1988) and Ali (1979), who test whether quinella pricing is arbitrage-linked to exacta pricing. Instead, we ask whether the same model that explains pricing of win bets can *jointly* explain the pricing of exacta and quinella bets.

TABLE 3 Testing the Fit of Each Model without Conditional Independence, Correcting for Heteroskedasticity

Specification	(1) Preferences	(2) Perceptions	(3) Bake-Off
Panel A: Exacta Bets ($n = 52,322$)			
Preferences-based model predictions (risk-love)	1.0078 (0.0023)		−0.0506 (0.0424)
Perception-based model predictions (misperceive probabilities)		1.0492 (0.0024)	1.1017 (0.0441)
Constant	−0.0015 (0.0001)	0.0013 (0.0001)	0.0015 (0.0001)
Adjusted R^2	0.7859	0.7884	0.7884
Panel B: Quinella Bets ($n = 52,774$)			
Preferences-based model predictions (risk-love)	1.002 (0.0017)		0.263 (0.037)
Perception-based model predictions (misperceive probabilities)		0.9191 (0.0015)	0.6783 (0.0339)
Constant	0.0005 (0.0001)	0.0028 (0.0001)	0.0022 (0.0001)
Adjusted R^2	0.8688	0.8697	0.8698

NOTE: Dependent variable: price of a contract paying \$1 if the combinatoric bet wins.

construction—conditionally independent of the characteristics of all other horses in the race.

Before deriving the predictions, we will explore the setup and basic intuition. Given two horses with odds $F/1$ (which we will also call the favorite) and $L/1$ (the longshot) and $F < L$, each model makes a unique prediction about the odds of the exacta and quinella. However, each observation with given odds of the first and second place horse does not have the same exacta and quinella odds. We can use the quinella (exacta) price and our models to make a better prediction about the price of the exacta (quinella). This section uses all four of these odds in concert to make predictions about the probability of the favorite coming first (given that F and L came first and second in some order).

For the derivation below, note that the exacta F-L occurs with probability $p_F \cdot p_{L|F}$ and the L-F exacta occurs with probability $p_L \cdot p_{F|L}$, and hence the corresponding quinella occurs with probability $p_F \cdot p_{L|F} + p_L \cdot p_{F|L}$ (where p_F is the probability of F winning, and $p_{L|F}$ denotes the probability of F coming second given that L won the race). Also, the L-F exacta has odds of $E_{LF}/1$, and the L-F quinella (which is the same as the F-L quinella) has odds $Q/1$.

Consequently, these two models have different implications for how frequently we expect to observe the L-F exacta winning, relative to the F-L exacta winning. Further, these comparisons yield distinct predictions even *within* any set of apparently similar

Preferences Model

(Risk-lover, unbiased expectations)

Exacta: $p_{LPF|L}U(E_{LF}) = 1$

$$\Rightarrow p_{F|L} = \frac{U(L)}{U(E_{LF})}$$

Quinella:

$$[p_{LPF|L} + p_{FPL|F}]U(Q) = 1$$

$$\Rightarrow p_{L|F} = U(F) \left(\frac{1}{U(Q)} - \frac{1}{U(E_{LF})} \right)$$

Hence

$$\frac{p_{LPF|L}}{p_{LPF|L} + p_{FPL|F}} = \frac{U(Q)}{U(E_{LF})}$$

Perceptions Model

(Biased expectations, risk-neutral)

Exacta: $\pi(p_L)\pi(p_{F|L})(E_{LF} + 1) = 1$

$$\pi(p_{F|L}) = \frac{L+1}{E_{LF}+1} \Rightarrow p_{F|L} = \pi^{-1} \left(\frac{L+1}{E_{LF}+1} \right)$$

Quinella:

$$[\pi(p_L)\pi(p_{F|L}) + (p_F)\pi(p_{L|F})](Q+1) = 1$$

$$\pi(p_{L|F}) = (F+1) \left(\frac{1}{Q+1} - \frac{1}{E_{LF}+1} \right) \Rightarrow p_{F|L} = \pi^{-1} \left(\frac{(F+1)(E_{LF}-Q)}{(E_{LF}+1)(Q+1)} \right)$$

$$\begin{aligned} & \frac{p_{LPF|L}}{p_{LPF|L} + p_{FPL|F}} = \frac{\pi^{-1} \left(\frac{L+1}{E_{LF}+1} \right)}{\pi^{-1} \left(\frac{L+1}{E_{LF}+1} \right) + \pi^{-1} \left(\frac{1}{F+1} \right) \pi^{-1} \left(\frac{(F+1)(E_{LF}-Q)}{(E_{LF}+1)(Q+1)} \right)} \end{aligned}$$

aces (those whose first two finishers are at $L/1$ and $F/1$ with the quinella paying $Q/1$).

Thus, we can include a full set of fixed effects for L , F , Q and their interactions in our statistical tests.³² The residual after partialing out these fixed effects is the marginal effect of including quinella and exacta odds in our predictions. Since much of the variation in the quinella and exacta payoffs comes from the first place odds, which we have already used in our tests, we are here interested in the improvement in prediction when using the bettor's full choice set.

Figure 8 compares these predicted values with their actual values, where each are reported as deviations from their baseline (or L - F - Q cell mean). It is important to note that by focusing only upon comparisons between the first two horses, we eliminate parametric assumptions about "the race for second place" (conditional independence).

Clearly the preference-based model does a much better job in predicting the winning exacta, given horses that finish in the top two positions (and their odds). Indeed, the predictions of the perceptions-model are robustly positively correlated with actual outcomes ($\rho = 0.12$; $n = 50,216$), while the preferences model yields predictions that are perversely negatively correlated with actual outcomes ($\rho = -0.16$). The results of a fixed effects regression on these predictions can be found in Table 4.

Table 4 requires some explanation. The first thing to note is that if we run this regression using OLS without fixed effects, the results look much the same as before. The perceptions model has more explanatory power, and a coefficient close to one in the bake-off. The preferences model has a coefficient near zero. However, we are interested here in the incremental effect of adding the quinella and exacta odds, hence we have used a fixed effects model.

³²Because the odds L , F , and Q are actually continuous variables, we include 100 fixed effects for each, one for each percentile of the distribution of each.

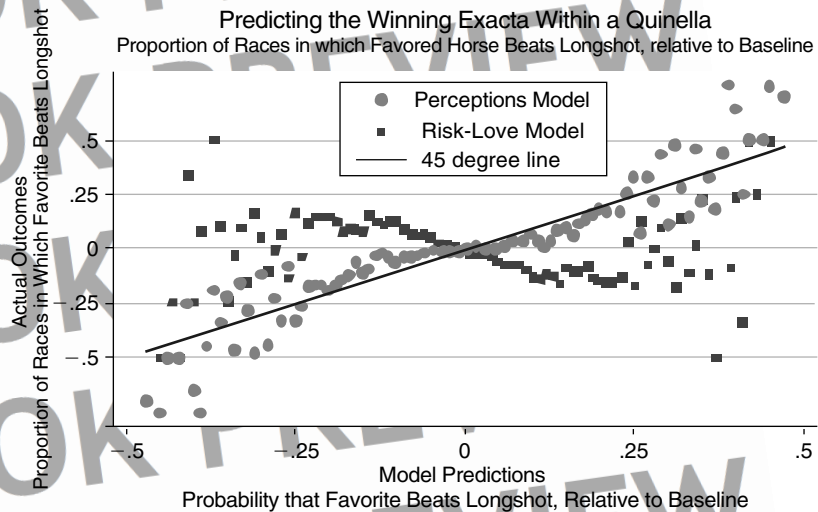


Chart shows model predictions and outcomes relative to a fixed-effect regression baseline. Baseline controls for saturated dummies for: (a) The odds of the favored horse; (b) The odds of the longshot (c) The odds of the quinella; and (d) A full set of interactions of all three sets of dummy variables

FIGURE 8 Predictions including quinella and exacta odds—demeaned.

NOTE: For each race we compute the likelihood of an F-L exacta given an F-L quinella. These predictions are made under the two models outlined in the text, using as inputs data on the odds of each horse ($F/1$, $L/1$), their quinella ($Q/1$) and the winning exacta ($E/1$). We then compute the mean predictions and outcomes for all races within the same $\{F, L, Q\}$ cell. Subtracting these means yields the model predictions and outcomes relative to these fixed effects. For the purposes of the plot, we round these residuals to the nearest percentage point (shown on the x-axis), and the y-axis shows actual win percentages for races in each bucket.

TABLE 4 Fixed Effects Regression of Models Using Quinella and Exacta Odds

Specification	(1) Preferences	(2) Perceptions	(3) Bake-Off
Panel A: Exacta Bets ($n = 50,216$)			
Preferences-based model predictions (risk-love)	−0.7582 (0.0126)		−8.2994 (0.0222)
Perception-based model predictions (misperceive probabilities)		0.4909 (0.0134)	8.5581 (0.0212)
Constant	-3.52×10^{-9} (0.0015)	-3.64×10^{-9} (0.0015)	-5.69×10^{-9} (0.6103)
Adjusted R^2	0.0347	0.0132	0.0352

NOTE: Dependent variable—dummy variable for whether favorite came first.

The perceptions model has a moderate coefficient when presented alone in the regression. The preferences-based model is actually systematically biased the wrong way. The bake-off provides us with no useful information, except that the models make, in general, highly correlated predictions.

A concrete example provides further intuition of why the models make different predictions and why the perceptions model fares better. Consider all races in which F was at 4/1, and L was at 9/1, while the quinella was at 20/1 (which is close to actuarially fair). If the F-L exacta is at odds of 40/1, then given the utility function shown in Figure 3, a risk-loving bettor is willing to bet this high-risk F-L exacta, even if its chances of winning are quite low. Given the odds of the quinella, and the fact that the probability of the F-L exacta is lower than implied by the odds (because of the extra risk premium associated with it), this implies that the risk-loving bettor must believe that L-F is (relatively) quite likely. This turns out to be a poor prediction, and indeed empirically, F-L is typically more likely than L-F as we would expect. This explains the anomalous negative performance of the risk-love model, particularly around the region where either exacta is about as likely as the other. By contrast, the perceptions model performs well because the relative misperceptions of bettors are approximately constant for most horses, leading them to misprice the F-L exacta and quinella in roughly similar ways, which yields well-calibrated estimates of the relative likelihood of the L-F exacta.

The intuition above leads us to wonder about the usefulness of an unbiased risk-neutral model. This model explains the data in this test as well as the preferences model, even in the fixed effects form. However, an unbiased risk-neutral model cannot account for the fact that people gamble or the favorite-longshot bias in the first place, so even though it performs well in this test, we can eliminate it from consideration as an explanation of the behavior we observe at the track.

These tests imply that while a preference-based model can be constructed to account for the pricing of win bets, it yields inaccurate implications for the pricing of exacta and quinella bets. Moreover, its predictions of the relative frequency of exacta outcomes are on average often negatively related to actual outcomes. By contrast, the perceptions-based model is consistent with the pricing of exacta, quinella, and trifecta betting, and as this section showed, also consistent with the relative pricing of exacta and quinella bets. Moreover, these results are robust to a range of different approaches to testing the theory.

7. CONCLUSION

This chapter summarizes all of the theories of the favorite-longshot bias in gambling. However, the term bias is somewhat misleading here. That the rate of return to betting on horses at long odds is much lower than the average return to betting on favorites simply falsifies a model that bettors maximize a function that is linear in probabilities and linear in payoffs.

By examining these theories and placing them within the larger economics literature, we are able to divide them into three categories. The first are theories that involve preferences, the second are theories that involve the weighting of probabilities of a bet winning, or misperceptions. The third is observationally equivalent to the second, and involves different classes of bettors that have different information sets. For

compactness, we referred to the former as explaining the data with preferences, while we refer to the latter as explaining the data with perceptions. Neither label is particularly accurate.

Employing a new dataset that is an order of magnitude larger than anything considered to date, we document a set of stylized facts from the win market. As with other authors, we note a substantial favorite-longshot bias. However, two features of the win market reported by other authors are notably absent. Namely, we fail to find positive expected payoffs on extreme favorites. We are also unable to replicate results that posit that the risk attitudes (or misperceptions) are significantly different in the last race of the day. These features have either disappeared since the time other authors originally wrote about them, or more likely, were not statistically significant in the first place.

This chapter also contributes to the literature that tries to determine whether neo-classical or behavioral theories better explain aggregate behavior in the field. Our innovation is to consider combinatoric bets, such as exactas, quinellas, and trifectas. Because the underlying risk, or set of beliefs (depending on the relevant theory) is traded in both the win and combinatoric betting markets, we can derive testable implications of both sets of theories. Our results suggest that the perception-based theories yield greater explanatory power than the preference-based theories. Indeed, while both are significant explanations of the pricing of exotic bets, the preference-based model adds little explanatory power beyond the perceptions model. These results are robust to a range of alternative approaches.

Of course, neither model is literally correct, and thus there is plenty of scope for related work on these issues. Thus, rather than stating a strong conclusion, we would simply argue that our results suggest it seems likely that nonexpected utility theories are the more promising candidate for explaining racetrack betting, and perhaps this is cause for optimism that they may also explain anomalies in other domains of decision-making under uncertainty.

References

- Ali, M. M. 1977. Probability and Utility Estimates for Racetrack Bettors, *The Journal of Political Economy* 85, 803–815.
- Ali, M. M. 1979. Some Evidence of Efficiency in a Speculative Market, *Econometrica* 47, 387–392.
- Asch, P., and R. E. Quandt. 1987. Efficiency and Profitability in Exotic Bets, *Economica* 54, 289–298.
- Asch, P., and R. E. Quandt. 1988. Betting Bias in “Exotic” Bets, *Economics Letters* 28, 215–219.
- Asch, P., and R. E. Quandt. 1990. Risk Love, *Journal of Economic Education* 4, 422–426.
- Asch, P., B. G. Malkiel, and R. E. Quandt. 1982. Racetrack Betting and Informed Behavior, *Journal of Financial Economics* 10, 187–194.
- Bassett, G. W., Jr. 1981. Point Spreads vs. Odds, *The Journal of Political Economy* 89, 752–768.
- Bird, R., M. McCrae, and J. Beggs. 1987. Are Gamblers Really Risk Takers? *Australian Economic Papers* 26, 237–253.
- Bradley, I. 2003. The Representative Bettor, Bet Size and Prospect Theory, *Economics Letters* 78, 409–413.
- Brown, W. O., and R. D. Sauer. 1993. Fundamentals or Noise? Evidence from the Professional Basketball Betting Market, *The Journal of Finance* 48, 1193–1209.
- Brunk, G. G. 1981. A Test of the Friedman-Savage Gambling Model, *The Quarterly Journal of Economics* 96, 341–348.

- Busche, K., and C. D. Hall. 1988. An Exception to the Risk Preference Anomaly, *The Journal of Business* 61, 337–346.
- Cain, M., D. Law, and D. Peel. 2000. The Favorite-Longshot Bias and Market Efficiency in UK Football Betting, *Scottish Journal of Political Economy* 47, 25–36.
- Camerer, C. F. 2001. Prospect Theory in the Wild: Evidence from the Field, in Daniel Kahneman and Amos Tversky (eds.), *Choices, Values and Frames*. Cambridge: Cambridge University Press, Cambridge, UK.
- Coleman, L. 2004. New Light on the Longshot Bias, *Applied Economics* 36, 315–326.
- Conlisk, J. 1993. The Utility of Gambling, *Journal of Risk and Uncertainty* 6(3), 255–275.
- Crafts, N. F. R. 1985. Some Evidence of Insider Knowledge in Horse Race Betting in Britain, *Economica*, 207, 295–304.
- Dowie, J. 1976. On the Efficiency and Equity of Betting Markets, *Economica* 52, 295–304.
- Fama, E. F. 1970. Efficient Capital Markets: A Review of Theory and Empirical Work, *The Journal of Finance* 25, 383–417.
- Figlewski, S. 1979. Subjective Information and Market Efficiency in a Betting Market, *The Journal of Political Economy* 87, 75–88.
- Fishburn, P. C. 1980. A Simple Model for the Utility of Gambling, *Psychometrika* 45, 435–448.
- Friedman, M., and L. J. Savage. 1948. The Expected-Utility Hypothesis and the Measurability of Utility, *The Journal of Political Economy* 56, 279–304.
- Gabriel, P. E., and J. R. Marsden. 1990. An Examination of Market Efficiency in British Racetrack Betting, *The Journal of Political Economy* 98, 874–885.
- Gabriel, P. E., and J. R. Marsden. 1991. An Examination of Market Efficiency in British Racetrack Betting: Errata and Corrections, *The Journal of Political Economy* 99, 657–659.
- Gander, J. M., W. H. Dare, C. R. Brown, and R. A. Zuber. 1998. Informed Traders and Price Variations in the Betting Market for Professional Basketball Games, *The Journal of Finance* 53, 385–401.
- Gander, J. M., R. A. Zuber, and R. S. Johnson. 2001. Searching for the Favorite-Longshot Bias Down Under: An Examination of the New Zealand Pari-Mutuel Betting Market, *Applied Economics* 33, 1621–1629.
- Gander, J., R. Zuber, T. O'Brien, and B. Russo. 1988. Testing Rationality in the Point Spread Betting Market, *The Journal of Finance* 43, 995–1008.
- Gandhi, A. K. 2007. Rational Expectations at the Racetrack: Testing Expected Utility Using Prediction Market Prices. Mimeo, University of Wisconsin-Madison.
- Golec, J., and M. Tamarkin. 1992. The Degree of Inefficiency in the Football Betting Market, *Journal of Financial Economics* 30, 311–323.
- Golec, J., and M. Tamarkin. 1995. Do Bettors Prefer Longshots Because They Are Risk-Lovers, or Are They Just Overconfident? *Journal of Risk and Uncertainty* 11, 51–64.
- Golec, J., and M. Tamarkin. 1998. Bettors Love Skewness, Not Risk, at the Horse Track, *The Journal of Political Economy* 106, 205–225.
- Gray, P. K., and S. F. Gray. 1997. Testing Market Efficiency: Evidence from the NFL Sports Betting Market, *The Journal of Finance* 52, 1725–1737.
- Griffith, R. M. 1949. Odds Adjustment by American Horse Race Bettors, *American Journal of Psychology* 62, 290–294.
- Hamid, S. S., A. J. Prakash, and M. W. Smyser. 1996. Marginal Risk Aversion and Preferences in a Betting Market, *Applied Economics* 28, 371–376.
- Harville, D. A. 1973. Assigning Probabilities to the Outcomes of Multi-Entry Competitions, *Journal of the American Statistical Association* 68, 312–318.
- Hausch, D. B., and W. T. Ziemba. 1995. Efficiency of Sports and Lottery Betting Markets, in R. Jarow, V. Maksimovic, and W. T. Ziemba (eds.), *Handbooks in Operations Research and Management Science*, Vol. 9: Finance, Elsevier B.V., Amsterdam, pp. 545–580.
- Hausch, D. B., W. T. Ziemba, and M. Rubinstein. 1981. Efficiency of the Market for Racetrack Betting, *Management Science* 27, 1435–1452.
- Henery, R. J. 1985. On the Average Probability of Losing Bets on Horses with Given Starting Odds, *Journal of the Royal Statistical Society, Series A (General)* 148, 342–349.
- Hurley, W., and L. McDonough. 1995. A Note on the Hayek Hypothesis and the Favorite-Longshot Bias in Pari-Mutuel Betting, *The American Economic Review* 85, 949–955.

- Hurley, W., and L. McDonough. 1996. The Favorite-Longshot Bias in Parimutuel Betting: A Clarification of the Explanation that Bettors Like to Bet Longshots, *Economics Letters* 52, 275–278.
- Jullien, B., and B. Salanié. 2000. Estimating Preferences Under Risk: The Case of Racetrack Bettors, *The Journal of Political Economy* 108, 503–530.
- Jullien, B., and B. Salanié. 2008. Empirical Evidence on the Preferences of Racetrack Bettors, in D. B. Hausch and W. T. Ziemba, *Handbook of Sports and Lottery Markets*. North-Holland, Amsterdam, pp. 27–49.
- Koessler, F., and A. Ziegelmeyer. 2002. Parimutuel Betting under Asymmetric Information, Mimeo, Univerite Louis Pasteur.
- Koessler, F., A. Ziegelmeyer, and M. H. Broihanne. 2003. The Favorite-Longshot Bias in Sequential Parimutuel Betting with Non-Expected Utility Players, *Theory and Decision* 54, 231–248.
- Losey, R. L., and J. C. Talbot, Jr. 1980. Back on Track with the Efficient Markets Hypothesis, *The Journal of Finance* 35, 1039–1043.
- Levitt, S. D. 2004. Why Are Gambling Markets Organised so Differently from Financial Markets? *The Economic Journal* 114, 223–246.
- Markowitz, H. 1952. The Utility of Wealth, *The Journal of Political Economy* 60, 151–158.
- Manski, C. F. 2006. Interpreting the Predictions of Prediction Markets, *Economics Letters* 91(3), 425–429.
- McGlothlin, W. H. 1956. Stability of Choices Among Uncertain Alternatives, *American Journal of Psychology* 69, 604–615.
- Metrick, A. 1996. March Madness? Strategic Behavior in NCAA Basketball Tournament Betting Pools, *Journal of Economic Behavior and Organisation* 30, 159–172.
- Ottaviani, M., and P. N. Sørensen. 2003. Late Informed Betting and the Favorite-Longshot Bias, Mimeo, Institute of Economics, University of Copenhagen.
- Piron, R., and L. R. Smith. 1994. Testing Risklove in an Experimental Racetrack, *Journal of Economic Behavior and Organization* 27, 465–474.
- Plott, C. R., J. Wit, and W. C. Yang. 2003. Parimutuel Betting Markets as Information Aggregation Devices: Experimental Results, *Economic Theory* 22, 311–351.
- Potters, J., and J. Wit. 1996. Bets and Bids: Favorite-Longshot Bias and Winner's Curse, Mimeo, Tilburg University.
- Quandt, R. E. 1986. Betting and Equilibrium, *The Quarterly Journal of Economics* 101, 201–208.
- Rosett, R. N. 1965. Gambling and Rationality, *The Journal of Political Economy* 73, 595–607.
- Sauer, R. D., V. Brajer, S. P. Ferris, and M. W. Marr. 1988. Hold Your Bets: Another Look at the Efficiency of Gambling Market for National Football League Games, *The Journal of Political Economy* 96, 206–213.
- Sauer, R. D. 1998. The Economics of Wagering Markets, *Journal of Economic Literature* 36, 2021–2064.
- Shin, H. S. 1991. Optimal Betting Odds Against Insider Traders, *The Economic Journal* 101, 1179–1185.
- Shin, H. S. 1992. Prices of State Contingent Claims with Insider Traders and the Favorite-Longshot Bias, *The Economic Journal* 102, 426–435.
- Shin, H. S. 1993. Measuring the Incidence of Insider Trading in a Market for State-Contingent Claims, *The Economic Journal* 103, 1141–1153.
- Snyder, W. W. 1978. Horse racing: Testing the Efficient Markets Model, *The Journal of Finance* 33, 1109–1118.
- Sobel, R. S., and S. T. Raines. 2003. An Examination of the Empirical Derivatives of the Favorite-Longshot Bias in Racetrack Betting, *Applied Economics* 35, 371–385.
- Sobel, R. S., and M. E. Ryan. 2008. Unifying the Favorite-Longshot Bias with Other Market Anomalies, in D. B. Hausch and W. T. Ziemba (eds.), *Handbook of Sports and Lottery Markets*. North-Holland, Amsterdam, pp. 137–160.
- Terell, D., and A. Farmer. 1996. Optimal Betting and Efficiency in Parimutuel Betting Markets with Information Costs, *The Economic Journal* 106, 846–868.
- Thaler, R. 1985. Mental Accounting and Consumer Choices, *Marketing Science* 4, 199–214.
- Thaler, R. H., and W. T. Ziemba. 1988. Anomalies: Parimutuel Betting Markets: Racetracks and Lotteries, *The Journal of Economic Perspectives* 2, 161–174.
- Thalheimer, R., and M. M. Ali. 1995. The Demand for Parimutuel Horse Race Wagering and Attendance, *Management Science* 41, 129–143.

- Tversky, A., and D. Kahneman. 1992. Advances in Prospect Theory: Cumulative Representation of Uncertainty, *Journal of Risk and Uncertainty* 5, 297–323.
- Vaughan Williams, L. 1999. Information Efficiency in Betting Markets: A Survey, *Bulletin of Economic Research* 51, 1–30.
- Vaughan Williams, L., and D. Paton. 1997. Why is there a Favorite-Longshot Bias in British Racetrack Betting Markets, *The Economic Journal* 440, 150–158.
- Vaughan Williams, L., and D. Paton. 1998. Why Are Some Favorite-Longshot Biases Positive and Others Negative? *Applied Economics* 30, 1505–1510.
- Weitzman, M. 1965. Utility Analysis and Group Behavior: An Empirical Study, *The Journal of Political Economy* 73, 18–26.
- Woodland, L. M., and B. M. Woodland. 1991. The Effects of Risk Aversion on Wagering: Point Spreads vs. Odds, *The Journal of Political Economy* 99, 638–653.
- Woodland, L. M., and B. M. Woodland. 1994. Market Efficiency and the Favorite-Longshot Bias: The Baseball Betting Market, *The Journal of Finance* 49, 269–279.
- Woodland, L. M., and B. M. Woodland. 2001. Market Efficiency and Profitable Wagering in the National Hockey League: Can Bettors Score on Longshots? *Southern Economic Journal* 67, 983–995.
- Woodland, L. M., and B. M. Woodland. 2003. The Reverse Favorite-Longshot Bias and Market Efficiency in Major League Baseball: An Update, *Bulletin of Economic Research* 55, 113–123.
- Ziembra, W. T., and D. B. Hausch. 1986. *Betting at the Racetrack*. Dr. Z Investments, Inc., San Luis Obispo, California.
- Zuber, R. A., J. M. Gandar, and B. D. Bowers. 1985. Beating the Spread: Testing the Efficiency of the Gambling Market for National Football League Games, *The Journal of Political Economy* 93, 800–806.

APPENDIX A: Pricing of Combinatoric Bets Using Conditional Independence

This appendix details how we come up with the predictions of our models using the assumption of conditional independence. It takes a slightly different form than the method of pricing exactas developed in the text but we hope that by sacrificing parsimony we will increase the reader's intuition for what we are doing.

The following formulae, derived in the text, are central for the derivations in this section.

Preferences Model

(Risk-lover, unbiased expectations)

$$U(O) = \frac{1}{p}$$

$$\Rightarrow O = U^{-1}\left(\frac{1}{p}\right)$$

Perceptions Model

(Biased expectations, risk-neutral)

$$\pi(p) = \frac{1}{1+O}$$

$$\Rightarrow O = \frac{1}{\pi(p)} - 1$$

The functions $U(O)$ and $\pi(p)$ are derived from Figure 1, and displayed in Figures 3 and 4, respectively. Given the odds of a horse winning, we can find the market's expectation of the true probability of that horse winning under either model.

An exacta is a bet on two horses to finish first and second in a particular order. In order to calculate the probability of an exacta winning, we use the assumption of conditional independence commonly found in the literature. This assumption states in particular that the probability of a horse finishing in some place in a race is independent of where the other horses finish. If we observe horse A finishing first (with first odds $A/1$), and we knew that it had a probability p_A of doing so, we can figure out the probability of horse B (with first place odds $B/1$) finishing second by determining the probability of it coming first in the race among the remaining horses. Of course, the fact that horse A came first provides us with no new information about the other horses, so we just renormalize the probability of the new race to 1, giving the conditional probability $p_{B|A} = p_B(1 - p_A)^{-1}$. Thus

$$p_E = \frac{p_A p_B}{1 - p_A}$$

$$\Rightarrow O_E = U^{-1}\left(\frac{1 - p_A}{p_A p_B}\right)$$

$$\pi_E = \frac{\pi(p_A)\pi(p_B)}{1 - \pi(p_A)}$$

$$\Rightarrow O_E = \frac{1 - \pi(p_A)}{\pi(p_A)\pi(p_B)} - 1 = O_A(O_B + 1) - 1$$

We can then test the predictions of the two models against the observed exacta odds. In Table 1 we do this, but we express both the predictions and the observed odds as the amount you have to pay to win \$1. This is simply given by $(1 + Odds)^{-1}$.

A quinella is a bet on two horses to finish first and second in any order. Thus, the probability of an A-B quinella is the same as the sum of the probabilities of the A-B

exacta and B-A exacta. A trifecta is a bet on three horses to finish one-two-three in a specific order. The probability of a trifecta is found by starting with the probability of an A-B exacta and applying the assumption of conditional independence once again to determine the probability of a horse C (with win odds $C/1$) crossing the line third after A and B have crossed first and second. This gives us the following relations:

$$p_Q = \frac{p_A p_B}{1 - p_A} + \frac{p_A p_B}{1 - p_B} \quad \pi_Q = \frac{\pi(p_A)\pi(p_B)}{1 - \pi(p_A)} + \frac{\pi(p_A)\pi(p_B)}{1 - \pi(p_B)}$$

$$p_T = \frac{p_A p_B p_C}{(1 - p_A)(1 - p_A - p_B)} \quad \pi_T = \frac{\pi(p_A)\pi(p_B)\pi(p_C)}{(1 - \pi(p_A))(1 - \pi(p_A) - \pi(p_B))}$$

The odds implied by each model can then be calculated by inverting as before.

APPENDIX B: Data

Our dataset consists of all horse races run in North America between 1992 and 2002. The data was generously provided to us by Trackmaster, a subsidiary of the jockey club. The data record the performance of every horse in each of its starts, and contains the universe of officially recorded variables having to do with the horses themselves, the tracks, and race conditions.

Our concern is with the pricing of bets. Thus, our primary sample consists of the 5,376,560 observations in 647,903 races for which win odds and finishing positions are recorded. We use these data, subject to the data cleaning restrictions below, to generate the data that allows us to map between odds and the true probability of a horse winning. We are also interested in pricing exacta, quinella, and trifecta bets. For about 10% of these observations (59,414 races covering 491,040 horse race starts) we also have the exacta and quinella payoffs for the actual winners of that race. (The prices of non-winning combinations are not recorded.) In 50,421 of these races (covering 429,493 horse race starts) we also observe the winning trifecta payoff.

Due to the size of our dataset, whenever observations were suspect, or difficult to deal with, we simply dropped the entire race from our dataset. Specifically, if a race has more than one horse owned by the same owner, rather than deal with “coupled runners,” we simply dropped the race. Additionally, if a race had a dead heat for first, second, or third place, the exacta, quinella, and trifecta payouts may not be well defined. Thus, we dropped these races. Certain observations on the winning odds were clearly wrong (e.g., when the odds were reported as zero), and so we simply dropped the entire race whenever the odds suggested that the sum of the probabilities implied by the odds was less than 118%, or greater than 128% (i.e., unless the track take implied by the odds is between 15% and 22%, we drop the data as likely to be faulty). After these steps, we are left with 4,867,857 valid observations on win bets from 588,175 races and 447,535

observations in 54,013 races include both valid win odds and payoffs for the winning exotic bets.

Finally, Figures 1, 6, and 7 show a mapping between odds and returns for different subsets of the data. For prices that are relatively common (such as 4/1), we had enough observations that we could reliably estimate the true probability. At more unusual levels we had to group together horses with similar odds. Our grouping algorithm chose the width of each bin so as to yield a standard error on the estimated rate of return in that bin that was less than 2%; we included all starts above 150/1 in a single final grouping. We used a consistent set of bins and data for all the results in our chapter, and linearly interpolated between bins when necessary.

CHAPTER 8

Unifying the Favorite-Longshot Bias with Other Market Anomalies

Russell S. Sobel

Department of Economics, West Virginia University, Morgantown, WV, USA.

Matt E. Ryan

Department of Economics, West Virginia University, Morgantown, WV, USA.

1. Introduction	138
2. Biases Found in the Previous Literature	138
3. What Causes the Favorite-Longshot Bias at the Racetrack?	140
3.1. <i>The Casual Bettor</i>	143
3.2. <i>The Serious or Regular Bettor</i>	144
3.3. <i>The Arbitrageur</i>	148
4. Is it Risk or Information?	150
5. Can the Model Explain the Biases in Other Markets?	156
6. Conclusion	158
<i>References</i>	158

1. INTRODUCTION

The favorite-longshot bias in racetrack wagering is perhaps the most well-documented anomaly in the literature on betting market efficiency. Dozens of studies have shown a consistent bias in which the expected return from betting on favorites exceeds the expected return from betting on longshots. This chapter focuses on the bigger picture with respect to this anomaly, namely to what extent it can be viewed as a bias that can be integrated with the differing biases found in other financial and betting markets—biases that are often in the opposite direction.

This chapter begins by reviewing the empirical findings on how the favorite-longshot bias differs across different types of bets, and different betting markets. The chapter then proceeds by discussing the different types of bettors, and also the different possible theories that have been presented to explain this bias, and whether they can be viewed in a broader context to provide a consistent explanation for these many biases. Finally, this chapter discusses the issue of arbitrage, and specifically why arbitrage may not eliminate these differential returns across bets.

2. BIASES FOUND IN THE PREVIOUS LITERATURE

Economists have spent considerable effort exploring the efficiency of market outcomes. Betting markets, because of their unique characteristics, provide an interesting case in which to conduct empirical studies of efficiency. At the top of the list of favorable characteristics is the outright abundance of data available—a facet that makes analyzing the stock market fruitful as well. However, Thaler and Ziemba (1988) note distinct advantages that betting markets have over financial markets:

The advantage of wagering markets is that each asset (bet) has a well-defined termination point at which its value becomes certain. The absence of this property is one of the factors that have made it so difficult to test for rationality in the stock market. Since a stock is infinitely lived, its value today depends both on the present value of future cash flows *and* on the price someone will pay for the security tomorrow. Indeed, one can argue that wagering markets have a better chance of being efficient because the conditions (quick, repeated feedback) are those which usually facilitate learning.

Despite the feedback and large-scale repeated-play aspect of betting markets, the search for market efficiency has led to the identification of an unexpected, and persistent, anomaly known as the favorite-longshot bias. In practice, weak-form market efficiency would imply that betting on racetrack favorites should be equally profitable as betting on longshots. Dozens of empirical studies of racetrack betting, however, have consistently found that the return to betting on favorites is higher than the return to betting on longshots. While Thaler and Ziemba (1988) provide perhaps the best overview of this body of literature, some of the more influential individual studies include Ali

(1977, 1979), Asch et al. (1982, 1984, 1986), Busche and Hall (1988), Figlewski (1979), Hausch et al. (1981), Losey and Talbott (1980), Snyder (1978), Swidler and Shaw (1995), Tuckwell (1983), and Ziemba and Hausch (1986). In some particular instances, such as discrepancies in win bets versus place and show bets (Hausch et al., 1981) as well as in exacta markets (Ziemba and Hausch, 1986), the return differential has actually been strong enough to produce deviations from weak-form market efficiency as well, in the form of profitable (positive expected value) wagers.

This unexpected bias identified in betting markets, while present in the vast majority of studies, does not uniformly hold true. Several notable studies have, in fact, found an exactly opposite bias, with favorites being overbet and longshots underbet, an opposite favorite-longshot bias. Swidler and Shaw (1995) find this opposite bias for a smaller Class II racetrack in Texas, and Busche and Hall (1988) find this opposite bias at a racetrack in Hong Kong that had much higher betting volumes than the U.S. racetracks used in other studies. More recently, Sobel and Raines (2003) have identified a situation in which the bias changes from the regular favorite-longshot bias to an opposite bias, under certain predictable conditions. Gramm and Owens (2005) find that the favorite-longshot bias diminishes for races with larger betting pools and more entrants. While these studies might appear at odds with the other literature on racetrack wagering, their results are actually more consistent with findings from studies of efficiency in markets other than racetrack wagering.

Studies of Las Vegas sports wagering have concentrated on betting on National Football League (NFL), National Basketball Association (NBA), National Hockey League (NHL), and Major League Baseball (MLB) games. In these markets, deviations from market efficiency have been found in about three-fourths of the studies, and in the vast majority of cases where it existed, it was an *opposite* favorite-longshot bias. In the terminology of this literature, bettors tend to overvalue favorites. Gandar et al. (1988) show that a strategy of betting on NFL longshots produces a return greater than betting on favorites for point spread betting on NFL games. Kochman and Badarinathi (1992) found through a simple analysis of role (underdog or favorite), location of the game (home or away), and month in which the game was played, that a rate of success above break-even could be achieved for wagering on NFL games between 1986 and 1990. Zuber et al. (1985) found an exploitable inefficiency in NFL point spread betting market during the 1983 regular season, while Lacey (1990) finds some profitable opportunities in certain betting rules in NFL point spreads from 1984 to 1986. In professional baseball, Woodland and Woodland (1994) find an opposite favorite-longshot bias in baseball betting against the odds, which they confirm still exists after including 10 additional years of data in their updated study, Woodland and Woodland (2003). In their study of NHL betting markets, Woodland and Woodland (2001) find a strong opposite favorite-longshot bias, which is confirmed in an updated and corrected version of this study by Gandar et al. (2004). Similar to the case of the regular favorite-longshot bias, these opposite favorite-longshot bias findings in professional sports betting markets are not always found. Sauer (1988), for example, finds efficiency in NFL over/under betting, while Johnson and Pawlukiewicz (1992) find efficiency in over/under betting in the NBA.

In the stock market, there are several findings related to the studies of betting summarized above. The first is the finding that portfolios of recent loser stocks seem to

outperform portfolios of recent winner stocks. To the extent that recent winner stocks are comparable to racetrack favorites, this is an opposite favorite-longshot bias. Several of the articles demonstrating (and providing explanations for) this bias in the stock market include De Bondt and Thaler (1985, 1987, 1990), Chan (1988), Bremer and Sweeney (1991), Brown and Van Harlow (1988), Lehmann (1990), and Howe (1986). The best overall summaries of the literature on stock market efficiency are De Bondt and Thaler (1989) and Thaler (1992).

More closely related is the finding of De Bondt and Thaler (1990) that professional stock market forecasters' predictions about company earnings tend to have an opposite favorite-longshot bias. They regress actual company earnings on forecasted earnings and find a slope coefficient significantly less than one. This implies that the companies with the highest forecasted earnings tended to be overestimates of true earnings, while the companies with the lowest forecasted earnings tended to be underestimates of true earnings. In studies of betting behavior, this type of bias would result in favorites being overbet and longshots being underbet. This phenomenon also manifests itself in variance comparisons, as the variance in the predicted values is significantly higher than the variance in the true-underlying values.

Perhaps most ironically, while economists have struggled to find an explanation for these biases, a similar opposite favorite-longshot bias has been found to exist in professional economists' *own* economic forecasts in the Livingston survey by Ahlers and Lakonishok (1983).

The link between deviations from market efficiency in these many markets has not been clearly made in the previous literature. They have been largely viewed as separate anomalies, each with its own potential explanation. However, the key to unlocking the mystery of why human behavior results in departures from market efficiency may lie precisely in developing a unified theory that simultaneously explains the phenomenon occurring in these different markets. In fact, with so-called anomalies, or departures from efficiency, appearing in every strand of economic research from international finance to labor economics, a unified theory is essential to our general understanding of how markets work. This chapter attempts to provide a starting point for that unified framework.

3. WHAT CAUSES THE FAVORITE-LONGSHOT BIAS AT THE RACETRACK?

Previous authors have attempted to provide theoretical explanations for the existence of the specific bias found in their study. Some of the explanations for the regular favorite-longshot bias in racetrack betting include risk-loving behavior on the part of bettors, poorly informed casual bettors who bet too evenly across all entrants, bettors getting direct utility from betting on longshots, a systematic tendency of individuals to overestimate the chances of low probability outcomes, and bettors having preferences over the skewness of the payouts. Ali (1977) suggests that this bias could simply be a general feature of any odds-based pari-mutuel system. In contrast to many of the utility-based theories grounded in the idea of a risk-loving bettor, Golec and Tamarkin

(1998) conclude that after controlling for skewness, the data are also consistent with risk-aversion. Theories explaining the favorite-longshot bias cover a broad spectrum, to say the least.

Despite their range, previous theories for the favorite-longshot bias can be generally grouped into two categories: risk-preference theories and information-perception theories. The first group of theories attributes the bias to a preference for risk among gamblers. Utility curves can be constructed so as to generate the desired effect; Friedman and Savage (1948) posited that consumers have a globally risk-averse utility function, yet over the relevant range concerning gambling and betting, the utility function becomes locally risk-loving. Weitzman (1965) introduced the concept of the “representative bettor,” along with this representative bettor behaving as an economic actor. Seeing as any rational actor would take the higher expected return over the lower one, and thus equilibrate the market of odds via expected returns, the representative bettor must be logically risk-loving in shifting his bets disproportionately (in profit terms) toward longshots. Starting with the conclusion that bettors are risk-loving, Quandt (1986) shows that the favorite-longshot bias must exist in order to clear the market. Because of the additional utility derived from taking high-risk (high variance) bets, the payoff of such a bet must necessarily be smaller for the high-risk longshot bet than the low-risk favorite bet.

The second group of theories, the information-perception theories, places the activities of the bettor as a reactionary to new information at the time of betting, as opposed to being naturally predisposed to take riskier bets. This literature generally assumes that bettors do not perfectly absorb information. Snowberg and Wolfers (2008), for example, point to “studies by cognitive psychologists that suggest that people are systematically poor at discerning between small and tiny probabilities (and hence they will price each similarly). Further, certain events are strongly preferred to extremely likely events, leading to even very likely events to be underpriced.” Ultimately, it is the representative bettor’s inability to process information correctly that leads to a favorite-longshot bias in the information-perception theories.

One common aspect in these explanations is the utilization of a representative bettor model in which all bettors are assumed to be acting in one particular manner. However, Sobel and Raines (2003) unambiguously show that this is not the case. The authors analyze nearly 2,800 races at two dog tracks in West Virginia.¹ By analyzing attendance, and the breakdown of types and sizes of bets placed on different days of the week at the same tracks, and on the exact same set of racing entrants, they find that some days, in

¹There exist some differences between betting on dog racing as compared to more traditional horse racing. While the structure of the betting, types of bets, track take percentages and information provided to bettors are all identical to horse racing, dog racing tends to exhibit a far higher percentage of serious bettors due to the fact that the incidence of exotic bets is four to five times higher than at horse tracks. For further explanation of serious bettors, see Section 3.1. In addition, dog tracks are much less costly to run than horse tracks. The typical attendance found in studies of horse racing is generally around 15,000 per night; the average dog track attendance in the Sobel and Raines dataset is approximately 1,300. The data exclude races in which entrants were scratched from the race, so that all races examined include eight entrants. However, races with fewer entrants are analyzed separately in the original paper and also provide evidence in favor of their hypothesis. In addition, races in which there were ties in the finish were also excluded to avoid problems in computing objective probabilities.

TABLE 1 Favorite vs. Longshot Betting: Weekends vs. Weekday Nights

	Weekdays	Weekends	Ratio weekdays/weekends
Average wager per person per race	\$6.28	\$5.71	1.100
Exotic bets (% of all bets)	17.3%	13.2%	1.311
Average percentage of win pool bet on favorite	28.9%	26.6%	1.086
Average percentage of win pool bet on longshot	3.8%	5.1%	0.745
Favorite/longshot betting ratio	7.5	5.2	1.442

NOTE: Data derived from Sobel and Raines (2003). The favorite/longshot betting ratio is the average percent of win bets on the favorite divided by the average percent of win bets on the longshot.

particular weeknights, have a substantially larger percentage of serious bettors, while the weekends have a higher portion of casual bettors. This observation follows from the fact that serious bettors are characterized by larger average wagers and a higher incidence of exotic bets. Casual bettors, on the other hand, tend to wager less, and wager on simpler bets.² Table 1 shows the breakdown of favorite and longshot wagers during the weekdays and the weekends in the Sobel and Raines dataset, along with data on average wager size and exotic bet frequency.³

As can be seen in Table 1, the average bet per person per race is 10% higher on weekdays than on weekends, and the proportion of exotic bets is over 30% greater. Both of these outcomes are consistent with a higher proportion of casual bettors at the track on weekends than on weeknights. The result of this difference in the favorite and longshot betting patterns is also clearly visible. On the average weekday night, for every \$1 bet on the longshot, there is \$7.50 bet on the favorite. On the weekends, however, this falls substantially to \$5.20 bet on the favorite for every \$1 bet on the longshot. This discrepancy cannot be explained by a difference in the characteristics of the races themselves, because at the two tracks studied the *exact same* racing entrants are used across days. This visible difference, then, is entirely due to a difference in the types of bettors at the track—a difference that is not arbitrated away.

Figure 1 shows a plot of the weekday subjective betting percentages against the weekend subjective betting percentages for all racing entrants (not just the favorite and longshot). Each point is also for a specific grade of race. The observation in the upper right hand corner, for example, is the average betting on the first favorite in grade AA races on weekdays plotted against the same value for weekends. There is a clear rotation

²Casual bettors also tend to leave the track earlier, allowing later evening races to actually have a different proportion in the bettor mix than earlier races.

³At both tracks studied, races are offered on Monday evening, Wednesday evening, Thursday evening, Friday evening, Saturday afternoon, Saturday evening, and Sunday afternoon. In addition, one track, Wheeling Downs, also offered a Wednesday afternoon race. The weekdays subsample includes all races on worknights (Monday, Wednesday, and Thursday nights), while the weekend subsample includes Friday evening, and both races on Saturday. Wednesday afternoon and Sunday races are not included in either subsample, but are included in all data showing combined numbers for all races. Wednesday afternoons are special events at which large tour groups are present, appearing more like a weekend. Interestingly, Sunday at one track appears more typical of a weekend, while at the other track more typical of a weekday night, in terms of average attendance and betting patterns. For these reasons it was included in neither subgroup.

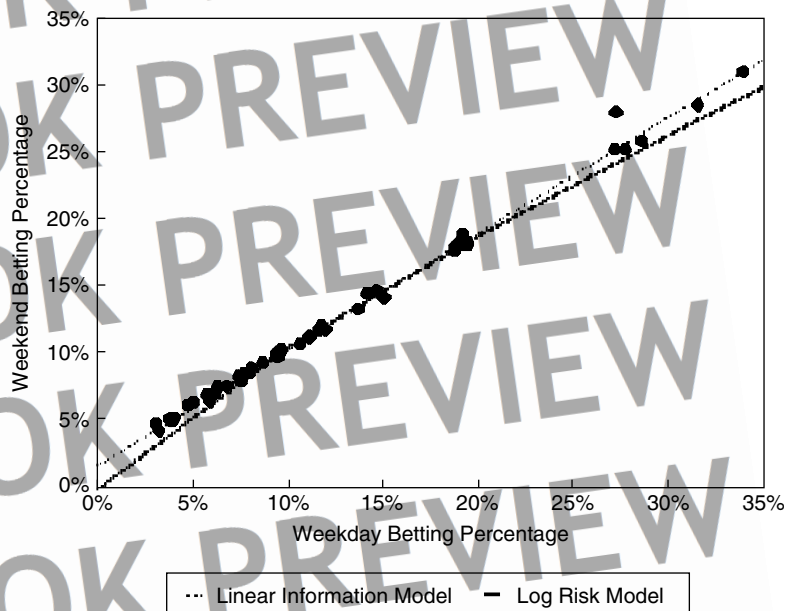


FIGURE 1 Weekday vs. weekend betting: all dogs by grade.

NOTE: Taken from Sobel and Raines (2003). Individual observations are the percentages bet on the favorite, second favorite, and so on, to the longshot, by the grade of the race. The observation in the upper right hand corner, for example, is the average betting on the first favorite in grade AA races on weekdays plotted against the same value for weekends. Lines show fitted regressions from the linear information model and the double-log risk model.

of the bets, showing an outcome on weekends that might best be described as the weekday bets, just with an added proportion of uniform bets across all racing entrants and grades. Statistical testing shows clearly that the objective probabilities across the two tracks, different grades, and different racing conditions are identical. This difference in betting percentages occurs despite the exact same probability structure facing the bettors.

Given the clear presence of heterogeneous bettors, and the potential for heterogeneous bettor models to explain the favorite-longshot bias, we now turn to an explicit discussion of the different types of bettors generally participating in betting markets. We identify three such categories: the casual bettor, the serious or regular bettor, and the arbitrageur.

3.1. The Casual Bettor

The casual bettor does not attend the track frequently, and is thus less able to use information to make an informed bet.⁴ Information used for a wealth-maximizing bet can

⁴Piron and Smith (1995) and Sobel and Raines (2003) are the only studies that attempt to explicitly account for casual bettors.

come from a variety of sources, including previous race performance (and quality of opponents), track conditions and locale, as well as initial race odds and preliminary wagering. The casual bettor can simply be viewed as being able to only fully integrate some of this information. In a Bayesian framework, the casual bettor simply has more weight on a uniform prior in the construction of the posterior probabilities. In many cases these bettors are at the track accompanying friends or family and have other non-wealth-maximizing objectives, or could even be deriving utility from the act of gambling itself. Casual bettors will make bets in a more arbitrary fashion, and are also more likely to choose simpler and less risky types of bets. Arbitrary bets could be based on any predisposed preference—the race position or number of the entrant, the color the entrant is wearing, or even the entrant's name or appearance. By definition, these bets will tend to be more evenly distributed across all participants than the true probabilities would suggest. Thus, the actions of this group of bettors results in a pool of bets that are placed too heavily on longshots and too lightly on favorites, which by itself can generate a regular favorite-longshot bias in which the return to betting on the favorites is higher than the return to betting on longshots. These casual bettors will overbet any racing entrant with an objective probability lower than the average probability of $1/N$ (where N is the number of racing entrants), and will underbet any racing entrant with an objective probability higher than average. The result is a linear rotation of the betting proportions around $1/N$. In a game theoretic approach, making money by playing against these causal bettors is the focus of our second group, which we term serious or regular bettors.

3.2. The Serious or Regular Bettor

The serious or regular bettor is the type of person who attends the track regularly, is an informed participant in the betting market, places bets with a strong understanding of the participants in the races, can digest the vast amount of information provided in the racing forms, and believes he has a knack for picking the entrant most likely to win based on past performance. He will be someone who consistently picks the winning entrants or combinations more frequently than the casual bettors at the track. This type of bettor gets utility from the action of picking the winning entrant, or winning money, more often through the betting night than other serious and casual bettors in his social group at the track. Making money by playing against causal bettors who cannot accurately identify the likely winning entrants is the goal of serious or regular bettors.

This type of bettor is likely to place bets to win, place, or show on who he thinks is likely to win the race, and to bet on the likely combinations of the best entrants in the multi-entrant, exotic bets. By doing so, the actions of the serious or regular bettor generally lead to an opposite favorite-longshot bias in which the favorites are overbet and yield a lower return. These bets are often supplemented with much riskier exotic bets, including picking the top two or three finishers. These bets are often placed with a wheel or box option, in which the bettor picks his favorite one or two entrants, and then buys a single bet on every possible combination of those two entrants near the top,

with all other entrants in the remaining places. Almost by definition, these types of bets cause a favorite-longshot bias, because they place a single bet on each of many possible combinations, not all of which are equally likely to occur.

A better term for these types of bettors may be winner pickers, as this is their betting strategy's main focus. These bettors examine and understand the odds system, and frequently utilize the displayed odds prior to the race as a supplement to their own unique information to further help their ability to pick the entrants most likely to win. These serious or regular bettors sometimes have inside information not available to all bettors—for example, they could be closely associated with the owners or trainers. Further, they may think their analytical abilities based on their experience allow them to pick the true winners better than the other bettors against which they compete. Again, they use this information to attempt to pick the *winning* entrant, and often wait to place their bets until near the close of the betting period so as to not reveal their actions (through the displayed odds) to other bettors. Gandar et al. (2001) show that bets placed closer to the end of betting more closely correspond to the true objective probabilities than early bets.

These serious bettors, however, as is more often assumed, are not arbitrageurs. Serious bettors and arbitrageurs are two distinctly different groups. Sobel and Raines consistently find in their data that the number of clearly visible arbitrage opportunities present at the track is actually higher on days dominated by more serious or regular bettors.⁵

Table 2 shows the estimates of how the favorite-longshot bias in the win betting market differs between weekdays and weekends.⁶ On the weekday races, dominated by a larger proportion of serious bettors, there is a strong opposite favorite-longshot bias that weakens as the betting pool rotates more heavily toward longshot betting on weekends. Had the effect been slightly larger on weekends, the bias would turn into a regular favorite-longshot bias on weekends. On weekdays, despite the presence of a large proportion of more serious bettors, there is a clearly arbitragable positive rate of return on longshot betting in the win market. One important implication of this finding is that the biases are not the *result* of arbitrage equilibriums, but rather are allowed to persist because of its absence.⁷

To explore for any time-of-day effects, the data were grouped into the first 13 races and the final two races of the day. An immediate problem is that races vary in grade from AA to M. The data suggest that favorites win a higher proportion of AA races than any other grade. In fact, the proportion of races that a favorite wins declines as the grade of the race becomes lower. This is relevant because in almost 90% of the days at both

⁵In their study this is weekday evenings when attendance is lower and when exotic betting is much heavier. This result also holds for "Superfecta Sundays" when these types of bettors dominate the track.

⁶In cases where two racing entrants tied for being the favorite (having identical odds and thus subjective probabilities), each was viewed as the first favorite with weight one-half and the second favorite with weight one-half.

⁷In Ali's (1977) model, for example, the favorite-longshot bias is explained as a natural result of bettors making bets relative to the market odds, when those odds diverged from the bettor's guess about the true win probabilities, thus making it a result of arbitrage.

TABLE 2 Objective vs. Subjective Win Probabilities: By Group, Removing Ties

Favorite	All days (2,558 races)			Weekdays (952 races)			Weekends (1,066 races)		
	Objective probability of win (π)	Subjective probability of win (ρ)	Rate of return from bet	Objective probability of win (π)	Subjective probability of win (ρ)	Rate of return from bet	Objective probability of win (π)	Subjective probability of win (ρ)	Rate of return from bet
1	27.64%	25.27%	-23.72%	28.33%	24.53%	-27.41%	27.36%	25.69%	-22.08%
2	18.45%	18.04%	-18.89%	18.82%	20.12%	-9.47%	18.23%	16.01%	-27.91%
3	14.26%	14.82%	-13.90%	14.22%	14.06%	-18.18%	14.25%	15.76%	-8.13%
4	11.55%	11.53%	-18.45%	11.48%	12.41%	-10.79%	11.57%	11.02%	-22.43%
5	9.49%	9.27%	-17.61%	9.33%	9.45%	-16.67%	9.58%	9.80%	-11.73%
6	7.78%	8.84%	-7.72%	7.60%	7.67%	-19.73%	7.85%	9.04%	-6.28%
7	6.23%	6.48%	-12.64%	5.96%	5.64%	-20.09%	6.36%	7.18%	-7.94%
8	4.61%	5.76%	17.48%	4.26%	6.13%	36.96%	4.80%	5.50%	17.93%

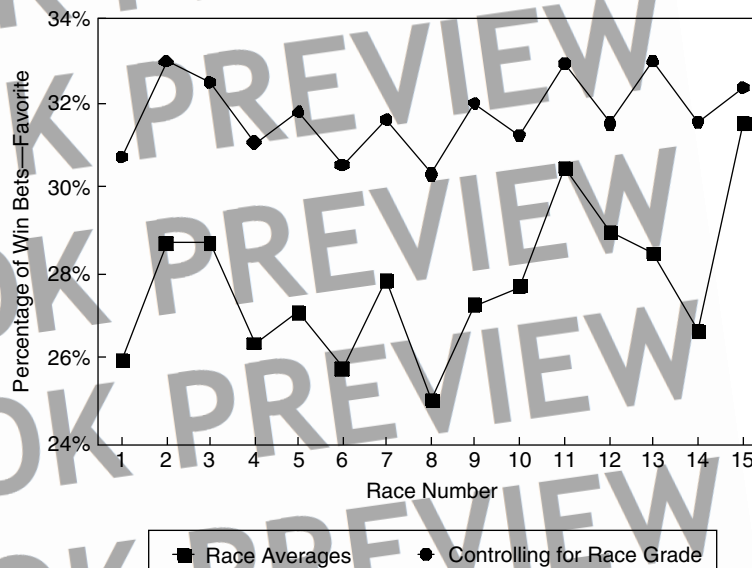


FIGURE 2 Time of day effects on favorite betting. NOTE: Taken from Sobel and Raines (2003).

tracks, the final two races were grade AA. Because favorites win a larger share of grade AA races, and the final race tends to be grade AA, it is natural to expect betting on the favorite to be higher in the final race irrespective of risk, wealth, or mental accounting. To adjust for these grade effects, we performed a simple race level regression controlling for race number (a constant was omitted) and grade (AA was the excluded group). Figures 2 and 3 show the raw race averages and the averages after controlling for grade for the favorite and longshot, respectively, for each and every race of the day.

Both figures clearly show the importance of controlling for grade effects in the last few races. The raw data show that betting on the favorite rises significantly in the last race, which disappears after controlling for the fact that it is a grade AA race. Looking at the corrected data, betting on the favorite remains fairly constant throughout the racing day. The notable exception is in the final two races, where betting on the longshot reverses its slight downward trend and rises up again, especially in the last race of the day. This last race effect appears to be isolated to the longshot, with no similar last race effect happening for the favorite.⁸

⁸This is also evidence against a risk-preference theory of the favorite longshot bias. As the average bettor's wealth declines throughout the evening (due to track take), risk-preference would suggest they move to more risk-averse bets, but they don't. One plausible explanation of the general trend over the last half of the evening is that attendance falls significantly through the evening, and it is disproportionately casual bettors who leave early, implying that the later races are dominated by serious bettors who bet more heavily on the favorites. Thus, the pattern through the course of the racing day is more supportive of the information theory (casual/serious bettor differences) than of the risk theory.

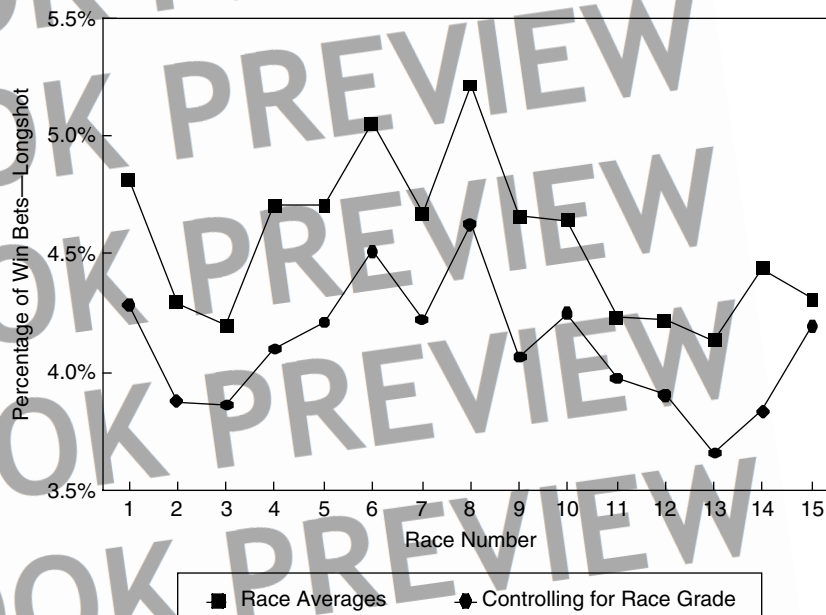


FIGURE 3 Time of day effects on longshot betting. NOTE: Taken from Sobel and Raines (2003).

3.3. The Arbitrageur

The arbitrageur is the typical type of agent modeled in economic theory. He or she takes into account all available information, and attempts only to maximize his or her return on investment with no other utility considerations, and risk neutral preferences. As in other markets, arbitrageurs are the agents that drive markets toward efficiency. Profit opportunities are quickly identified and acted on, eliminating any deviations from efficiency. Most important to this type of bettor is the information conveyed in the market odds at the track that constantly change to reflect betting intensities, and how these compare, *for all entrants*, to his or her best estimate of the true probabilities. The arbitrageur cares little about who might win the race, and more about which entrant has the highest expected return based on any divergence he or she sees.

The actions of arbitrageurs lead to market efficiency, in the weak sense. A market is weakly efficient as long as no persistent profit opportunities exist that could be exploited based on all available information. While there is no doubt that these types of individuals exist in society, and account for the efficiency of many other markets, the persistence of the favorite-longshot bias, and of its opposite, imply that either: (1) they simply do not participate to a great enough extent in these markets, or (2) they don't have the necessary information or at least not in a timely enough manner to utilize it before the betting window closes for the race.

There is substantial evidence in Sobel and Raines (2003) that these arbitrageurs are absent to a large extent at the two tracks they study based on the clear and persistent presence of arbitrage opportunities at these tracks. These occurrences actually increase on nights (or during later races) dominated by the serious or regular bettors. Their data suggest that these arbitrage opportunities are less frequent on higher-attendance weekend nights, not because arbitrageurs are there to arbitrage these away, but because of the presence of casual bettors who tend to bet on longshots. The serious bettors at these tracks appear to allocate their time and money toward betting in the exotic bets, and attempting to pick winners better than everyone else, with little regard for attempting to arbitrage away even visible opportunities. In fact, evidence from other studies appears to be mounting that the vast majority of bettors tend to make bets independent of the track odds.

Camerer (1998), for example, finds that track bettors did not respond to large changes in track odds. Camerer went to a local track and made large bets on single horses to change their odds. He then withdrew the bets just before the betting on the race closed. By comparing the remaining bets in these races with bets made on similar horses when he did not interfere, he was able to see how the large changes in odds influenced other bettors. He found no statistically significant difference. He interprets this result as that it did not manipulate the bettors who saw through the incorrect odds, but in fact arbitrage would suggest that bettors should have responded to this horse being overbet by betting on other entrants, who have now become underbet. The bettors did not. His results also suggest that these average racetrack bettors do not actively search out or participate in arbitrage activities. In a similar study, Hanson et al. (2006) find experimentally that prediction market prices are unaffected by agents wishing to distort them. Wolfers and Zitzewitz (2004) survey studies in price manipulation and conclude that no effect exists on prices beyond the short term. A historical study by Strumpf and Rhode (2003) finds, as well, that efforts to manipulate prices are ultimately unsuccessful.

It is important to note that arbitrage involves both discovering profit opportunities, which often requires data and time-intensive statistical analysis, and then acting on this information. This is precisely the type of entrepreneurship modeled by Kirzner (1973). The entrepreneur is someone who spots previously unnoticed profit opportunities and acts on them. This not only takes a certain type of individual to perform, but also requires a market return from their actions.

One must keep in mind that the arbitrage opportunities available at racetracks are not only minor, but are also uncovered. Unlike triangular exchange rate arbitrage, race-track bettors must still take risk to exercise arbitrage. In addition, like any other pure return-maximizing strategy, returns will be driven down to zero *economic* profit, not zero accounting profit, on the time and investment needed to conduct the arbitrage itself. While the return to betting on favorites exceeds the return on longshots in most studies, the absolute level of the return on favorites must be high enough to overcome the track take which is usually between 15% and 20%. However, using experimental methods, Hurley and McDonough (1995) find that changing the track take has no effect on the severity of the favorite-longshot bias. In addition, the track minimum payout of \$1.10 on

a \$1 bet creates a lower bound on the return from a bet on an extreme favorite, perhaps at a level that is positive. Swidler and Shaw (1995) point out that the total daily monetary return on the arbitrage opportunities they identify would not be enough to cover the cost of admission, parking, the program, and a minimum wage rate per hour given the price setting ability of large individual bets. One might similarly ask why economists who have found and published numerous examples of profitable arbitrage opportunities in the stock market have not used their own money to arbitrage these inefficiencies away. The reason, of course, is the opportunity cost of economists' time and the risk involved. With so many arbitrage opportunities abounding in the world, it would seem to be consistent to say the arbitrageurs have plenty of work and can concentrate on the least costly and most covered opportunities, while leaving the smaller inefficiencies abounding in other markets.

As Leeson et al. (2006) discuss, the presence of forgone profit opportunities is much more likely than the presence of persistent losses in markets. Economic losses diminish resources under the control of the decision-maker, and lead to business failure. When a potential profit opportunity goes unexploited, while there is an opportunity cost, it does not create the same strong incentive for self correction. In fact, an individual may never know he or she passed up a potentially profitable opportunity, but will certainly know if he or she earns a loss. According to this theory, then, the weaker incentive for market correction can lead to a greater presence and persistence of unexploited profit opportunities in markets.

One final explanation for the lack of arbitrage is that the bettors capable of this activity simply do not have the necessary information or at least not in a timely enough manner to utilize it before bets close for the race. Because the returns on bets in economic studies are based on the final odds, these are not the odds visible during the entire betting period to potential bettors. At some tracks, for example, late money bets, accompanied by the delay in updating the odds on the tote board, and the unpredictable length of time needed to wait in line to place the bet, means that it is almost impossible to make a bet based on the final market odds.

4. IS IT RISK OR INFORMATION?

Thus we are left with two potential theories to explain the favorite-longshot bias (or its opposite). It is either the result of an equilibrium outcome driven by arbitrageurs with a preference for risk, or it is the result of a weighted average of casual and regular bettors' strategies without offsetting arbitrage occurring. And while the risk-preference model dominated the literature for decades, modern evidence seems to clearly point in the other direction.

The most convincing evidence on this comes from Sobel and Raines (2003), who gathered not only overall odds and payout data for the win market, but also data on all other markets, as well as the day of the week, track attendance, number of racing entrants in the race, the grade (or quality level) of the race, and the race number. In addition, data were collected on the various types of bets placed, from simple to exotic.

This breadth of supplemental data allowed them to see how the favorite-longshot bias changed across different bets existing at the same time at the track, as well as how these biases changed across different nights that had significantly different proportions of serious and casual bettors.

They conclude that the bias can be explained (and its presence in certain situations, and the opposite in others) as a reflection of changing combinations of serious and casual bettors. This information-based model of serious and casual bettors has clearly different empirical predictions from the risk-preference model, most notably the linear versus non-linear nature of the betting percentages relative to their true values, and also in terms of how the bias would change based on the number of racing entrants in the race, and across different types of bets.

Figure 4 shows the relationships between objective (true) probabilities of an entrant winning the race and the subjective probability of the entrant winning, gauged by true betting percentages. In an efficient market, subjective probabilities match perfectly with the objective probabilities, which would be reflected by a 45° line from the origin.

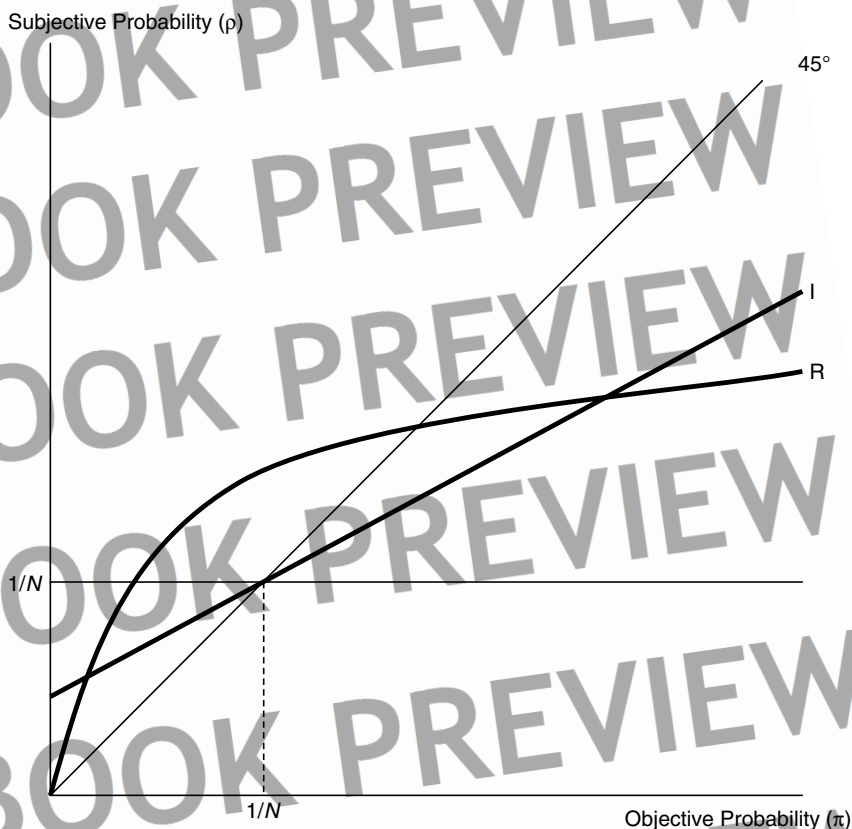


FIGURE 4 Information vs. risk in subjective and objective probabilities.

No bias exists in this case. The existence of less-informed bettors, who bet too evenly across racing entrants, can be illustrated as a rotation of this line around the average probability of $1/N$. This is shown by the line I in the figure. The degree to which casual bettors dominate the betting market will determine the degree to which the line rotates.

The risk model does not involve a linear rotation as is true in the case of the information-based model. Instead, pre-existing desires for riskier bets dictate betting patterns. A typical risk-preference would result in a betting pattern illustrated by the curve R in Figure 4 according to the relationship $\rho_i = \lambda \pi_i^\tau$, where ρ_i represents the subjective probability for racing entrant i , π_i represents the objective probability for racing entrant i , and λ and τ are parameters of the utility function that measure preference for risk that can be estimated using betting market outcomes. The existence of the favorite-longshot bias in this theory comes from a desire of bettors to choose the riskier bets associated with longshots.

Now, for comparison, it is worthwhile to briefly derive the mathematical properties of an information-based model for risk-neutral bettors with Bayesian updating. Assuming that prior to arriving at the track, each bettor has no knowledge about the racing entrants, thus all entrants are equally likely to win in the diffuse prior. The mean of the bettor's prior belief about the entrants probability of win, ρ_P , must be $\rho_P = 1/N$. Once at the track, the bettor collects a set, k , of information signals regarding the racing entrants' true probability of win. Let ρ_I denote the mean probability implied by the information signals received by the bettor. Using Bayes' theorem, the mean of the bettor's posterior probability of win distribution, ρ , on which the bettor makes bets, is given by $\rho = (1 - h) \rho_P + h \rho_I$, where h is a precision parameter that allocates the relative weight placed on the new information and the prior belief. The precision parameter $h = h(k, \sigma_P, \sigma_I, c)$ is a function of the information received, the variance of the priors and the information signals, and of the degrees of freedom, or complexity of the bet, c . Without assuming a specific form for h , the relationship with its arguments is intuitively obvious

$$\frac{\partial h}{\partial k} > 0, \quad \frac{\partial h}{\partial \sigma_I} < 0, \quad \frac{\partial h}{\partial c} < 0.$$

This simply says that as a bettor becomes more information rich, he or she will choose to place less weight on prior beliefs. However, if there is high variance in the new information, or the complexity of processing the information is great, then the bettor will place more weight on prior beliefs. Let the relationship between the entrant's true probability of win and the mean of the information signals conveyed to the bettor through the information be given by $\rho_I = \nu + \phi\pi$. The posterior mean equation from which the bettor places his or her bets is $\rho = \alpha + \beta\pi$, where $\alpha = (1 - h)/H + h(\nu + \phi)$ and $\beta = h\phi$. Strong form market efficiency would exist if $\beta = 1$ and $\alpha = 0$. The relation $0 < \beta < 1$ gives the traditional favorite-longshot bias, while $\beta > 1$ produces an opposite favorite-longshot bias.

How do these differing model predictions fit actual betting data? Figure 5 shows data on the objective and subjective probabilities for all quinella bets at the tracks studied

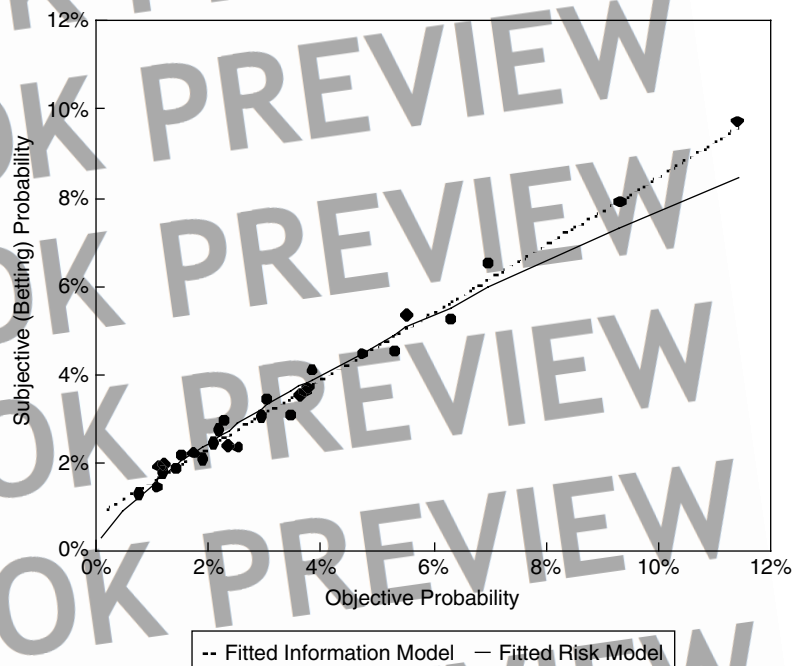


FIGURE 5 Risk model vs. information model in quinella bets: subjective vs. objective probabilities. *Source:* Sobel and Raines (2003). (Observations are for each possible quinella combination, illustrating the objective and subjective probabilities of each.)

by Sobel and Raines, along with fitted risk and information model regression lines. In Figure 5 the relationship in the data is almost perfectly linear, in rejection of the risk explanation that implies a non-linear relationship between objective and subjective probabilities. The linear relationship suggested by the information model does a particularly good job in fitting the data at the two extremes relative to the risk model.

To this point we have focused on a comparison of the subjective and objective probabilities. We now turn our attention to how this relationship appears once converted into the rates of return on bets versus objective probabilities. Let π_i denote the true win probability of racing entrant i . Let ρ_i denote the betting market's subjective estimate of the entrant's probability of win. Griffith (1949), and McGlothlin (1956), were the first to observe that this subjective probability is the percent of the total betting pool that is bet on this particular racing entrant i . The track payout per dollar if the bet wins, $\$i$, is given by $\$i = O_i + 1$, where O_i is the track odds, which are defined in terms of the underlying subjective probabilities as $O_i = (1 - \rho_i)/\rho_i$.⁹ Further substitution and simplification produces $\$i = 1/\rho_i$. The expected return per dollar bet, $R = \$i \times \pi_i + 0 \times (1 - \pi_i) = \$i \times \pi_i$. The rate of return, r , for a \$1 bet, is $r = (R - 1)/1$,

⁹Here we abstract from the track take for simplicity. Inclusive of a track take of t , the track odds are $O = (1 - \rho')/\rho'$ where $\rho' = \rho^*(1 - t)$.

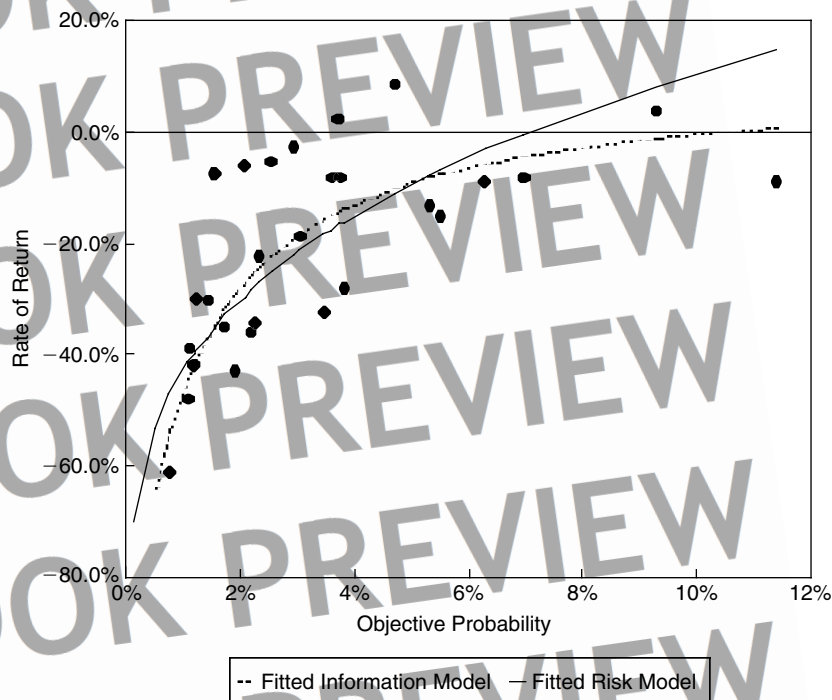


FIGURE 6 Risk model vs. information model in quinella bets: rate of return vs. objective probabilities. NOTE: Taken from Sobel and Raines (2003). Observations are for each possible quinella combination, illustrating the objective probabilities and the corresponding actualized rates of return.

which through substitution yields $r = (\pi_i - \rho_i)/\rho_i$. Figure 6 shows the data from Figure 5 transformed using this relationship.

In the transformed data, the information model and the risk model now both predict a non-linear relationship, however these non-linear relationships are clearly different. Again, the information model clearly fits the data better than a risk-preference model.

Additional evidence in favor of the information-based explanation is found by examining bets of different complexity that exist simultaneously at the track. More complex, exotic bets function to turn informed regular bettors into bettors with less information in a Bayesian-type information model. Information-perception models are grounded in the ability of the bettor to digest information and to make an appropriate bet based on this information. As such, more complicated bets should see the same smoothing effect when compared to easier ones. Sobel and Raines' data show that the relative lack of information available for the more complex bets creates a greater smoothing effect among the more complex bets.

Table 3 displays the regression results for each of five different types of bets for which data were available and present at the same time at the track. The regressions

TABLE 3 Objective vs. Subjective Probability Regressions: All Bets

Betting market regressions. Dependent variable is subjective probability (standard errors shown in parentheses)					
	Win	Place	Show	Quinella	Perfecta
Constant	-0.018** (0.006)	-0.003 (0.005)	0.049** (0.013)	0.008** (0.001)	0.003** (0.001)
Objective probability	1.143** (0.041)	1.013** (0.020)	0.869** (0.032)	0.770** (0.019)	0.788** (0.035)
R ²	0.9924	0.9977	0.9919	0.9849	0.9062
Betting combinations	8	8	8	28	56
Observations per betting combination	2558	2558	2558	2459	893
T-ratio for slope = 1	3.503**	0.645	-4.082**	-12.262**	-6.136**
Estimated rotation point	12.59%	25.10%	37.51%	3.55%	1.45%
Predicted rotation point (null prior)	12.50%	25.00%	37.50%	3.57%	1.79%

Source: Sobel and Raines (2003). (**indicates statistical significance at the 1% level, * at the 5% level. The test of market efficiency is not just that the slope equals one, but also that the intercept term equals zero. The *F*-statistics for this joint test strongly rejected this hypothesis for all markets except the place market. The predicted rotation point is the average probability across all betting combinations (i.e., one of eight entrants win, two of eight place, three of eight show, one of 28 quinellas win, and one of 56 perfectas win). The rotation point estimated by the model is equal to the constant divided by one minus the slope coefficient. This is the probability at which any bet above has the opposite bias as any bet below in the information model.)

are ordered so that the complexity of the bet, in terms of the number of parameters or data points needed to estimate the objective probability, increases as we move rightward among the columns in the table. The more complex bets show lower regression coefficients—in other words, the slope of the line becomes flatter. This is the exact result predicted by an information-based model—more flattening due to the smoothing effect of casual bettors. A further strike to the risk-preference theories is the fact that the least risky market (the show bet) and the most risky one (the perfecta) demonstrate the usual favorite-longshot bias, while the middle-risk bets show either little bias or an opposite favorite-longshot bias. These biases, however, line up in order when they are ranked by complexity.

Also, Sobel and Raines examine how the bias changes as the number of racing entrants changes. Within a pure risk model, the over (or under) betting of a particular entrant is entirely dependent on the entrant's true probability of win. This is not true in a model with casual bettors causing the bias by betting too evenly across racing entrants. Because casual bettors bet too evenly, the direction of the bias will depend on whether the racing entrant has an above (or below) average probability given the number of entrants in the race. They indeed find this to be the case, that a racing entrant with, say, a 1/7 probability of win would be overbet with six entrants, but underbet with eight

entrants.¹⁰ Returning to the bottom row in Table 4, these rotation points (the points at which it switches from entrants being overbet to underbet) are presented, along with the value predicted by the information model ($1/N$ according to a null prior). They coincide almost perfectly with the predictions of the information model. The risk model, on the other hand, has no prediction as to the rotation point, which would depend on the severity of the preference for risk. However, the risk model would unambiguously imply that the rotation point, whatever it is, should be *the same across all bets* existing at the track, given that it is a reflection of risk-preference related to the probability structure of the bets. This can clearly be rejected in the data.

5. CAN THE MODEL EXPLAIN THE BIASES IN OTHER MARKETS?

No firm understanding of why sports betting markets contain the opposite bias from most racetracks is known. According to our theory, however, one possible explanation is that Vegas sports betting markets tend to be dominated by a higher proportion of serious or regular bettors (and fewer casual bettors) than are racetracks. In football and basketball, the bets studied have been the point spread and over/under (total game score) bets. These bets do differ somewhat from racetrack bets made relative to market odds, however. Bets on baseball, like in racetrack wagering, are made relative to odds, and because the opposite favorite-longshot bias is found in baseball as well, a difference in the types of bets cannot be responsible for causing an opposite bias in sports wagering than the one present in racetrack wagering. In addition, the regular favorite-longshot bias in racetrack wagering has been found by Gabriel and Marsden (1990, 1991) in the UK where bets are made against a bookie-based fixed odds system, rather than the pure market odds structure as in the U.S.

A systematic difference in the betting behavior of serious and casual bettors might also explain why some racetrack studies find opposite favorite-longshot biases in contrast to most other studies on racetrack betting. The higher betting per person at the Hong Kong track suggests that it is dominated more heavily by serious bettors than the average U.S. racetrack used in other studies, and similarly, one might argue that the smaller Class II racetrack draws a relatively higher proportion of serious bettors than the tracks used in other studies.

If a difference between the behavior of serious and casual bettors is accountable for the flipping of a favorite-longshot bias to its opposite, it might also suggest that the stock market findings are due to a relatively high proportion of serious to casual investors. An uninformed casual investor will frequently invest in mutual funds, or

¹⁰Finally, another degree to which data can be broken down is the time of day that the bet is placed. After controlling for the quality level of the race, later evening races become more heavily dominated by serious bettors (as casual bettors are more likely to leave the track early). As expected, the proportion of bets on the favorite relative to the longshot grows throughout the evening as casual bettors leave the track.

somehow otherwise allow an informed middleman to manage his or her portfolio. Possible explanations for the stock market bias proposed by previous authors have been that investors overreact to information, that the bias is due to risk-preferences (an explanation that includes the small/large firm risk differential explanation), and the presence of uninformed noise traders in the market.

While stock market research has found an opposite favorite-longshot bias in stocks and earnings forecasts, some other studies have found a regular favorite-longshot bias in options pricing. For example, the longshot or shorter maturity options have been found to be overpriced by Rubinstein (1985, 1987), Shastri and Wethyavivorn (1987), and Tompkins, Ziemba, and Hodges (2008) who survey other references. Could the much more complex calculations necessary for options pricing create the different bias? Thaler (1992) has suggested that differences between biases found across different types of racetrack bets might be due to differing complexities of the bets. In addition, Elton et al. (1982) show a similar example regarding complexity affecting behavior in the stock market with respect to futures transactions. The idea is that, to some degree, as bets become more complex, all bettors become less informed relative to the level necessary to obtain accurate probability estimates. That is the essence of the place and show system, and other exotic wagers, in Hausch, Ziemba, and Rubinstein (1981) and Ziemba and Hausch (1986).

The general theory presented in this chapter suggests a starting point for a unified model of these anomalies. This model is a three-group heterogeneous bettor model. The first group—the casual bettors—generally tends to overbet longshots (and underbet favorites) with their strategy of using selected, and perhaps highly unreliable, signals to select their bets (due to their significantly limited information set). The second group—the serious bettors—generally tends to overbet favorites (and underbet longshots) with their strategy of attempting to use their knowledge and human capital to pick the winners. The third, and final, group—the arbitrageurs—are individuals who invest resources to both discover (usually through intensive data analysis) and exploit profit opportunities, but only when these opportunities generate a higher than normal (risk adjusted) rate of return from their resource investment.

In addition to allowing for heterogeneous bettors, our theory also incorporates heterogeneous bets, differing in their level of complexity. Employing Bayesian logic, a change in the complexity of a bet has a mathematically equivalent effect to changing the information level of the bettor. On more complex bets, even serious bettors become relatively more uninformed, causing them—just like the casual bettors—to bet too evenly across possible outcomes. Thus markets with simple bets and/or dominated by serious bettors will tend toward an opposite favorite-longshot bias (like sports betting), while markets with complex bets and/or dominated by casual bettors will tend toward a regular favorite-longshot bias (like exotic bets). The presence of arbitrageurs can eliminate these divergences from efficiency when they are identified and exploited, but at a cost that makes this type of activity unprofitable in most narrow and limited betting markets.

6. CONCLUSION

The ideas presented in this chapter lay a framework for a unified theory of deviations from market efficiency. If a unified model can explain these anomalies, it may allow great insights into human behavior and further the refinement of our economic models. It appears that information, complexity, and human analytical capital are jointly responsible. Highly informed and educated bettors faced with simple situations appear to systematically overvalue favorites and undervalue longshots—they attempt to pick winners better than everyone else. As complexity increases, or the number of less informed (or less able) bettors increases, this bias moves toward the other direction, eventually overtaking the underlying opposite bias.

Opportunities to further test our model abound in other areas where these inefficiencies have been found to exist.¹¹ In addition, the fundamental model presented here has broad, sweeping implications for the efficiency of product markets where consumers are faced with uncertainty with regard to things such as product quality or price differences across firms, and for public sector economics in models of rent seeking, voter information, and electoral competition.

References

- Ahlers, D., and J. Lakonishok. 1983. A Study of Economists' Consensus Forecasts, *Management Science* 29, 1113–1125.
- Ali, M. M. 1977. Probability and Utility Estimates for Racetrack Bettors, *Journal of Political Economy* 85, 803–815.
- Ali, M. M. 1979. Some Evidence on the Efficiency of a Speculative Market, *Econometrica* 47, 387–392.
- Asch, P., B. G. Malkiel, and R. E. Quandt. 1982. Racetrack Betting and Informed Behavior, *Journal of Financial Economics* 10, 187–194.
- Asch, P., B. G. Malkiel, and R. E. Quandt. 1984. Market Efficiency in Racetrack Betting, *Journal of Business* 57, 165–175.
- Asch, P., B. G. Malkiel, and R. E. Quandt. 1986. Market Efficiency in Racetrack Betting: Further Evidence and a Correction, *Journal of Business* 59, 157–160.
- Bremer, M. A., and R. J. Sweeney. 1991. The Information Content of Extreme Negative Rates of Return, *Journal of Finance* 46, 747–754.
- Busche, K., and C. D. Hall, 1988. An Exception to the Risk-Preference Anomaly, *Journal of Business* 61(3), 337–346.
- Brown, K. C., and W. Van Harlow. 1988. Market Overreaction: Magnitude and Intensity, *Journal of Portfolio Management* 14, 6–13.
- Camerer, C. 1998. Can Asset Markets be Manipulated? A Field Experiment with Racetrack Betting, *Journal of Political Economy* 106(3), 457–482.
- Chan, K. C. 1988. On the Return of the Contrarian Investment Strategy, *Journal of Business* 61(2), 147–163.
- De Bondt, W. F. M., and R. H. Thaler. 1985. Does the Stock Market Overreact? *Journal of Finance* 40, 793–805.
- De Bondt, W. F. M., and R. H. Thaler. 1987. Further Evidence on Investor Overreaction and Stock Market Seasonality, *Journal of Finance* 42, 557–581.

¹¹Try asking someone the name of the longest book they have ever read. Then ask them to try and remember (without looking) how many pages it was. Our model predicts that they will be very likely to overestimate the true number of pages.

- De Bondt, W. F. M., and R. H. Thaler. 1989. A Mean-Reverting Walk Down Wall Street, *Journal of Economic Perspectives* 3, 189–202.
- De Bondt, W. F. M., and R. H. Thaler. 1990. Do Security Analysts Overreact? *American Economic Review Papers and Proceedings* 80(2), 52–57.
- Elton, E., M. Gruber, and J. Rentzler. 1982. Intra Day Tests of the Efficiency of the Treasury Bills Futures Market, Working paper No. CSFM 38, Columbia University Business School, New York.
- Figlewski, S. 1979. Subjective Information and Market Efficiency in a Betting Model, *Journal of Political Economy* 87(1), 75–88.
- Friedman, M., and L. J. Savage. 1948. The Utility Analysis of Choices Involving Risk, *Journal of Political Economy* 56(4), 279–304.
- Gabriel, P. E., and J. R. Marsden. 1990. An Examination of Market Efficiency in British Racetrack Betting, *Journal of Political Economy* 98, 874–885.
- Gabriel, P. E., and J. R. Marsden. 1991. An Examination of Market Efficiency in British Racetrack Betting: Errata and Corrections, *Journal of Political Economy* 99, 657–659.
- Gander, J., R. Zuber, T. O'Brien, and B. Russo. 1988. Testing Rationality in the Point Spread Betting Market, *Journal of Finance* 43(4), 995–1008.
- Gander, J., R. A. Zuber, and R. S. Johnson. 2001. Searching for the Favorite-Longshot Bias Down Under: An Examination of the New Zealand Pari-Mutuel Betting Market, *Applied Economics* 33(13), 1621–1629.
- Golec, J., and M. Tamarkin. 1998. Bettors Love Skewness, Not Risk, at the Horse Track, *Journal of Political Economy* 106(1), 205–225.
- Gramm, M., and D. H. Owens. 2005. Determinants of Betting Market Efficiency, *Applied Economics Letters* 12(3), 181–185.
- Griffith, R. M. 1949. Odds Adjustments by American Horse-Race Bettors, *American Journal of Psychology* 62, 290–294.
- Hanson, R., R. Oprea, and D. Porter. 2006. Information Aggregation and Manipulation in an Experimental Market, *Journal of Economic Behavior & Organization* 60(4), 449–459.
- Hausch, D. B., W. T. Ziemba, and M. Rubinstein. 1981. Efficiency of the Market for Racetrack Betting, *Management Science* 27(12), 1435–1452.
- Howe, J. S. 1986. Evidence on Stock Market Overreaction, *Financial Analysts Journal* 41, 74–77.
- Hurley, W., and L. McDonough. 1995. A Note on the Hayek Hypothesis and the Favorite-Longshot Bias in Parimutuel Betting, *American Economic Review* 85(4), 949–955.
- Johnson, R. S., and J. E. Pawlukiewicz. 1992. The Efficiency of the Over-Under Betting Market for NBA Basketball Games, *Journal of Economics* 18, 97–100.
- Kirzner, I. M. 1973. *Competition and Entrepreneurship*. University of Chicago Press, Chicago, IL.
- Kochman, L., and R. Badarinathi. 1992. Pricing Inefficiencies in the Football Betting Market: Anthology, *Atlantic Economic Journal* 20(3), 104.
- Lacey, N. J. 1990. An Estimation of Market Efficiency in the NFL Point Spread Betting Market, *Applied Economics* 22(1), 117–129.
- Leeson, P. T., C. J. Coyne, and P. J. Boettke. 2006. Does the Market Self-Correct? Asymmetrical Adjustment and the Structure of Economic Error, *Review of Political Economy* 18(1), 1–12.
- Lehmann, B. N. 1990. Fads, Martingales, and Market Efficiency, *Quarterly Journal of Economics* 105, 1–28.
- Losey, R. L., and J. C. Talbott. 1980. Back on Track with the Efficient Markets Hypothesis, *Journal of Finance* 35, 1039–1043.
- McGlothlin, W. H. 1956. Stability of Choices among Uncertain Alternatives, *American Journal of Psychology* 69, 604–615.
- Piron, R., and L. R. Smith. 1995. Testing Risklove in an Experimental Racetrack, *Journal of Economic Behavior & Organization* 27(3), 465–474.
- Quandt, R. E. 1986. Betting and Equilibrium, *Quarterly Journal of Economics* 101(1), 201–208.
- Rubinstein, M. 1985. Nonparametric Tests of Alternative Option Pricing Models, *Journal of Finance* 40, 455–480.
- Rubinstein, M. 1987. Derivative Assets Analysis, *Journal of Economic Perspectives* 1, 73–93.
- Sauer, R. D. 1988. Hold Your Bets: Another Look at the Efficiency of the Gambling Market for National Football League Games: Comment, *Journal of Political Economy* 96(1), 206–213.

- Shastri, K., and K. Wethyavivorn. 1987. The Valuation of Currency Options for Alternative Stochastic Processes, *Journal of Financial Research* 10, 282–293.
- Snyder, W. W. 1978. Horse Racing: Testing the Efficient Markets Model, *Journal of Finance* 33(4), 1109–1118.
- Snowberg, E., and J. Wolfers. 2008. Examining Explanations of a Market Anomaly: Preferences or Perceptions? in D. B. Hausch and W. T. Ziemba (eds.), *Handbook of Sports and Lottery Markets*. North-Holland, Amsterdam, pp. 103–133.
- Sobel, R. S., and S. T. Raines. 2003. An Examination of the Empirical Derivatives of the Favourite-Longshot Bias in Racetrack Betting, *Applied Economics* 35(4), 371–385.
- Strumpf, K., and P. Rhode. 2004. Historical Presidential Betting Markets, *Journal of Economic Perspectives* 18(2), 127–141.
- Swidler, S., and R. Shaw. 1995. Racetrack Wagering and the “Uninformed Bettor”: A Study of Market Efficiency, *The Quarterly Review of Economics and Finance* 35(3), 305–314.
- Thaler, R. H., and W. T. Ziemba. 1988. Anomalies: Parimutuel Betting Markets: Racetracks and Lotteries, *Journal of Economic Perspectives* 2(2), 161–174.
- Thaler, R. H. 1992. *The Winner's Curse: Paradoxes and Anomalies of Economic Life*. Princeton University Press, Princeton, NJ.
- Tompkins, R. G., W. T. Ziemba, and S. D. Hodges. 2008. The Favorite-Longshot Bias in S&P 500 and FTSE 100 Index Futures Options: The Return to Bets and the Cost of Insurance, in D. B. Hausch and W. T. Ziemba (eds.), *Handbook of Sports and Lottery Markets*, North-Holland, Amsterdam, pp. 161–180.
- Tuckwell, R. H. 1983. The Thoroughbred Gambling Market: Efficiency, Equity, and Related Issue, *Australian Economic Papers* 22, 106–118.
- Weitzman, M. 1965. Utility Analysis and Group Behavior: An Empirical Study, *Journal of Political Economy* 73(1), 18–26.
- Wolfers, J., and E. Zitzewitz. 2004. Prediction Markets, *Journal of Economic Perspectives* 18(2), 107–126.
- Woodland, L. M., and B. M. Woodland. 1994. Market Efficiency and the Favorite-Longshot Bias: The Baseball Betting Market, *Journal of Finance* 49(1), 269–279.
- Woodland, L. M., and B. M. Woodland. 2001. Market Efficiency and Profitable Wagering in the National Hockey League: Can Bettors Score on Longshots? *Southern Economic Journal* 67(4), 983–995.
- Woodland, L. M., and B. M. Woodland. 2003. The Reverse Favourite-Longshot Bias and Market Efficiency in Major League Baseball: An Update.” *Bulletin of Economic Research* 55(2), 113–123.
- Ziemba, W. T., and D. B. Hausch. 1986. *Betting at the Racetrack*. San Luis Obispo, CA. Dr. Z Investments, Inc.
- Zuber, R. A., J. M. Gandar, and B. D. Bowers. 1985. Beating the Spread: Testing the Efficiency of the Gambling Market for National Football League Games, *Journal of Political Economy* 93(4), 800–806.

CHAPTER 9

The Favorite-Longshot Bias in S&P 500 and FTSE 100 Index Futures Options: The Return to Bets and the Cost of Insurance

Robert G. Tompkins

Hochschule für Bankwirtschaft, Germany.

William T. Ziemba

Sauder School of Business, UBC, Vancouver, Canada, Mathematical Institute, Oxford University, and ICMA Centre, University of Reading, UK.

Stewart D. Hodges

School of Business, University of Warwick, Coventry, UK.

1. Introduction	162
2. Methodology	166
3. Results	170
3.1. Results for Quarterly Options on Stock Index Futures	174
3.2. Results for Monthly Options on Stock Index Futures	175
4. Conclusions	178
References	179

Abstract

This chapter examines whether the favorite-longshot bias that has been found in gambling markets (particularly horse racing) applies to options markets. We investigate this for all options on the S&P 500 futures and the FTSE 100 futures for the 17+ years from March 1985 to September 2002. Calls on the S&P 500 with both three months and one month to expiration display a relationship between probabilities and mean returns that are very similar to the favorite bias in horse racing markets. There are slight profits from deep in-the-money and at-the-money calls on the S&P 500 futures and increasingly greater losses as the call options are out-of-the-money. For three-month and one-month calls on the FTSE 100 futures, the favorite bias is not found, but a significant longshot bias has existed for the deepest out-of-the-money options. For the put options on both markets, and for both three-month and one-month horizons, investors overpay for all put options as an expected cost of insurance to protect against downside risk. The patterns of mean returns is analogous to the favorite-longshot bias in racing markets.

JEL Classifications: C15, G13

Keywords: Longshot bias, gambling, option prices, implied volatilities

1. INTRODUCTION

Griffith (1949), McGlothlin (1956), Snyder (1978), Ali (1979), and others have documented a favorite-longshot bias in racetrack betting.¹ High probability–low payoff gambles have high expected value and low probability–high payoff gambles have low expected value. For example, a 1/10 horse having more than a 90% chance of winning has an expected value of about \$1.03 per \$1 bet, whereas a 100/1 horse has an expected value of about 14 ¢. The favorite-longshot bias exists in other gambling markets such as sports betting; see Hausch et al. (1994) for a survey of results.

In Ziemba and Hausch (1986), the expected return per dollar bet versus the odds levels are studied for over 300,000 horse races. The North American public underbets favorites and overbets longshots. This bias has appeared for many years across all sizes of racetrack betting pools. The effect of these biases is that for a given fixed amount of money bet, the expected return varies with the odds level; see Figure 1. For bets on extreme favorites, there is a positive expected return. For all other bets, the expected return is negative. The favorite-longshot bias is monotonic across odds or, equivalently, the probability of winning and the drop in expected value is especially large for the

¹While the horse racing favorite-longshot bias is quite stable and pervasive, there exist exceptions in Asian racetrack markets (Busche and Hall, 1988; and Busche, 1994). The favorite-longshot bias literature is surveyed in Hausch et al. (1994, 2008) where many papers are reprinted including the early studies of Griffith (1949) and McGlothlin (1956). See also the survey of Sauer (1998). Recent papers consistent with the usual bias are Hurley and McDougough (1996), Sobel and Raines (2003), and Ottaviani and Sørensen (2003) plus the chapters in this volume.

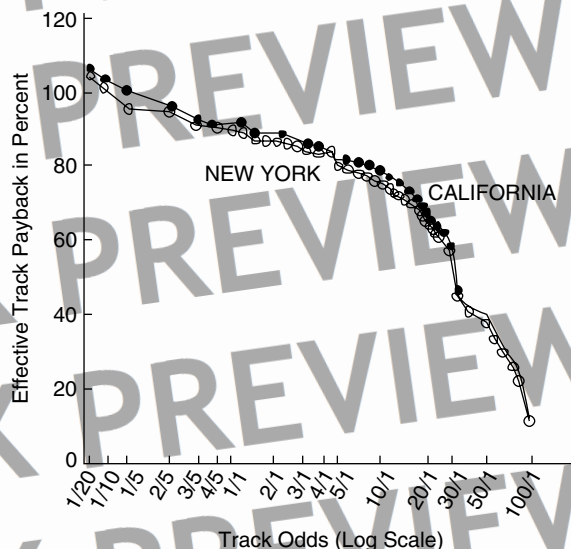


FIGURE 1 The effective track payback less breakage for various odds levels in California and New York for 300,000 plus races over various years and tracks.
Source: Ziemba and Hauseh (1986).

lower probability horses. The effect of differing track take/transaction costs is seen in the California versus New York graphs.

Thaler and Ziemba (1988) suggest a number of possible reasons for this bias. These include bettors' overestimation of the chances that longshot bets will win as in Kahneman and Tversky (1979). Tversky and Kahneman (1983) argue that bettors might overweight small probabilities of winning when the potential payout is large (in calculating their utility). Bettors may derive utility simply from the hope associated with holding a ticket on a longshot, as it is more fun to pick a longshot to win over a favorite and this has more bragging rights. Transaction costs also play a role. Finally, they suggest that some bettors may choose horses for irrational reasons, such as the name of the horse. Other explanations are that the bias results from the complexity of the wagers and the information available to bettors and not from risk-preferences; see Sobel and Raines (2003). Ottaviani and Sørensen (2003), Hurley and McDonough (1996), Quandt (1996), and Shin (1991, 1992) provide theoretical models that attempt to explain the bias. See also the chapters in this volume, especially Ottaviani and Sørensen (2008). The reasons for the effect varying over time are a combination of several factors. These include: (1) the utility bettors gain from betting on longshots and the associated preferences over the skewness of returns; (2) the systematic tendency for bettors to overestimate the chances of low probability outcomes, and underestimate high probability outcomes, and (3) the type and information aspects of informed and noise bettors in the race involved. Sobel and Raines (2003) show that the bias is steeper for lower quality races compared to higher quality races even on the same day at the same track. Consistent with this, Ziemba and Hausch (1987) show that the bias for the

Kentucky Derby is much less than is typical. The 1986 graph in Figure 1 is more flat in 2007 until the tail dropoff for high odds horses. On this, see, in this volume, the chapters by Snowberg and Wolfers (2008) and Ziemba (2008).

Puts and calls on stock index futures represent leveraged short or long positions on the index and their behavior might have similar features to racetrack bets. Demand for options comes from both hedging and speculation. The primary use of put options is for hedging.

For the call options, the most obvious hedging demand is to sell them against existing holdings of equity. This covered call strategy tends to depress the price of (especially out-of-the-money) call options. If this were the sole mechanism for dealing in call options, this should result in an increase in the expected return for purchasers of out-of-the-money call options. Coval and Shumway (2001) considered the expected and actual returns for options on the S&P 500. They find that call options (in a fairly narrow range around the current underlying S&P 500 price) have a higher expected return relative to the underlying S&P 500 index market. While this result is consistent with a leverage effect (the beta of options being much larger than the beta of the S&P 500), the return remains less than what it should be if leverage were the sole factor. Coval and Shumway (2001) do not consider the investment in deep out-of-the-money call options on the S&P 500 as we do here.²

We find significant expected losses for such deep out of the money call options. This could be due to speculative activity similar to that for longshot horse race bets. However, Bollen and Whaley (2003) showed that buyer-initiated trading in index puts dominates the market. Because there are few natural counter-parties to these trades (apart from hedge funds), the implied volatilities of these options rise and the implied volatilities of the corresponding call options rise due to put-call parity. However, they show that the primary choice of buyer-initiated index put trading occurs for the nearest out-of-the-money put options. They also stated, “since portfolio insurers generally buy OTM puts rather than ITM puts,” this implies that relatively speaking the demand for in-the-money puts is less and given that they argue that option mispricing is due to supply and demand imbalances at different strike prices, then in-the-money puts would be relatively less expensive. By put-call parity, this implies that the costs of the out-of-the-money call options would be relatively less expensive and offer a higher return. Nevertheless, our results indicate that deep out-of-the-money call options are overpriced.

Rubinstein (1994) pointed out that the implied volatilities for options on the S&P 500 changed after the 1987 stock market crash with the prices of out-of-the-money put options rising and the prices of out-of-the-money call options falling (relative to the price of the at-the-money option). This implied volatility skew (or smile) effect has been an active area of research. Buraschi and Jackwerth (2001, p. 523) conclude, “returns on out-from-the-money options are driven by different economic factors

²During January 1990 to October 1995, which was the period of the Coval and Shumway (2001) analysis, the average underlying S&P 500 futures price was approximately 430. They examined puts with strikes 15 points below and calls with strikes 10 points above. This implies an average percentage difference of strike prices that were 3.5% below the current price for puts and 2.3% higher for calls. In this study, we examined all traded options with ranges $\pm 43.3\%$ of the current S&P 500 price.

than those relevant for at-the-money options.” We consider the returns for such deep away-from-the-money put options (as opposed to the near-the-money options considered by Coval and Shumway, 2001).

With market imperfections (such as transaction costs or other frictions that disallow riskless hedges to be constructed in continuous time) or incomplete markets, option prices are no longer uniquely determined by arbitrage, and may be determined (within limits) by supply and demand. Dumas et al. (1998) suggest that the behavior of market participants may be the reason for the existence of smiles. They state: “with institutional buying pressures for out-of-the-money puts and no naturally offsetting selling pressure, index put prices rise to a level where market makers are eventually willing to step in and accept the bet that the index level will not fall below the exercise price before the option’s expiration (i.e., they sell naked puts) . . . option series clienteles may induce patterns in implied volatilities, with these patterns implying little in terms of the distributional properties of the underlying index” (p. 21).

Figlewski (1989) suggests that volatility smiles exist because of the demands of option users. He suggests that the higher prices (and resulting higher implied volatilities) associated with out-of-the-money options exist because people simply like the combination of a large potential payoff and limited risk. He likens out-of-the-money options to lottery tickets with prices such that they embody an expected loss. Nevertheless, this does not dissuade some from purchasing them.³ This would suggest that investors might be acting irrationally. Poteshman and Serbin (2002) show that this is the case for the exercise of exchange-traded stock options. They conclude that the early exercise of American calls on stocks during the period of 1996–1999 was in many instances “clearly irrational without invoking any model or market equilibrium.” If investors act irrationally in this regard, it is also possible they also act irrationally when assessing the value of the option and could display similar irrational behavior to other speculative endeavors such as gambling.

We examine the returns from investing in call and put options on stock index futures markets and assess whether the mean returns are biased for high leverage situations, as they are in various betting markets. To test the hypothesis that options display such biases requires a sufficient number of independent observations in actively traded markets and a broad enough range of strike prices where such low probability options are quoted. We use stock index futures options data, as these markets have existed for a sufficiently long period of time to yield enough independent exercise cycles and the range of offered strike prices allow the entire probability spectrum to be spanned. These instruments may be dominated by institutional investors buying portfolio insurance (as suggested by Bollen and Whaley, 2003). Given that such speculative behavior may be more likely in option markets with more retail activity, it could be helpful to examine individual stock option markets in parallel. However, either such individual stock option markets may not have been as actively (and consistently) traded as the stock

³The purchase of overpriced out-of-the-money puts may be justified by the desire of investors to provide a smoother risk profile in volatile markets.

index options, yield sufficient independent expiration cycles,⁴ and often do not offer a sufficiently wide range of strike prices to examine extreme probability events, we exclude their analysis here. In any case, Bakshi et al. (2003) suggest that the exclusion of individual stock options in our study will not seriously affect our general conclusions. They consider the skew effect for individual equity options written on the S&P 100 (OEX) cash index and its 30 largest equity components. Using 1991–1995 data, they show that the risk-neutral index skews exist and are a consequence of risk aversion and fat tailed distributions. Given that they report that individual stock skews are flatter than the index skews, this suggests that the option pricing bias is not greater in markets with more retail customers (i.e., individual stock options) but is in fact less. These equity markets have less systematic risk premium than the index markets and that is reflected in the less steep skew.

As equity index option markets have a wider range of available strike prices and trade on a monthly expiration cycle, yielding more independent trials than for stock options, we restrict our analysis solely to these markets. We examine two markets that have slightly different levels of retail trading activity. Our first is the S&P 500 futures options market. According to the Marketing Department of the Chicago Mercantile Exchange (and from Large Position reports from the Commodity Futures Trading Commission) virtually all trading activity for options on the S&P 500 futures comes from institutional traders. For options on the Financial Times Stock Exchange (FTSE) 100 futures traded at the London International Financial Futures Exchange (LIFFE) there is more retail involvement. Press releases from LIFFE report that retail involvement in these options comprise up to 10% of the total volume (similar to that of the individual stock options traded on the LIFFE). This market will provide some insights into the impacts of non-professional trading on the favorite-longshot bias.

Section 2 presents data sources and the methodology for the transformation of option prices into odds, so that the results can be compared to the horse racing literature. Section 3 presents results for the S&P 500 and FTSE 100 options markets. Section 4 concludes.

2. METHODOLOGY

To investigate whether a favorite-longshot bias exists in option markets requires a transformation of option prices into odds. In the Black Scholes (1973) equation, $N(d_2)$ is the forward price of a digital option that pays \$1 if $F > X$. It is the (risk neutral) odds at which investors can bet on this event. For a put option, the digital that pays \$1 if $X < F$ is $N(-d_2)$. As with the racing studies, one must collect a large sample of independent events, determine the odds of certain events occurring, invest a fixed amount in each bet (say \$1), and examine the a posteriori payoff of that bet. A pool of bets with the same odds must be aggregated and the mean payoff returns calculated. Our data

⁴Individual Stock Options are only offered on a quarterly cycle. Given that Stock Index options have had monthly expirations since 1987, there are more independent observations to test the hypotheses.

is the publicly available settlement prices for the futures contracts and all call and put options on the S&P 500 and FTSE 100 index markets on those dates when the options had either exactly one month or three months to expiration.⁵ The period of analysis was the 17.5 years from March 1985 to September 2002 and yielded 69 independent quarterly observations for the S&P 500 and FTSE 100 futures.⁶ S&P 500 futures began trading on April 21, 1982 and puts and calls on the S&P 500 began trading on January 28, 1983. The early years had little volume and few strike prices. Hence, our dataset covers the vast bulk of options trading in the S&P 500. For the monthly observations (serial options), there were 187 independent observations for the S&P 500 futures and 124 observations for the FTSE 100 index options markets. The data were obtained from the Chicago Mercantile Exchange for the S&P 500 futures and options. These option contracts are American style options on futures. The data for the FTSE 100 futures and options were obtained from the LIFFE for the European style options on futures from 1992 to 2002 and from Gordon Gemmill for the American style options on futures prior to 1992. The interest rate inputs were obtained from the British Bankers Association (U.S. Dollar or British Pound LIBOR).

Monthly and quarterly data were used instead of daily data to ensure independence of the observations and final outcomes. We identified all expiration dates for all available options over the sample period. On that day, we recorded the settlement levels of the futures contract (the nearest to the expiration of the futures contract and possibly the cash index if that date was a simultaneous expiration of the futures and options contract), and all available option prices on this nearby futures contract that had either one month or three months to expiration.

Given that settlement prices were used, it was not necessary to conduct the standard filtering procedures such as butterfly arbitrages; see Jackwerth and Rubinstein (1996). However, we did remove all options with prices below 0.05 (as for a trade to take place the offer price must be at least 0.05). With 17 years of quarterly data, we had 69 quarterly observations in our analysis with an average of 39.1 available strike prices per observation for the options on the S&P 500 and 30.8 strikes for options on the FTSE 100. For the monthly expirations, the average number of strike prices available for the S&P 500 options was 39.0 and 28.6 for the FTSE 100.

The first step is to calculate a measure of the odds of options finishing in the money (analogous to the odds in horse racing). Since the options are American, the Barone-Adesi and Whaley (1987) approximation has been used to recover the implied volatilities, which have then been substituted into the Black (1976) formula to calculate the pseudo-European option probabilities [$N(d_2)$ and $N(-d_2)$]. For the European style

⁵The pit committee of the CME determines the settlement prices rather than by market transactions and this could impact our results (especially for OTM options). However, the actual price at the end of the trading day could be a bid, mid, or offer price. Given that our analysis considers the payoffs from purchasing options, if the actual price that could be dealt at was the bid or mid price (rather than the offer price we implicitly assume), the payoffs of the options would be reduced accordingly. Therefore, our estimates of the wealth relatives for buying OTM options are more likely to be overly optimistic.

⁶To examine the impact of the 1987 crash, we also analyzed the post-crash period of our dataset. The results were not materially different (apart from small reductions in the mean wealth relatives of out-of-the-money put purchases).

options on the FTSE 100, the Black (1976) implied volatilities were directly used.⁷ In all markets, the implied volatilities for each option were used to calculate the odds. To make a more consistent comparison with horse race betting, the premium for the options were expressed in forward value terms. Using Black's (1976) formula

$$C_{fv} = FN(d_1) - XN(d_2) \quad (1a)$$

$$P_{fv} = XN(-d_1) - FN(-d_2), \quad (1b)$$

where

$$d_1 = \frac{\ln\left(\frac{F}{X}\right) + \frac{1}{2}\sigma^2(T-t)}{\sigma\sqrt{(T-t)}}$$

and

$$d_2 = d_1 - \sigma\sqrt{(T-t)}.$$

As we only observe the current option prices C_{pv} and P_{pv} , we transform these to the results in Equations (1a) and (1b) by multiplying the observed prices by $e^{r(T-t)}$ (where r is the LIBOR interpolated between adjacent standard maturities as reported by the British Bankers Association on the observation date, t , and T is the expiration date).

The terminal payoffs of the options are

$$C_T = \text{MAX}(F_T - X, 0) \text{ and } P_T = \text{MAX}(X - F_T, 0), \quad (2)$$

respectively. We calculate the wealth relatives of the ratios of these to the initial option forward values: in the absence of risk premiums these would be expected to average to one.

An important issue in averaging them is how the wealth relative on each option should be weighted. In our data sample, the number of strikes available increases with time. We would therefore lose efficiency if we weighted all options equally, as this would correspond to investing increasing amounts over time, where, for a given day the returns on options at different strikes are not independent. Our first principle is therefore to weight each monthly or quarterly period equally, by investing a fixed amount of money (e.g., \$1) at each date.

To achieve the same investment amount for the alternative option contracts, the number of options purchased equals

$$Q_C = \$1/C_{fv} \text{ and } Q_P = \$1/P_{fv}, \quad (3)$$

respectively, for all calls and puts. Equation (3) suggests that for higher priced options (e.g., in-the-money), the quantity purchased will be small and for lower priced options

⁷To be strictly comparable with horse racing odds, we should calculate the cost of a digital option under the distribution implied from option prices. However, it is not clear which of the variety of parametric and non-parametric approaches for the determination of implied distributions would be most appropriate. In any case, this would introduce another possible source of error.

(e.g., out-of-the-money), the number of options purchased will be large. We interpret the in-the-money options as the favorites and the out-of-the-money options as the longshots.

Unlike horse racing, the (risk neutral) probabilities of payoff in the options markets are not expressed as odds but in a continuous probability range from 0% to 100% (and at random points). In horse racing, while the bets are expressed as odds, such bets actually represent a continuous probability range for all bets between discrete categories (and are rounded down). As examples, 9/5 bets cover all ranges from 1.80 to 1.99 to 1 and 5/2 bets covers all bets from 2.00 to 2.49 to 1.

To determine expected wealth relative at fixed “odds” levels [$N(d_2)$ or $N(-d_2)$], we use interpolation to estimate what strike and option price would apply. Within the range of “odds” that exist on a given day,⁸ we linearly interpolate the implied volatility between adjacent strikes. With each wealth relative estimated thus, we form a simple average of wealth relatives from non-overlapping periods, and can therefore easily perform significance tests.⁹

Standard significance tests (such as a one-tailed t -test) may be inadequate when the sample distribution is not normal. The holding period return distributions of options tend to be quite positively skewed, and particularly so for out-of-the-money options, and when a risk premium on the underlying increases (for calls) or reduces (for puts) the (objective) probability of exercise. Care is therefore needed in testing the significance of the mean wealth relative to any given null hypothesis. To address this, we conducted Monte Carlo simulations to obtain the distribution of the realized mean wealth relatives for samples of suitable sizes (60 for quarterly and 160 for monthly horizons).¹⁰ These simulations were done under Black Scholes assumptions, with and without a risk premium, and for one-month and three-month times to expiration. The confidence intervals obtained in this way were noticeably different from the t -test intervals that would have been applicable for a normal (or nearly normal) distribution.¹¹

⁸We only interpolate and do not extrapolate beyond the range of traded strikes.

⁹We compute mean expected payoffs for various odds (probability of finishing in the money) bets to be comparable to the racetrack literature. One could beta risk adjust these bets to possibly separate out risk from behavioral biases. Coval and Shumway (2001) show how to do this. They find that there are then negative expected returns from buying puts and calls. This is consistent with our story that long run expected profits accrue to option sellers rather than option buyers. However, the Shumway and Coval results show that if you risk adjust the expected return from buying in-the-money 3 month calls is most likely negative and not positive as shown in Figure 3.

¹⁰The Monte Carlo simulation entailed simulating 10,000 times the average wealth relative for 60 (quarterly) and 160 (monthly) payoffs for call and put options with $N(d_2)$ from 0.05 to 0.95 in 0.05 increments. The standard error of the wealth relative was determined and the appropriate confidence levels were determined. For the inclusion of the risk premium, the negative continuous dividend adjustment to the Merton (1973) model proposed in the following section was used to determine the ratio of the expected wealth relative compared to the Black Scholes (1973) option price.

¹¹Nevertheless, because the confidence intervals most affected are for the right hand tails of out-of-the-money options (which tended not to be observed empirically) the use of the simulated intervals makes little difference to our results.

3. RESULTS

The first step is to examine what the payoffs of call and put options would be under the Black Scholes (1973) model. Although the presence of a risk premium on the equity index does not affect the option valuation, it will affect the pattern of realized wealth relatives. When risk premiums exist (e.g., in equity markets, see Constantinides, 2002), the expected return for the investment in options will differ from the \$1 investment. Similar in spirit to Coval and Shumway (2001), we examined the expected theoretical returns for call and put options using the Black Scholes (1973) formula with no risk premiums and risk premiums of 2%, 4%, and 6%. This is done by using -2% , -4% , and -6% , respectively, as the continuous dividend rate, using the Merton (1973) dividend adjustment, in the Black Scholes formula. The ratios of the option prices are determined and plotted as a function of money. This can be seen in Figure 2 for call and put options. The calls lie above the \$1 investment and the puts lie below the \$1 investment.

Consistent with the theoretical results of Coval and Shumway (2001), who show that in a very general setting, call options written on securities with expected returns above the risk-free rate should earn expected returns that exceed those of the underlying security and put options should earn expected returns below that of the underlying security. They also show that under very general conditions, these divergent expected returns would be increasing with the strike price (degree of out-of-the-moneyness). With this guidance as to how we would expect option returns to behave as a function of the Black Scholes (1973) model with risk premiums, we can now assess the returns actually observed for options on the S&P 500 and FTSE 100 futures. The results appear in Tables 1 and 2 for three-month options. The call and puts options appear on the left-hand and right-hand sides, respectively. For both, the first column is the odds of finishing in the money as measured by $N(d_2)$ or $N(-d_2)$. The next column indicates the number

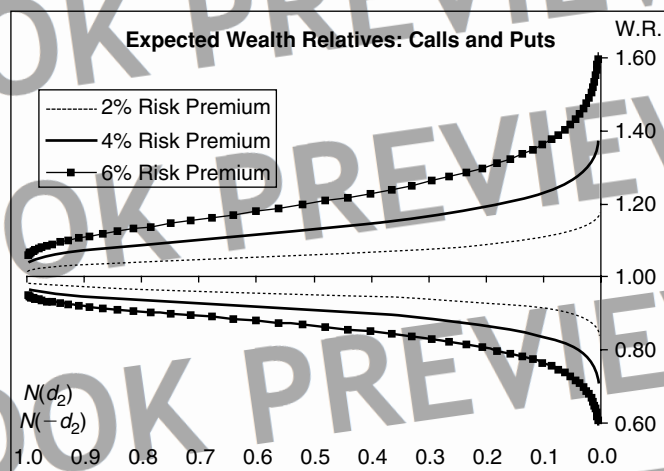


FIGURE 2 Expected wealth relatives for call and put options with alternative risk premium levels.

TABLE 1 Mean Return per \$1 Bet vs. Odds Levels: Three-Month Options on S&P 500 Futures, 1985–2002

Call options on the S&P 500 futures					Put options on the S&P 500 futures				
Odds (%)	#Obs	Average payoff	SD of payoff	<i>t</i> -test vs. \$1	Odds (%)	#Obs	Average payoff	SD of payoff	<i>t</i> -test vs. \$1
.95–1.00	47	1.0010	0.3204	0.02	.95–1.00	37	0.8998	0.4493	–1.35*
.90–.95	60	1.0561	0.4605	0.95	.90–.95	44	0.8662	0.5872	–1.50*
.85–.90	66	1.1231	0.5704	1.76**	.85–.90	50	0.8426	0.7265	–1.53*
.80–.85	67	1.1407	0.6990	1.66**	.80–.85	54	0.7937	0.8120	–1.86**
.75–.80	63	1.0938	0.5953	1.25	.75–.80	53	0.8137	0.8950	–1.51*
.70–.75	64	1.1366	0.7732	1.41*	.70–.75	51	0.7879	0.9979	–1.51*
.65–.70	62	1.1461	0.8648	1.33*	.65–.70	53	0.7702	0.9648	–1.73*
.60–.65	59	1.1311	0.9972	1.01	.60–.65	54	0.6215	1.0258	–2.70****
.55–.60	58	1.1727	1.1154	1.18	.55–.60	50	0.8225	1.2458	–1.01
.50–.55	54	0.9890	1.0410	–0.08	.50–.55	56	0.5807	1.1377	–2.76****
.45–.50	56	1.1365	1.3925	0.73	.45–.50	51	0.7344	1.4487	–1.31*
.40–.45	58	1.2063	1.6012	0.98	.40–.45	56	0.6785	1.5367	–1.57*
.35–.40	51	0.9770	1.7015	–0.10	.35–.40	56	0.4744	1.2383	–3.19****
.30–.35	54	0.9559	1.6041	–0.20	.30–.35	62	0.6257	1.6791	–1.76**
.25–.30	59	1.2923	2.7539	0.81	.25–.30	64	0.6316	1.8231	–1.62*
.20–.25	53	1.1261	2.5378	0.36	.20–.25	65	0.6426	1.9854	–1.45*
.15–.20	55	0.8651	2.0742	–0.48	.15–.20	64	0.6696	2.2441	–1.18
.10–.15	56	1.2262	3.6982	0.46	.10–.15	66	0.6602	2.6359	–1.05
.05–.10	53	1.5085	5.3370	0.69	.05–.10	66	0.6432	3.4256	–0.85
.00–.05	39	0.0123	0.1345	–44.89****	.00–.05	57	0.7525	5.6025	–0.33
All options	69	1.1935	2.4124	0.67	All options	69	0.6212	2.5247	–1.25

of observations we have for that particular 5% band (i.e., days for which \$1 could be invested). The average payoff for a \$1 investment in that particular option band appears next and is followed by the standard deviation of the option payoffs within the band. The final column is a modified one tailed *t*-test of the hypothesis that the mean return is equal to the initial investment of \$1 using

$$t = (\bar{X}_i - \$1\lambda) / (s_i / \sqrt{n}) \quad (4)$$

where

$$\bar{X}_i = \sum_{j=1}^n X_{i,j} / n \text{ and } X_{i,j} \quad (5)$$

$X_{i,j}$ is the wealth relative of the j th option in the i th continuous “odds” range, and λ is the equity risk premium. Critical levels for the t -test were determined using a Monte Carlo simulation. When the hypothesis is rejected at a 90% level or above, the t -statistic appears in **bold** print, “*”, “**”, “***”, or “****” on the t -statistic indicates that the level of significance is greater than the 90%, 95%, 97.5%, or 99% level, respectively.¹²

Figures 3 and 4 provide a graphical view of the mean returns, related to in-the-moneyness, in these markets. These are not plots of the data in Tables 1 and 2, but were calculated by our continuous interpolation method. This was done solely to allow

TABLE 2 Mean Return per \$1 Bet vs. Odds Levels: Three-Month Options on FTSE Futures, 1985–2002

Call options on the FTSE futures					Put options on the FTSE futures				
Odds (%)	#Obs	Average payoff	SD of payoff	t -test vs. \$1	Odds (%)	#Obs	Average payoff	SD of payoff	t -test vs. \$1
.95–1.00	32	1.0294	0.3215	0.52	.95–1.00	29	1.0019	0.5058	0.02
.90–.95	38	1.0485	0.4830	0.62	.90–.95	38	0.8995	0.6101	–1.02
.85–.90	41	1.1025	0.5901	1.11	.85–.90	36	0.8564	0.7274	–1.19
.80–.85	43	1.1033	0.7033	0.97	.80–.85	37	0.9628	0.8862	–0.25
.75–.80	44	0.9531	0.6601	–0.47	.75–.80	40	0.9709	0.9221	–0.20
.70–.75	49	0.9473	0.7491	–0.49	.70–.75	37	0.9201	1.0829	–0.45
.65–.70	47	1.1151	1.0764	0.73	.65–.70	40	1.0430	1.1861	0.23
.60–.65	49	0.8999	0.7903	–0.89	.60–.65	43	0.8264	1.1006	–1.03
.55–.60	44	1.1142	1.1296	0.67	.55–.60	38	0.9276	1.3428	–0.33
.50–.55	45	0.9505	1.2324	–0.27	.50–.55	39	0.8525	1.3050	–0.71
.45–.50	44	1.0148	1.1783	0.08	.45–.50	48	0.8615	1.5273	–0.63
.40–.45	41	0.8594	1.1062	–0.81	.40–.45	43	0.8764	1.7370	–0.47
.35–.40	43	1.1381	1.8821	0.48	.35–.40	48	0.7311	1.4967	–1.25
.30–.35	43	0.6177	1.1931	–2.10***	.30–.35	44	1.0169	2.2145	0.05
.25–.30	47	1.0396	2.1356	0.13	.25–.30	53	0.7216	2.2611	–0.90
.20–.25	38	0.8813	1.9081	–0.38	.20–.25	49	0.6252	1.9079	–1.37
.15–.20	0	0.4773	1.3779	–2.40***	.15–.20	48	1.0081	3.3628	0.02
.10–.15	42	0.9025	2.6841	–0.24	.10–.15	46	0.4131	1.9507	–2.04***
.05–.10	37	0.1421	0.7891	–6.60****	.05–.10	44	0.3600	2.2526	–1.88**
.00–.05	35	0.1877	1.1102	–4.32****	.00–.05	38	0.0893	1.0420	–5.39****
All options	70	0.9983	1.4668	–0.01	All options	70	0.6016	1.6203	–2.05***

¹²The confidence intervals used were based on a risk premium of 1.75% per quarter, which was the average realized risk premium for the two stock index markets over the period of our analysis. Thus, λ in Equation (4) is 1.0175 for quarterly options and 1.00583 for monthly options.

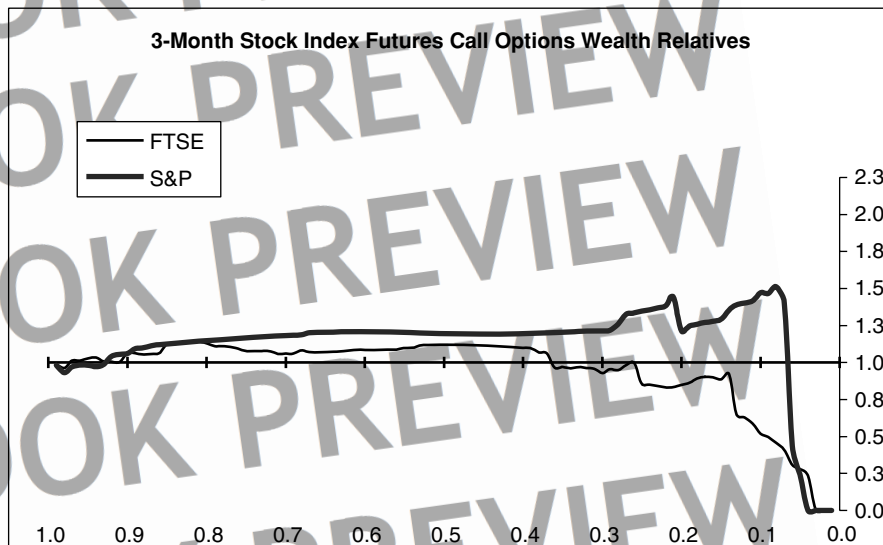


FIGURE 3 Mean return per dollar bet vs. odds levels: three-month stock index calls, 1985–2002.

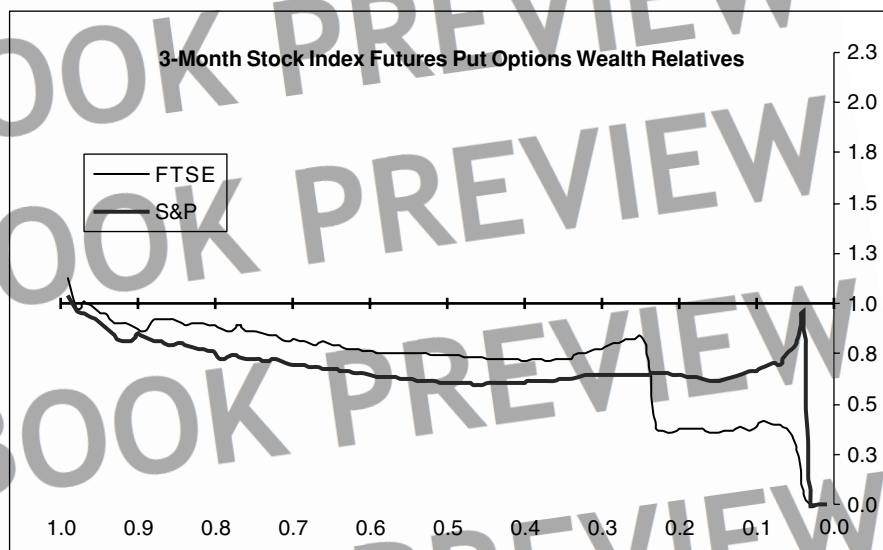


FIGURE 4 Average return per dollar bet vs. odds levels: five-month stock index puts, 1985–2002.

continuous, smooth curves to be drawn (and does not alter our interpretation of the results for the ranges in Tables 1 and 2).

3.1. Results for Quarterly Options on Stock Index Futures

For the call options on the S&P 500 futures, we find a similar favorite-longshot bias as in horse racing. The deep in-the-money call options in the probability ranges of 65% to 90% have a mean return of \$1.058. For the remaining ranges from 5% to 65%, we cannot reject the hypothesis that the return is significantly different from the \$1 investment. For the deepest out-of-the-money calls (0–5%) the mean returns are only 1.23 cents per dollar invested. We reject the hypothesis of an expected return of \$1 for the lowest 5% at a 99% level or above. This result supports the hypothesis of Figlewski (1989) that out-of-the-money call options are seen by investors as lottery tickets and investors overpay for deep out-of-the-money call options on the S&P 500 futures. Thus, the literature on “excessive optimism” in the assessment of risky situations may apply here; see Kahneman and Tversky (1979) and Tversky and Kahneman (1983).

For the call options on the FTSE 100 futures, for call options in the range between 80% to 100% probabilities, there is no significant difference between the expected payoff and the initial \$1 investment. Likewise, for most of the range from 35% to 80%, we cannot reject the hypothesis that the return is significantly different from the \$1 investment. However, for most of the out-of-the-money calls with probabilities less than 35%, we reject the hypothesis of an expected return of \$1 at above a 99% confidence level.

The put options on both the FTSE 100 and S&P 500 futures (essentially) all have negative mean returns. Moreover, the mean payoff is decreasing as the probabilities decrease, analogous to the horse racing favorite-longshot bias. This is also consistent with the contentions of Rubinstein and Jackwerth (1996), Dumas et al. (1998), and Bollen and Whaley (2002) that investors view put options as insurance policies and are willing to accept an expected loss to protect their holdings of equity against downside risk losses. To provide a clearer comparison between our results and those of Ziemba and Hausch (1986), the figures use similar axes: probabilities equal the reciprocal of the odds plus one. This can be seen for sets of stock index options in Figures 3 and 4.

This allows direct comparison to Figure 2, that presents the theoretical relationship between an option’s expected returns and risk premiums. If risk premium was causing call options returns to return more than the \$1 investment, we would expect Figure 3 to resemble the upper portion of Figure 2. When the returns are expressed as wealth relatives, out-of-the-money options offer a lower rate of return—exactly the opposite of what we expect. Therefore we conclude that the mechanism at work is not the risk premium argument of Coval and Shumway (2001) but a favorite-longshot bias.

In Figure 3, in-the-money call options yield more than the \$1 invested in each option. This is not surprising, given the existence of a risk premium for the equity market.¹³ However, the overall pattern is surprising: we would expect all calls to offer a higher rate of return, and for this to increase as the odds lengthen, as in Figure 2. For put options on the stock index futures in Figure 4, the mean return tends to decrease, as the option

¹³In the probability ranges from 45% to 55%, our results are similar to those of Coval and Shumway (2001).

is further out of the money. This is more consistent with Figure 2, but still suggests some anomalous behavior. For the S&P 500 put option returns, it is the in-the-money ones that are significant, whereas for the FTSE 100 options, only the out-of-the-money options are significant. In all cases, the returns on the longshot options are much more variable than on the favorites. Thus a much larger deviation of the sample mean from one is required, for a given number of observations, in order to reject the hypothesis.

3.2. Results for Monthly Options on Stock Index Futures

An enlargement of the data for the index options occurs when one considers options on futures with monthly expirations. This also allows a comparison with the three-month terms to expiration discussed above. The results appear in Tables 3 and 4 for the one-month calls and puts for the S&P 500 futures and FTSE 100 futures, respectively.

TABLE 3 Mean Return per \$1 Bet vs. Odds Levels: One-Month Options on S&P 500 Futures, 1985–2002

Call options on the S&P 500 futures					Put options on the S&P 500 futures				
Odds (%)	#Obs	Average payoff	SD of payoff	t-test vs. \$1	Odds (%)	#Obs	Average payoff	SD of payoff	t-test vs. \$1
.95–1.00	187	1.0092	0.2506	0.50	.95–1.00	187	0.9792	0.4949	–1.87**
.90–.95	187	0.9938	0.3923	–0.22	.90–.95	187	0.9883	0.6677	–2.01***
.85–.90	187	1.0029	0.4877	0.08	.85–.90	187	0.9989	0.7746	–1.18
.80–.85	187	0.9796	0.5925	–0.47	.80–.85	187	0.9544	0.8778	–1.67*
.75–.80	187	1.0064	0.6762	0.13	.75–.80	187	0.9880	0.9814	–0.61
.70–.75	187	0.9346	0.7612	–1.18	.70–.75	187	0.9437	0.9879	–1.54
.65–.70	187	0.9693	0.8699	–0.48	.65–.70	187	0.9520	1.0734	–0.74
.60–.65	187	0.9656	0.9497	–0.50	.60–.65	187	0.9193	1.2257	–0.97
.55–.60	187	0.9196	1.0671	–1.03	.55–.60	187	0.8867	1.2654	–1.48*
.50–.55	187	0.9586	1.1004	–0.51	.50–.55	187	0.9217	1.4020	–0.72
.45–.50	187	0.8954	1.2820	–1.12	.45–.50	187	0.8146	1.4427	–1.54*
.40–.45	187	0.9204	1.3652	–0.80	.40–.45	187	0.9064	1.6557	–1.17
.35–.40	187	0.9671	1.5108	–0.30	.35–.40	187	0.7672	1.6412	–2.53***
.30–.35	187	0.8673	1.6712	–1.09	.30–.35	187	0.7987	1.9158	–1.64*
.25–.30	187	0.9927	1.8245	–0.05	.25–.30	187	0.7454	2.0390	–2.38***
.20–.25	187	0.7939	1.8764	–1.50*	.20–.25	187	0.6910	2.1639	–2.55***
.15–.20	187	0.9257	2.5402	–0.40	.15–.20	187	0.6101	2.3346	–3.28***
.10–.15	187	0.7585	2.8601	–1.15	.10–.15	187	0.5303	2.4037	–3.11****
.05–.10	187	0.6940	3.5704	–1.17	.05–.10	187	0.4039	2.3630	–3.52****
.00–.05	187	0.50958	3.9119	–1.71**	.00–.05	187	0.0508	0.9785	–14.49****
All options	187	0.9668	2.1085	–0.22	All options	187	0.5033	1.3827	–4.92****

TABLE 4 Mean Return per \$1 Bet vs. Odds Levels: One-Month Options on FTSE Futures, 1985–2002

Call options on the FTSE futures					Put options on the FTSE futures				
Odds (%)	#Obs	Average payoff	SD of payoff	t-test vs. \$1	Odds (%)	#Obs	Average payoff	SD of payoff	t-test vs. \$1
.95–1.00	123	0.9595	0.2694	–1.67*	.95–1.00	123	1.0105	0.4717	0.25
.90–.95	123	0.9719	0.4011	–0.78	.90–.95	123	0.9847	0.6005	–0.28
.85–.90	123	0.9596	0.5020	–0.89	.85–.90	123	1.0229	0.6874	0.37
.80–.85	123	0.9474	0.6020	–0.97	.80–.85	123	0.9235	0.7736	–1.10
.75–.80	123	0.9761	0.6480	–0.41	.75–.80	123	0.9760	0.9099	–0.29
.70–.75	123	0.8576	0.7525	–2.10**	.70–.75	123	1.0093	1.0292	0.10
.65–.70	123	0.9296	0.8458	–0.92	.65–.70	123	0.9501	1.0715	–0.52
.60–.65	123	0.8632	0.8191	–1.85**	.60–.65	123	0.8984	1.1686	–0.96
.55–.60	123	0.8866	1.0456	–1.20	.55–.60	123	0.9579	1.1831	–0.39
.50–.55	123	0.8295	0.9372	–2.02**	.50–.55	123	0.8033	1.2349	–1.77**
.45–.50	123	0.9129	1.2141	–0.80	.45–.50	123	0.8161	1.4092	–1.45*
.40–.45	123	0.7647	1.2268	–2.13**	.40–.45	123	0.9409	1.5550	–0.42
.35–.40	123	0.7588	1.1234	–2.38**	.35–.40	123	0.8699	1.6963	–0.85
.30–.35	123	0.8685	1.6097	–0.91	.30–.35	123	0.7072	1.7646	–1.84**
.25–.30	123	0.4707	1.1119	–5.28***	.25–.30	123	0.8041	2.0297	–1.07
.20–.25	123	0.7006	2.0045	–1.66**	.20–.25	123	0.5855	2.0360	–2.26**
.15–.20	123	0.4952	1.4297	–3.92***	.15–.20	123	0.5423	2.4428	–2.08**
.10–.15	123	0.4779	2.4364	–2.38***	.10–.15	123	0.5878	2.8156	–1.62*
.05–.10	123	0.4920	3.6893	–1.53*	.05–.10	123	0.4872	3.3026	–1.72**
.00–.05	123	0.3427	4.8288	–1.51*	.00–.05	123	0.2968	3.4337	–2.27**
All options	2,460	0.7926	2.0670	–4.98***	All options	2,460	0.6535	2.4630	–6.98***

For both the S&P 500 and FTSE 100 option markets, the deep in-the-money one-month calls have a mean wealth relative close to one. The further the options are out of the money, the lower the mean payoff, as shown in Figures 5 and 6 using our interpolation method to give returns for odds spaced at every 1%. The pattern is quite striking for both markets: the payoff decays monotonically and is similar to the racetrack longshot bias shown in Figure 1. However, the only cases of mean returns significantly below one are for the FTSE 100 options, for which there is more retail activity. We have the usual problem of the large measurement error in expected returns measured with limited observations over short horizons.

For all the put options, the pattern of mean returns for the one-month puts is extremely close to those found for the three-month put options. The deepest in-the-money puts pay on average the initial bet. Losses increase as the puts are further out of the money, displaying a similar longshot bias to Figure 1.

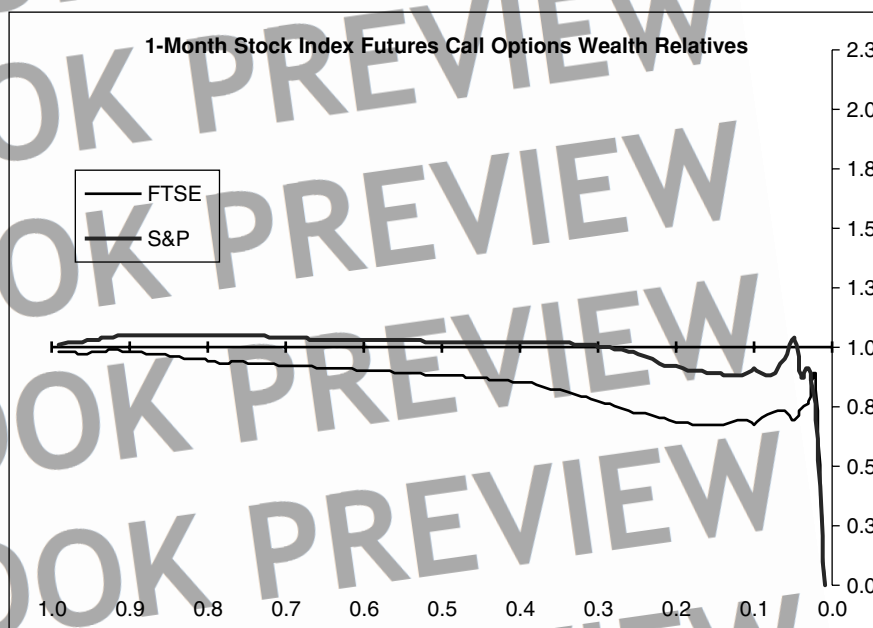


FIGURE 5 Mean return per dollar bet vs. odds levels: one-month stock index calls, 1985–2002.

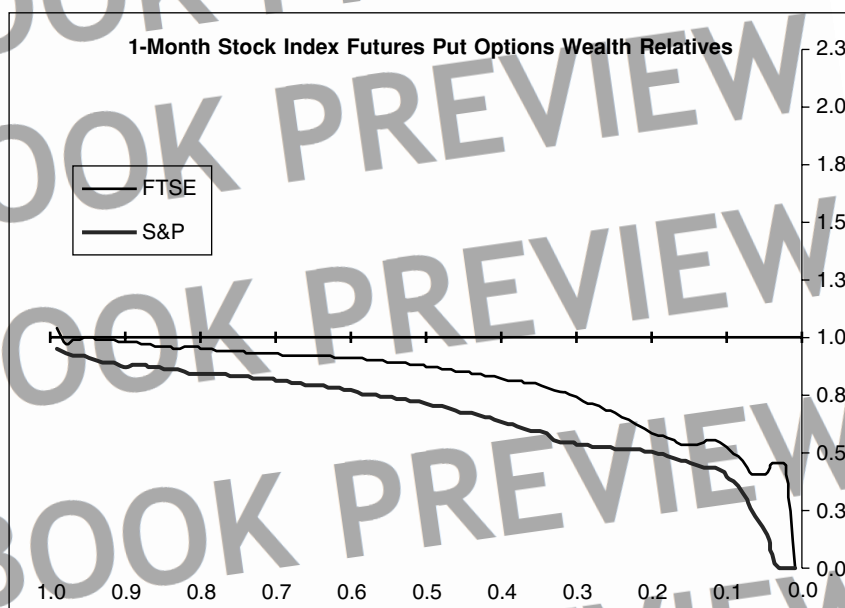


FIGURE 6 Mean return per dollar bet vs. odds levels: puts on stock index futures, 1985–2002.

Figures 5 and 6 show the mean return for one-month options on the S&P 500 and FTSE 100 futures across continuous probability bandwidths. In Figure 5, S&P 500 call options return the \$1 invested in each option on average. For the FTSE 100, the options tend to return slightly less than the initial investment. Therefore, there is no evidence of a favorite bias as the expected wealth relative is either equal to the initial investment or is statistically significantly below the initial investment. However, for both markets a significant longshot bias exists for the out-of-the-money calls in the probability ranges from 0% to 15% (the return is significantly less than the initial investment at a 90% level). However, the degree of the loss for the deep out-of-the-money calls in the 0% to 15% range is smaller than for the three-month options seen in Tables 1 and 2. This is not surprising as the expected losses occur at an almost steady rate over time, and we have only a third of the previous time to expiration.

The one-month put option returns appear in Figure 6. As with Figure 4 for the three-month put options, the mean return tends to decrease, as the option is further out of the money. For these options, the shape of the average return function is smoother than the three-month pattern. One possible explanation for this comes from Bollen and Whaley (2002). They indicate that the greatest concentration of trading in stock index put options is for put options with one month or less to expiration. Therefore, with more actively traded put options across the entire maturity spectrum, there is less need to interpolate.

4. CONCLUSIONS

The motivation for this research was to assess whether the favorite-longshot bias that has been found in horse racing and other gambling markets applies to options markets. The choice of stock index options was made due to a previous conjecture by Figlewski (1989) that deep OTM stock index call options are seen by investors as the equivalent of low cost/high payoff gambles and Dumas et al. (1998) that stock index put options are purchased at higher prices due to the need for insurance. We investigated the favorite-longshot bias for options on the S&P 500 Index Futures and FTSE 100 Index Futures for the 17+ years, March 1985 to September 2002.

The deep OTM index call options on the S&P 500 futures and FTSE 100 futures have negative mean returns. During the period of 1985–2002, the mean payback from the purchase of three-month call options in the probability range of 0–5% was less than 1.23 and 18.77 cents for every \$1 invested in the options for the S&P 500 and FTSE 100, respectively. Deep in-the-money three-month calls and one-month calls on the S&P 500 provide a mean return higher than the initial investment similar to the favorite-longshot bias in racetrack markets.

For the put options on the S&P 500 and FTSE 100, we find evidence consistent with the hypothesis of Dumas et al. (1998) that investors pay more for puts than they are subsequently worth. The degree of overpaying for these options increases monotonically as the probability of finishing in the money decreases. These results are also consistent with Coval and Shumay (2001), for options in the similar strike price range

they considered in their research. We present an empirical study of these index option markets rather than a theoretical analysis of why the biases exist. We know that the put skew bias steepens when there is a large drop in the underlying index and flattens when prices rise and that the shape of the bias varies over time. The one-month versus three-month figures document this. The bias in the puts is related to portfolio insurance protection against downside risks. Harvey and Siddique (2000) call this skewness preference. The deep out-of-the-money calls seem to be, as Figlewski (1989) argued, seen as lottery tickets which have low expected returns. The in-the-money and out-of-the-money three-month calls seem to reflect the equity risk premium as theoretically shown in Figure 2. The one-month calls in this range, which are highly price dependent on steep option time decay, are fairly priced for the S&P 500 and negatively priced for the FTSE 100. Investors' aversion to downside risk shown in overpriced puts, especially following the 1987 crash, is consistent with our data; see also Rubinstein (1994) and Bollen and Whaley (2002). This is similar to the pattern observed for the favorite-longshot bias, and is the expected cost of insurance.

The month call options on the S&P 500 and the FTSE 100 have similar patterns, but with magnitudes closer to one. Only for the in-the-money calls on the S&P 500 is a favorite bias found. The deep in-the-money calls on the FTSE 100 pay an average return very close to the initial bet. For the out-of-the-money options, there is a reduction in the expected return (like a longshot bias). However, this is not as extreme as for the three-month options, and only statistically significant for the FTSE 100 options.

References

- Ali, M. M. 1979. Some Evidence on the Efficiency of a Speculative Market, *Econometrica* 47, 387–392.
- Bakshi, G., N. Kapadia, and D. Madan. 2003. Stock Return Characteristics, Skew Laws, and the Differential Pricing of Individual Equity Options, *Review of Financial Studies* 16(1), 101–143.
- Barone-Adesi, G., and R. E. Whaley. 1987. Efficient Analytical Approximation of American Option Values, *Journal of Finance* 42, 301–320.
- Black, F. 1976. The Pricing of Commodity Contracts, *Journal of Financial Economics* 3, 167–179.
- Black, F., and M. Scholes. 1973. The Pricing of Options and Corporate Liabilities, *Journal of Political Economy* 81, 637–654.
- Bollen, N. P. B., and R. E. Whaley. 2004. Does Net Buying Pressure Affect the Shape of Implied Volatility Functions? *Journal of Finance* 59, 711–753.
- Buraschi, A., and J. C. Jackwerth. 2001. The Price of a Smile: Hedging and Spanning in Options Markets, *The Review of Financial Studies* 14, 495–527.
- Busche, K., and C. D. Hall. 1984. An Exception to the Risk-Preference Anomaly, *Journal of Business* 61, 337–346.
- Busche, K. 1994. Efficient Market Results in an Asian Setting, in D. B. Hausch, V. S. Lo, and W. T. Ziemba. (eds.), *Efficiency of Racetrack Betting Markets*. Academic Press, San Diego, CA, pp. 615–616.
- Constantinides, G. M. 2002. Rational Asset Prices, *Journal of Finance* 57, 1567–1591.
- Coval, J. D., and T. G. Shumway. 2001. Expected Option Returns, *Journal of Finance* 56, 983–1009.
- Dumas, B., J. Fleming, and R. E. Whaley. 1998. Implied Volatility Functions: Empirical Tests, *Journal of Finance* 53, 2059–2106.
- Griffith, R. 1949. Odds Adjustments by American Horse Race Bettors, *American Journal of Psychology* 62, 290–294.
- Figlewski, S. 1989. What Does an Option Pricing Model Tell Us About Option Prices, *Financial Analysts Journal* 45, 12–15.

- Harvey, C., and A. Siddique. 2000. Conditional Skewness in Asset Pricing Tests, *Journal of Finance* 55, 1263–1295.
- Hausch, D. B., V. S. Lo, and W. T. Ziemba. (Eds). 1994. *Efficiency of Racetrack Betting Markets*, Academic Press, San Diego, CA; Second Edition (2008), World Scientific, Singapore.
- Hurley, W., and L. McDonough. 1996. A Note on the Hayek Hypothesis and the Favorite-Longshot Bias in Parimutuel Betting, *American Economic Review* 85(4), 949–955.
- Jackwerth, J. C., and M. Rubinstein. 1996. Recovering Probability Distributions from Option Prices, *Journal of Finance* 51, 1611–1631.
- Kahneman, D., and A. Tversky. 1979. Choices, Values and Frames, *Econometrica* 47, 263–291.
- McGlothlin, W. 1956 Stability of Choices among Uncertain Alternatives, *American Journal of Psychology* 69, 604–615.
- Merton, R. 1973. The Theory of Rational Option Pricing, *Bell Journal of Economics* 4, 141–183.
- Ottaviani, M., and P. N. Sørensen. 2003. Late Informed Betting and the Favorite-Longshot Bias, Working paper, London Business School.
- Potesman, A. M., and V. Serbin. 2002. Clearly Irrational Financial Market Behavior: Evidence from the Early Exercise of Exchange Traded Stock Options, *Journal of Finance* 58, 37–70.
- Quandt, R. E. 1986. Betting and Equilibrium, *Quarterly Journal of Economics* 10(1), 201–207.
- Rubinstein, M., 1994. Implied Binomial Trees, *Journal of Finance* 49, 771–818.
- Sauer, R. D. 1998. The Economics of Wagering Markets, *Journal of Economic Literature* 36, 2021–2064.
- Shin, H. S. 1991. Optimal Betting Odds Against Insider Traders, *Economic Journal* 101(408), 1179–1185.
- Shin, H. S. 1992. Prices of Contingent Claims with Insider Trades and the Favorite-Longshot Bias, *Economic Journal* 102(411), 426–435.
- Sobel, R. S., and T. Raines. 2003. An Examination of the Empirical Derivatives of the Favorite-Longshot Bias in Racetrack Betting, *Applied Economics* 35, 371–385.
- Snyder, W. W., 1978. Horse racing: Testing the Efficient Markets Model, *Journal of Finance*, 33, 1109–1118.
- Thaler, R. H., and W. T. Ziemba. 1988. Anomalies—Parimutuel Betting Markets: Racetracks and Lotteries, *Journal of Economic Perspectives* 2, 161–174.
- Tversky, A., and D. Kahneman. 1983. Extensional vs. Intuitive Reasoning: The Conjunction Fallacy in Probability Judgment, *Psychological Review*, 90, 293–315.
- Ziemba, W. T. 2008. Efficiency of Racing, Sports, and Lottery Betting Markets, in D. B. Hausch and W. T. Ziemba (eds.), *Handbook of Sports and Lottery Markets*, North-Holland, Amsterdam, pp. 183–221.
- Ziemba, W. T., and D. B. Hausch. 1986. *Betting at the Racetrack*. Norris M. Strass, New York.
- Ziemba, W. T., and D. B. Hausch. 1987. *Dr Z's Beat the Racetrack*. William Morrow, New York.

DOWNLOAD FULL VERSION HERE: <https://bettingbooks.org/12>

PART IV: Weak Market Efficiency

Download More Sports Betting Books at: www.bettingbooks.org

DOWNLOAD FULL VERSION HERE: <https://bettingbooks.org/12>

This page intentionally left blank

Download More Sports Betting Books at: www.bettingbooks.org

CHAPTER 10

Efficiency of Racing, Sports, and Lottery Betting Markets

William T. Ziemba

Sauder School of Business, UBC, Vancouver, Canada, Mathematical Institute, Oxford University, and ICMA Centre, University of Reading, UK.

1. Introduction	184
2. Extent of Gambling in the U.S.	187
3. Racetrack Betting Markets	190
3.1. <i>Introduction to Racetrack Betting</i>	190
3.2. <i>Win Market</i>	190
3.3. <i>Place and Show Markets</i>	195
3.4. <i>Place and Show Probabilities</i>	196
3.5. <i>Optimal Capital Growth</i>	201
3.6. <i>Implementing the System and Empirical Results</i>	203
3.7. <i>Does the System Still Provide Profits?</i>	206
3.8. <i>Exotic Markets</i>	207
3.9. <i>Cross-Track Betting</i>	208
4. The Football Betting Market	209
5. The Basketball Betting Market	211
6. Lotteries	212
6.1. <i>Introduction to Lotteries</i>	212
6.2. <i>Inefficiencies with Unpopular Numbers</i>	214
References	217

Acknowledgments

This chapter is based on and updated from Hausch and Ziemba (1995). Special thanks go to Donald Hausch for his previous collaborations and comments on this update. Thanks go to Werner De Bondt, Victor Lo, and Raymond Sauer for their helpful comments and suggestions.

Abstract

This is a survey of efficiency in racing, sports, and lottery markets. The win market is efficient but exhibits a favorite-longshot bias. The place and show markets, which involve more possible finishes, allow inefficiencies by using the win probabilities. These biases are discussed for U.S. and Hong Kong markets. The Kelly capital growth criterion is useful to implement a model to exploit these inefficiencies. Exotic markets involve even more complex sets of bets. Finally, possible inefficiencies in cross-track betting are discussed. Football and basketball betting markets are largely efficient. Lotteries provide interesting markets with one way to potentially exploit them being the use of unpopular numbers.

Keywords: market efficiency, horserace betting, sports betting, lottery wagers

1. INTRODUCTION

Financial economists have long been interested in the efficiency of financial markets. Kendall (1953) examined the behavior of industrial share prices and spot prices for cotton and wheat. By analyzing serial correlations, prices appeared to follow a random walk. In the 1960s the focus was on defining efficiency and performing tests for a range of efficiency notions. Roberts (1967) defined weak, semi-strong, and strong form efficiencies as holding when stock market prices reflect all price information, all publicly available information, and all information, respectively. Most financial markets have been shown to be efficient in the weak form. The evidence for semi-strong is mixed; see Sung and Johnson (2008) for a survey of the racing literature. The strong form is largely inefficient. See Fama (1970) for a survey of this work in financial markets. The exceptions, termed anomalies, include seasonal patterns such as the small firm January effect, turn-of-the-month and year effects, holiday effect, day of week, time of day, the Value Line enigma, and cross-sectional regularities that apply to stocks with low price to earnings ratios or with earnings surprises, and so forth. See the surveys by Hawawini and Keim (1995, 2000), Ziemba (1994), and Keim and Ziemba (2000) for more details.

Fama (1991) updated his earlier survey. Tests for return predictability focus on forecasting returns using variables such as interest rates and dividend yields. Event studies formalize the semi-strong form idea by testing whether or not there are adjustments of prices to specific public announcements. Finally, the strong form concept is studied through tests for private information. The evidence is that future returns are predictable from past returns, dividend yields, and term structure variables. On the face of it, this is a violation of weak-form efficiency. But, as suggested by Roll (1977), since every test of efficiency must be a joint one with a maintained equilibrium hypothesis of price formation (e.g., the capital asset or arbitrage pricing models), this violation is confounded by the joint hypothesis problem of whether there is a rational variation over time in expected returns or systematic deviations from fundamental value. While

arguments can be made that increased returns may be occurring because of increased risk, which is difficult or impossible to measure accurately, there is very strong evidence that most or all of the gains in securities markets have occurred during the seasonally anomalous periods. Ritter and Chopra (1989) and Cadsby (1992) show, for example, that the only periods where risk as measured by the capital asset pricing model is rewarded with equity returns are precisely at the anomalous periods such as the day before holidays, the turn of the month, in the first two weeks of January for small stocks, and so on. Ariel (1987), Lakonishok and Smidt (1988), and Hensel et al. (2000) showed that essentially all the stock market's gains during the twentieth century in the U.S. occurred in the first half of the month. Event studies are more straightforward and less controversial, since they are able to provide more clear-cut evidence of the effect of new information. Regarding strong-form tests, there is considerable evidence that corporate insiders have private information that is not fully reflected in current prices.

Sports and betting markets are well suited for testing market efficiency and bettor rationality. This is because vast amounts of data are available, in the form of prices (for devising technical systems) and other information (for devising fundamental systems), and each bet has a specified termination point when its final asset value is determined. For rationality tests, markets with this latter property offer an advantage over markets, like securities markets, where the current value depends upon future events and current expectations of future values. Also, some wagering markets have characteristics that reduce the problematic nature of the aforementioned joint hypothesis test. For instance, Dana and Knetter (1994) note that point spread bets on National Football League games all have identical risk and return characteristics, as well as similar horizons. This allows a test of efficiency without specifying the bettors' utility functions.

The special properties of sports betting and lotteries might lead one to speculate that they are even more efficient than financial markets. However, there is another aspect to these markets that confounds the notion of rationality: for them to be offered, the average bettor must lose. Indeed, given the transaction costs involved in these markets (e.g., about 13–30% for horse racing and about 50% for lotteries), the average losses are large. This has not stopped the search for profitable wagering systems, though, and there are some notable successes. For example, Thorp (1961, 1962) demonstrated that card-counters can win playing blackjack. This survey of research on horse racing, sports betting on football and basketball, and lotteries reports numerous studies of efficiency in these markets. Several profitable systems are also described, though.¹ The continued success of these winning systems tends to be related to some complicating factor in its development or execution. For instance, the system may involve short odds and complex probability estimation (e.g., place and show wagering at the racetrack), it may rely on syndicates of bettors (e.g., cross-track horse race betting), it could require extremely long time horizons (e.g., lotteries), or extensive data collection and statistical work (e.g., fundamental handicapping systems for horse racing). The winning systems described are, of course, just a subset of the winning systems used in practice. The incentives to

¹Beyond the academic work surveyed here, evidence abounds of individuals who have successfully beaten the odds. See, for example, Akst (1989), Benter (1994, 2001), and Beyer (1978).

disclose details of a winning system may not be sufficient in some cases given that such an action typically reduces the system's profitability as others employ it. Finally, we also discuss optimal betting strategies for exploiting inefficiencies when they are present.

Sports betting and lotteries involve substantial transactions costs. Because it directly affects prices, the take—what the gambling establishment keeps for its operation—is properly accounted for in all of the analyses discussed in this survey. Another cost in these markets is for information (e.g., tip sheets at the racetrack). Costly information requires a redefinition of efficiency, one where prices are said to reflect information to the point where the cost of additional information just equals the benefit of acting on that information. But because they are difficult to measure, this survey ignores information costs (and other transaction costs beyond the take). Thus, our general findings of efficiency gain further support after the introduction of these other costs. In racing, the advent of rebates has driven the take from 13% to 30% for various bets to about 10% for large bettors. Just like shopping at a wholesale market for food, large bettors get a quantity discount. The rebaters and the large bettors share a discount given by the track for the signals (outcomes). The net effect of this is that when these discount bets are blended with the rest of the bets, the effective track take for the small bettor is more than the posted take. Rebates at a lower level are available to small bettors.

Figure 1 provides a taxonomy for the types of games that we discuss. Games are classified by: (1) whether the chance of winning is purely luck or can be influenced with skill; and (2) whether the payoff upon winning is predetermined or can be improved with skill. Luck-luck games allow no possibility of discovering a profitable strategy, and so as markets they are trivially efficient. On the other hand, there need not be a guarantee of efficiency for luck-skill games such as lotto (where we discuss a strategy of betting unpopular numbers, which does not affect the probability of winning but does affect the payoff upon winning) and skill-luck games (which are relatively uncommon). Blackjack, a skill-skill game, allows a profitable strategy. For horse racing, another skill-skill game, we review findings that certain forms of wagers are efficient while others are not.

Unlike most financial securities markets, the average lottery and sports betting participant must lose. We may, indeed, choose to differentiate gambling and investing by

CHANCE OF WINNING			
		COMPLETE LUCK	SKILL INVOLVED
P A Y O F	COMPLETE LUCK	Scratch lottery games with fixed payment	Example: Pay \$1 for a chance to pick all winners of hockey games on a particular day. From those who have all correct selections, one name is randomly drawn and awarded \$100,000.
	SKILL INVOLVED	Lotto, such as 6/49, with some or all pari-mutuel payoffs	Sports pool games Horse racing Blackjack Sports betting

FIGURE 1 Taxonomy of games. *Source:* Adapted from Ziemba et al. (1986).

their expected returns, using the terms gambling when the expected profit is negative and investing when the expected profit, including all transactions costs and risk adjustments, is positive. Obviously, a willingness to assume risk in the face of negative expected returns is inconsistent with the traditional assumptions that (1) individuals maximize the expected utility of wealth, and (2) utility functions are concave, that is, risk aversion. Instead of the second assumption, Friedman and Savage (1948) assumed a utility function that is convex in a neighborhood of the individuals present wealth but concave over higher and lower wealths. Given their different payoff distributions, simultaneously purchasing lottery tickets and insurance can be consistent with this form of a utility function. Markowitz (1952) offered a functional form that eliminates some behavior admitted by Friedman and Savage's form that is not generally observed. He also pointed to the possibility of a utility function that recognizes the fun of gambling. Conlisk (1993) formalized this notion and found his model to be largely consistent with actual risk-taking behavior.

Other surveys on this topic include Clotfelter and Cook (1991) and three chapters in Part VIII of this volume. On lotteries, Thaler and Ziemba (1988) on horse racing and lotteries, Hausch and Ziemba (1992) and Vaughan Williams (2003, 2005) on sports betting and lotteries, and Hausch et al. (1994a, 2008) on horse racing. Lane and Ziemba (2004, 2008) study hedging strategies for jai alai, a sport not considered in this survey.

2. EXTENT OF GAMBLING IN THE U.S.

The level of gambling in the U.S. is enormous. Welles (1989) estimates \$240 billion as the total annual wagering by Americans, both legal and illegal; an amount that is growing at 10% per year. By 1992, the estimate is \$329.9 billion including \$29.9 billion of gross revenue for governments and gaming institutions, more than six times annual movie ticket sales in the U.S. (Kleinfeld, 1993). Table 1 presents gross gambling revenues from various gambling from 1985 to 1992. Pari-mutuel betting revenues have grown slowest, while revenues from Indian gaming have dramatically risen.

Lotteries accounted for about one-third of gross gambling revenues over this period; see Table 1. On average, the lottery returns are only about 54% of sales. For pari-mutuel wagering, the average payback is about 80%, with the state government collecting usually only a few percent. That tax rate can vary across the tracks within a state, though. For example, the three major thoroughbred tracks in south Florida have long feuded over the prime winter racing season when tourism is at a peak. Florida passed a bill that in 1994 granted the prime season to Gulfstream Park with a 3.0% state tax. Hialeah Park's season was taxed by the state at 1.15%, while Calder's tax was 2.4%. Thalheimer and Ali (1995, 2008) found that demand for racetrack betting is price elastic, which suggests that track-take revenue should increase with a reduction in the track-take fraction. By 1990, 34 states were offering pari-mutuel betting on thoroughbred or standardbred horses, or on greyhounds. Thalheimer and Ali (1995) also found that the presence of a state lottery lowered both attendance and the average bettor's wager. See Table 2.

TABLE 1 Gross Gambling Revenues from Various Gambling Activities (\$billions)

Year	Lottery	Casino ^a	Pari-mutuel ^b	Bingo	Indian gaming	Total ^c
1985	5.2	5.5	3.1	0.91	0.09	15.4
1986	6.3	5.7	3.2	0.94	0.10	16.9
1987	6.6	6.4	3.3	0.90	0.11	18.3
1988	8.4	7.1	3.5	0.88	0.10	21.5
1989	9.6	7.7	3.6	0.95	0.30	24.0
1990	10.3	8.7	3.7	1.0	0.48	26.2
1991	10.2	9.0	3.7	1.1	0.72	26.7
1992	11.5	10.1	3.7	1.1	1.5	29.9

^aDoes not include Indian gaming.

^bIncludes horse and greyhound racing, jai alai, and off-track betting.

^cBecause of other forms of gambling, columns sum to less than the total column.

Source: Kleinfeld (1993).

TABLE 2 State Lottery Startup Dates, FY 2002 Sales, Sales Per Capita FY 2004, and Rank

State	Startup	Sales (\$ million)	Per capita sales	Rank
Arizona	1981	294.82	64.75	38
California	1985	2,915.90	82.57	29
Colorado	1983	407.97	87.71	28
Connecticut	1972	907.9	259.68	9
Delaware ^b	1975	674.01	777.57	3
Florida	1988	2,330.36	178.56	18
Georgia	1993	2,449.36	309.66	7
Idaho	1989	92.67	79.21	31
Illinois	1974	1,590.15	134.78	24
Indiana	1989	626.31	118.17	26
Iowa	1985	181.22	71.20	37
Kansas	1987	190.08	82.12	30
Kentucky	1989	638.72	175.52	19
Louisiana	1991	311.62	75.50	35
Maine	1974	157.9	141.54	22
Maryland	1973	1,306.55	252.10	11
Massachusetts	1972	4,213.22	682.60	5
Michigan	1972	1,688.04	194.88	14
Minnesota	1990	377.36	76.12	34

TABLE 2 (continued)

Missouri	1986	585.19	137.97	23
Montana	1987	33.63	39.83	40
Nebraska	1993	73.91	53.17	39
New Hampshire	1964	212.9	183.23	17
New Jersey	1970	2,068.52	252.40	10
New Mexico	1996	133.97	78.64	32
New York	1967	4,753.62	304.24	8
North Dakota ^e			9.21	41
Ohio	1974	1,983.11	188.21	16
Oregon ^b	1985	816.94	249.55	12
Pennsylvania	1972	1,934.16	189.86	15
Rhode Island ^c	1974	1,171.10	373.04	1
South Carolina	2002	319.99	227.63	13
South Dakota ^c	1987	629.96	865.25	2
Tennessee ^d			72.82	36
Texas	1992	2,966.27	130.28	25
Vermont	1978	81.99	148.83	21
Virginia	1988	1,108.07	170.30	20
Washington	1982	438.61	78.06	33
Washington, DC	1982	211.13	441.06	6
West Virginia ^b	1986	848.63	718.81	4
Wisconsin	1988	427.57	87.94	27
U.S. ^a		42,153.43	184.25	—

^aExcludes states with no lottery: Alabama, Alaska, Arkansas, Hawaii, Mississippi, Nevada, North Carolina, Oklahoma, Utah, and Wyoming.

^bIncludes net VLT sales (cash in less cash out).

^cIncludes gross VLT sales (cash in).

^dSales began January 20, 2004.

^eSales began March 25, 2004.

Source: Garrett and Wagner (2003) and Tax Foundation (2008). North American Association of State and Provincial Lotteries; Census Bureau population data; Tax Foundation calculations.

Sports betting in Nevada was estimated at \$1.3 billion in 1988, and illegal action nationwide at more than \$26 billion (see Akst, 1989), with the largest event being the Super Bowl for the championship of American football. Super Bowl wagering in 2008 was estimated to be over \$50 billion. A study in Ontario (see Abbate, 1995) found that 69% of adults played the lottery in the last month. For other forms of gambling, the corresponding percentages were 12% for sports gambling, 9% for card gambling, 7% for bingo, 3% for casinos, and 2% for horse racing.

3. RACETRACK BETTING MARKETS

3.1. Introduction to Racetrack Betting

The racetrack is a market in miniature in which wagering, odds, the outcome of the race, and payoffs all occur over a period of 20–30 min, followed by a new market, the next race. The number of horses in a race usually ranges from six to 12 in the U.S., three to 14 in Hong Kong, six to 18 in Japan, and four to 20 in England. A variety of wagers is available to bettors. The simplest is a wager to win, which involves picking the horse that finishes first. A place (show) bet involves picking a horse that is at least second (third). Exotic wagers are based on two or more horses, such as the daily double—picking the winners of two consecutive races; quinellas—picking the top two finishers in a race; and exactas—picking the top two finishers in a race in the correct order. As in securities markets, the public directly establishes the odds in a pari-mutuel betting market; the more the public bets on a horse, the lower its odds and the lower its return upon winning. The track pools the public's wagers and returns a fraction Q —the track payback—of that total to the winners. The remainder—the track take—is shared by the track owners, the state government, jockey's fund, race purses, and so on. An additional transactions cost is breakage, where the track rounds down returns to the nearest nickel or dime on the dollar.

Weak-form efficiency means that, with access only to the public's odds, no bet allows a positive expected profit. A profitable betting system based on the public's odds is possible only if weak-form efficiency is violated (see Epstein, 1977). We also consider a stricter version—strictly-weak-form efficiency—which requires that with access to the public's odds, all bets have equal expected return, namely Q , for a loss of $1 - Q$.

There is an extensive literature on horse race betting. We begin in Section 3.2 with studies of win betting; they point to a weak-form efficient win market but one that is not strictly-weak-form efficient given a strong and stable betting bias by the public against favorites and for longshots. Section 3.3 discusses the evidence against weak-form efficiency of the place and show markets. It also describes theoretical and implementation aspects of wagering schemes that have been devised to exploit this inefficiency. Exotic markets and cross-track betting are discussed in Sections 3.4 and 3.5, respectively. This survey of horse racing is restricted to pari-mutuel betting, that is, where the payoffs are directly determined by the public's betting. In many countries outside of North America, betting is handled by bookies who offer fixed odds. This literature is discussed in Hausch et al. (1994a, 2008).

3.2. Win Market

Define W_i as the public's wager to win on horse i and $W = \sum_i W_i$ as the public's win pool. Then, ignoring breakage, QW/W_i is the payoff per dollar wagered on horse i to win if and only if horse i wins. Let $O_i = QW/W_i - 1$. The odds on horse i are expressed as O_i to one, or $O_i/1$, and the return on a \$1 wager to win on horse i is the original \$1 plus another O_i in profit. Similarly, odds of x to y (or x/y) means that a \$ y wager

will pay $\$x + y$, the original wager plus $\$x$. The quantity W_i/W can be interpreted as the public's subjective probability that horse i will win. If these subjective probabilities are indeed correct, then the gross expected return on a win bet on any horse i is Q , that is, strictly-weak-form efficiency. Based on data on over 50,000 races and 300,000 horses collected from numerous studies, Figure 2 shows the actual expected returns for various odds categories. (This extends the analysis of Snyder, 1978.) Expected returns are plotted against odds using transaction costs of 0.1533 ($=1 - Q$) which applies in California. The horizontal line indicates the point at which returns are the expected 0.8467 ($=Q$). While objective odds (actual outcomes) and the public's subjective odds are highly correlated, actual returns departing from the horizontal line indicates that strictly-weak-form efficiency is violated.

Figure 2 exhibits a clear favorite-longshot bias, with expected returns falling as odds lengthen, down to only 13.7¢ per dollar wagered at odds exceeding 100/1. This bias is strong and stable, having appeared in datasets collected over several decades and from tracks of all sizes and throughout the world. (Early studies include Griffith, 1949; McGlothlin, 1956; and Fabricand, 1965.) In England, a similar graph (see Ziemba and Hausch, 1986) pertains even for fixed odds systems where bookies construct this risk-preference situation to mirror bettor's desires to overbet longshots and underbet favorites. Exceptions to this bias are Busche and Hall (1988) and Busche (1994) for Hong Kong and Japan. While clearly violating strictly-weak-form efficiency, Figure 2 illustrates that the favorite-longshot bias is insufficient to allow for profitable

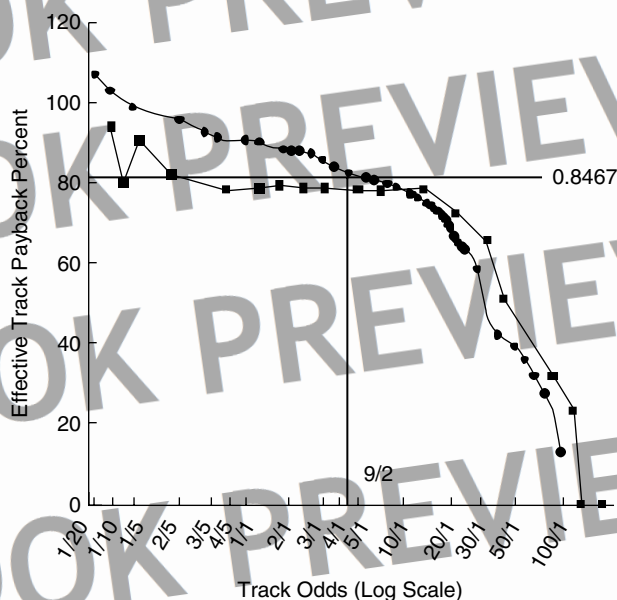


FIGURE 2 Effective track payback less breakage for various odds levels in California. Filled squares show more recent data. Sources: • Ziemba and Hausch (1986) and ■ Ziemba (2004).

wagers except on odds below 3/10. In this range, expected returns are positive and are about 4–5% for horses with the shortest odds. Such short odds are relatively rare; hence, for practical purposes, the win market, while not strictly-weak-form efficient, is weak-form efficient. This supports the notion that bettors, at least in aggregate, behave rationally in the sense that few profitable wagers remain. The filled squares in Figure 2 show more recent data (1997–2004) and in this period extreme favorites did not have positive expected value. Also, favorites have lower expected value than in the past. However, longshots are similarly overbet as in the data up to 1986. Rosett (1965) finds additional support for bettor rationality with evidence that a bettor will not make a simple bet over a more complicated one (a parlay) that is preferred in terms of overall probability of payoff.

McGlothlin (1956) studied the favorite-longshot bias over the course of a racing day, finding interesting departures from the usual bias for the last two races of the day. The penultimate race is frequently the feature race of the day, involving horses with well-documented records that have received more public scrutiny than usually occurs, and this is particularly so for the favorites. Interestingly, McGlothlin found little underbetting of these favorites. For the final race, McGlothlin calculated significantly positive expected returns for short-odds horses, atypically low returns for mid-range odds horses, and small but positive expected returns for the longest-odds horses. An explanation offered for these returns is that bettors who are losers—which describes the average bettor by the beginning of the last race of the day—use an end-game strategy to recover their day's losses. Such a strategy will tend to overlook favorites because of their low payoff, key in on mid-range odds horses because of their adequate returns together with a reasonable probability of success, and perhaps even involve underbetting of extreme longshots given their low probability of success. Metzger (1985) also found a loose pattern of increased underbetting on the favorite as the racing day progressed.

The favorite-longshot bias may be insufficient to allow for a practical and profitable technical betting scheme, but its existence is still worthy of explanation. Two approaches based on rational betting behavior have considered risk-seeking bettors and differences of opinion. The first of these generates the bias because risk-seeking bettors will demand a higher expected return for favorites, which have a lower variance of return than do longshots. Weitzman (1965) and Ali (1977) estimated the utility function of the representative bettor and showed it to be convex. Quandt (1986) proved that locally risk-seeking bettors are a necessary condition for the bias if the bettors have homogeneous beliefs, since a loss to the bettors in aggregate follows from a positive track take. Ali (1977) offered the second approach for generating the favorite-longshot bias. He considered races with two horses and assumed that the risk-neutral bettors hold heterogeneous beliefs about the likelihood of each horse winning, and all bettors wager an identical amount (say \$1) on the horse with the highest expected return on the basis of their belief. Ali defined π_1 as the true probability that horse 1 wins and supposed that each bettor's belief about that probability is a draw from a distribution that has π_1 as its median value. Suppose that $\pi_1 > 1/2$, that is, horse 1 truly is the favorite. Since each bettor wagers \$1, a pari-mutuel market belief of π_1 would require that a fraction π_1 of the bettors have beliefs exceeding π_1 . This would be unexpected, though, since π_1 is the

median belief. The pari-mutuel market payoff must be lower on average to attract the additional bettors needed to sustain that payoff. The pari-mutuel result is that horse 1 will tend to be the betting favorite but it will be underbet relative to its true probability. The reverse occurs with the other horse, the longshot, resulting in the favorite-longshot bias.² Blough's (1994) model incorporated both non-linear utility and heterogeneous beliefs, and with a restriction on the beliefs of bettors he extended Ali's analysis beyond two horses to an arbitrary number. He also developed an econometric test to distinguish between these two causes of the favorite-longshot bias but, as his data did not exhibit the bias, neither cause appeared to be present.

Behavioral explanations have also been offered for the favorite-longshot bias. Thaler and Ziemba (1988) mentioned several, including: (1) overestimation of small probability events; (2) exaggerating a longshot's contribution to expected utility (from prospect theory; see Kahneman and Tversky, 1979); and (3) bragging rights that are available to those who win with longshots but not to those winning with favorites. They also suggest that mental accounting (see Kahneman and Tversky, 1984) may be a promising way of modeling racetrack betting. It allows the bettor to be risk seeking in one domain while risk averse in another. The notion is that bettors have mental accounts and act as if the funds in these accounts are not fungible. Suppose bettors A and B have the same betting behavior. If A just lost his or her \$100 on a horse race while B just read in the newspaper's financial section of a \$100 loss in his or her stock portfolio, then mental accounting predicts different subsequent wagering behavior because B's loss is unrelated to racetrack betting.

Figure 2 is based on final odds—the odds on which win payoffs are determined. The public's odds can vary considerably over the betting period, though. Such price movements are of interest if they themselves reveal information. Asch et al. (1982) concluded that bettors late in the betting period do achieve higher returns than early bettors but were unable to devise a profitable strategy exploiting this. Asch et al. (1984, 1986) and Asch and Quandt (1986), using new data, showed that wagering on the horse with the highest win probability, when that probability is partially based on the public's marginal odds (i.e., odds calculated using only wagering late in the betting period), did not provide statistically significant profits.

Assuming that a bettor knows perfectly the true win odds of the horses in a race, Isaacs (1953) determined the betting scheme that maximizes expected profit. Isaacs accounted for the effect of one's bets on the odds and determined an algorithmic solution to the non-linear optimization problem.³ Rosner (1975) extended Isaacs by introducing logarithmic utility and a budget constraint, but ignored one's effect on the odds. Levin (1994) also extended Isaacs in several ways, for example, by introducing non-linear utility and a budget constraint. The public's win odds are just one source of information at the track. Others include the racing form, which provides past performances on all the horses entered in a race, predictions of expert handicappers that appear in

²This clearly is not a rational expectations equilibrium since bettors are not revising their estimates on the basis of the price that is offered.

³Below, we develop a similar optimization problem for place and show betting.

local newspapers, and tip sheets that can be purchased at the track. Semi-strong form efficiency can be studied by considering this information. Using a multinomial logit model to measure the information content of the forecasts of professional handicappers, Figlewski (1979) found their forecasts to contain considerable information but the public's odds discount almost all of it, in support of semi-strong efficiency. Snyder (1978) and a reanalysis of Losey and Talbott (1980) also showed that following the advice of professional handicappers is generally unprofitable. For harness racing, Ludlow (1994) tested a fundamental system. He considered several classification schemes for separating winners and nonwinners in one dataset, and then cross-validated them on another dataset.

Bolton and Chapman (1986) developed another fundamental system using a multinomial logit model to estimate win probabilities with 10 factors such as weight, post position, past performances, and jockey. They were essentially unable to demonstrate any significantly positive profits, though, in support of semi-strong form efficiency. Chapman (1994) extended this work to a 20-variable model and increased the dataset to 2,000 races from 200 in Bolton and Chapman (1986). Despite a very simple scheme of betting \$1 on the horse with the highest expected return as long as that return was positive, some evidence of profits was shown. Chapman also showed that the public's win odds contained considerable information beyond that contained in his 20 factors, but he did not investigate how its inclusion in his model would improve profits.

Benter (1994) provided some details of a very elaborate implementation of a multinomial logit model that is reported to have been successfully employed in Hong Kong for a number of years. Benter described the importance of defining factors that extract as much information as possible, stating that a model involving only simplistic specifications of factors does not provide sufficiently accurate estimates of winning probabilities. An example of the complexity of the factors that Benter uses is seen in his discussion of a distance preference factor to indicate the horse's demonstrated preference for a race of the distance that it will run in the upcoming race. For predicting races of 1–1.25 mi, Bolton and Chapman (1986) deal with the distance preference factor through a variable that equals one if a horse had run three or four of its last four races at distance levels of less than 1 mi, and zero otherwise (p. 1047). Benter (1994), as a result of a large number of progressive refinements, uses a variable defined as follows:

[F]or each of a horse's past races, a predicted finishing position is calculated via multiple regression based on all factors except those relating to distance. This predicted finishing position in each race is then subtracted from the horse's actual finishing position. The resulting quantity can be considered to be the unexplained residual which may be due to some unknown distance preference that the horse may possess plus a certain amount of random error. To estimate the horse's preference or aversion to today's distance, the residual in each of its past races is used to estimate a linear relationship between performance and similarity to today's distance. Given the statistical uncertainty of estimating this relationship from the usually small sample of past

racers, the final magnitude of the estimate is standardized by dividing it by its standard error. The result is that horses with a clearly defined distance preference demonstrated over a large number of races will be awarded a relatively larger magnitude value than in cases where the evidence is less clear.

A horse's post position is further information that is available to the bettor. Horses with inside post positions typically have an advantage in gaining positions near the rail during the race, a benefit that increases with the number of turns in the race; in fact, horses lose about one length every turn for each horse between them and the rail. Canfield et al. (1987) demonstrated that a post position bias exists favoring the inside posts and showed that the market correctly adjusted odds to reflect it, both in the win market and other markets. An exception is when the situation is nonobvious and the bet is complex, such as exacta wagering on rainy days when the inside bias may not be present.⁴ Such situations may allow profitable betting strategies. Betton (1994) also demonstrated a post position bias that was only partially reflected in the markets odds. She did not consider whether the bias was sufficient for profitable betting strategies.

Numerous trade publications purport to offer profitable systems that violate semi-strong form efficiency. Among the more "scientific" are Mitchell (1986), Quinn (1986, 1987, 1992), and Ziemba and Hausch (1986, 1987). Tests of strong-form efficiency are more difficult to conduct as they involve data that is generally not publicly available. Self-proclaimed accounts of successful gambling exploits do exist, though (e.g., Beyer, 1978, 1993). Also, Schnytzer and Shilony (1995) show that observing inside information in the odds can be beneficial. (See also Shin, 1992, 1993.)

3.3. Place and Show Markets

In the place and show markets, bets are profitable if the horse is at least second and third, respectively. Define P_j as the amount bet by the public on horse j to place and $P = \sum P_j$ as the place pool. Similarly, S_k is the public's show bet on horse k and $S = \sum S_k$ is the show pool. Then the payoff per dollar bet on horse j to place is

$$1 + \frac{QP - P_i - P_j}{2P_j} \quad (1)$$

if horses i and j are the first two finishers (in either order). Otherwise the payoff is zero. The first term of Equation (1) is the return of the dollar that was wagered. The second term recognizes that the track returns only OP of the pool, and from that amount the original wagers on horses i and j are returned. Of the remainder, $PQ - P_i - P_j$, half goes to the bettors of j , who share it on a per-dollar-wagered basis. The payoff on j is independent of whether j finishes first or second, but does depend on which horse

⁴Because tracks are beveled, rain may collect in the inside post positions.

it finishes with. A bettor on j to place generally prefers that a longshot, not a favorite, finishes with horse j . The payoff per dollar bet on horse k to show is similar:

$$1 + \frac{QS - S_i - S_j - S_k}{3S_k} \quad (2)$$

if horses i , j , and k are the top three finishers.⁵

One approach to place and show betting is to study whether profitable rules for those markets can be devised based on information in the win market. For instance, given the win market's bias for favorites, perhaps place and show bets on favorites—which are more likely to pay off than win bets—allow positive profits. Asch et al. (1984, 1986) studied this possibility using a logit model to estimate win probability based on the win odds, the marginal win odds (odds based on late betting only), and the morning line. Betting to win, place, and show on the horse with the highest win probability (together with additional screens), returns far exceeded those of the average bettor but were not positive with statistical significance. Betting to place or show based just on information in the win market can only be improved with some attention to information in the place and show markets, since it is the public's wagering in those markets that alone determine place and show payoffs. This was the approach of Hausch, Ziemba, and Rubinstein (Hausch et al., 1981, hereafter HZR), who demonstrated that weak-form inefficiencies can be identified in the place and show markets. There are three distinct aspects to their model: (i) determining place and show probabilities which together with Equations (1) and (2) allow expected returns to be calculated; (ii) using the optimal capital growth model for wagering; and (iii) using approximations for implementation in real time. All three aspects will be described.

3.4. Place and Show Probabilities

Suppose that horse i 's true win probability is (which could be determined on the basis of fundamental handicapping or, as HZR assumed, be derived from the public's wagering as W_i/W). Then Harville's (1973) formulas estimate the probability that horse i is first and horse j is second as:

$$q_{ij} \equiv \frac{q_i q_j}{1 - q_i} \quad (3)$$

Equation (3) follows from assuming that the likelihood that j will be second, conditional on i being first, is the probability that j would be first if i were not in the race, which can

⁵Equations (1) and (2) ignore breakage, which is incorporated by rounding down a payoff to the nearest nickel or, more typically, dime on the dollar. A further consideration is that U.S. tracks generally guarantee winners a minimum profit of 5%. So, even if $(SQ - S_i - S_j - S_k)/3S_i < 0.05$, which is termed a minus pool, show bettors on i receive a 5% return. Minus pools are most common when there is an extreme favorite. In that event, Hausch and Ziemba (1990b) show how risk-free arbitrage may be possible in the place and show markets.

be estimated to be $q_j/(1 - q_i)$. Similarly, Harville estimated the probability that horses i , j , and k finish first, second, and third, respectively, to be:

$$q_{ijk} \equiv \frac{q_i q_j q_k}{(1 - q_i)(1 - q_i - q_j)}. \quad (4)$$

Harville's formulas have also been discussed by Savage (1957) and Plackett (1975). With these ordering probabilities, place and show probabilities can be calculated. The probability that i places is the probability that i finishes first or second, which is:

$$q_i + \sum_{j \neq i} q_{ji}, \quad (5)$$

and the probability that i shows is:

$$q_i + \sum_{j \neq i} q_{ij} + \sum_{k \neq i} \sum_{j \neq i} q_{kji} \quad (6)$$

Dansie (1983) and Henery (1981) demonstrated that Harville's formulas, Equations (3) and (4), are implied when horses running times are independently and exponentially distributed. If horse i 's running time, T_i , is exponentially distributed with mean $1/\lambda_i$, then Equations (3) and (4) follow with $q_i = \lambda_i / \sum_j \lambda_j$. An alternative derivation supposes that T_i has the extreme-value distribution with location parameter, that is, i 's running time has density function:

$$f_i(t_i) = \exp(t_i - \theta_i) \exp[-\exp(t_i - \theta_i)], \quad -\infty < t_i < \infty$$

Then, if T_1, \dots, T_n are independent, Equations (3) and (4) follow with $q_i = \exp(\theta_i) / \sum_j \exp(-\theta_j)$. These two derivations are related because the extreme-value distribution is the logarithm of an exponential distribution, and the logarithm preserves order.

Underlying probability distributions not generating Equations (3) and (4) have been considered, too. Henery (1981) considered running times, T_i , that are independent and normally distributed as Equation (1). Then

$$q_{ij} = P(T_i < T_j < \min_{k \neq i,j} [T_k])$$

$$= \int_{-\infty}^{\infty} \Phi(t + \theta_j - \theta_i) \prod_{k \neq i,j} [1 - \Phi(t + \theta_j - \theta_k)] \phi(t) dt,$$

where $\Phi(\cdot)$ and $\phi(\cdot)$ are the standard normal cumulative distribution function (cdf) and density function, respectively. A similar expression holds for q_{ijk} . Thus, with (θ_i) , the ordering probabilities involve numerical integration. With (q_i) , possibly on the basis

of the public's win bets or from handicapping, (θ_i) is the solution of the following non-linear system of equations:

$$q_i = P(T_i < T_j < \min_{j \neq i} [T_j]) \quad i = 1, \dots, n$$

$$= \int_{-\infty}^{\infty} \prod_{k \neq i, j} \Phi(t + \theta_i - \theta_j) \Phi(t) dt.$$

Stern (1990) proposed another alternative to the Harville formulas. He assumed that running times are independent and gamma distributed with fixed shape parameter r . Such an underlying distribution is descriptive of a game where players score points according to independent Poisson processes and the winner is the first to score r points. With $r = 1$, Stern's model is Harville's model, and as $r \rightarrow \infty$, it converges to Henery's model.

Several empirical studies have considered these various models. Henery (1984) fit running times to the extreme-value distribution, finding that it best fit the faster running times. Harville (1973) showed that his formulas tend to over estimate the probability that favorites finish second or third and underestimate these same probabilities for longshots. This reverse favorite-longshot bias for the likelihood that horses finish second and third was also observed by HZR. The betting data in these two papers exhibits the favorite-longshot bias. One might conjecture that a natural consequence of this bias for first position is the reverse favorite-longshot bias for other positions, like second and third, as probabilities over a horse's finish positions must sum to one. However, Benter (1994), using Hong Kong data that does not exhibit the favorite-longshot bias, showed this conjecture to be incomplete at best in Hong Kong. Tables 3–5 show his results.

TABLE 3 Public's Estimate of Expected Win Probability Versus Actual Win Frequency (Based on 3,198 Races, Royal Hong Kong Jockey Club, September 1986–June 1993)

Range of estimated probabilities	Horses within range	Mean expected win probability	Actual win frequency observed	Z-statistic
0000–0.010	1,343	0.007	0.007	0.0
0.010–0.025	4,356	0.017	0.020	1.3
0.025–0.050	6,193	0.037	0.042	2.1
0.050–0.100	8,720	0.073	0.069	–1.5
0.100–0.150	5,395	0.123	0.125	0.6
0.150–0.200	3,016	0.172	0.173	0.1
0.200–0.250	1,811	0.222	0.219	–0.3
0.250–0.300	1,015	0.273	0.253	–1.4
0.300–0.400	716	0.339	0.339	0.0
>0.400	312	0.467	0.484	0.6

Source: Benter (1994).

TABLE 4 Harville Calculation of Expected Probability of Finishing Second Versus Actual Frequency (Based on 3,198 Races, Royal Hong Kong Jockey Club, September 1986–June 1993)

Range of estimated probabilities	Horses within range	Mean Harville probability of second	Actual second finish frequency observed	Z-statistic
0.000–0.010	962	0.007	0.010	0.9
0.010–0.025	3,449	0.018	0.030	5.3
0.025–0.050	5,253	0.037	0.045	2.8
0.050–0.100	7,682	0.073	0.080	2.3
0.100–0.150	4,957	0.123	0.132	1.9
0.150–0.200	3,023	0.173	0.161	–1.8
0.200–0.250	1,834	0.223	0.195	–3.0
0.250–0.300	1,113	0.272	0.243	–2.3
0.300–0.400	1,011	0.338	0.317	–1.4
>0.400	395	0.476	0.372	–4.3

Source: Benter (1994).

TABLE 5 Harville Calculation of Expected Probability of Finishing Third Versus Actual Frequency (Based on 3,198 Races, Royal Hong Kong Jockey Club, September 1986–June 1993)

Range of estimated probabilities	Horses within range	Mean Harville probability of third	Actual third finish frequency observed	Z-statistic
0.000–0.010	660	0.007	0.009	0.5
0.010–0.025	2,680	0.018	0.033	4.3
0.025–0.050	4,347	0.037	0.062	6.8
0.050–0.100	6,646	0.073	0.087	4.0
0.100–0.150	4,325	0.123	0.136	2.5
0.150–0.200	2,923	0.173	0.178	0.7
0.200–0.250	1,831	0.223	0.192	–3.4
0.250–0.300	1,249	0.273	0.213	–4.9
0.300–0.400	1,219	0.341	0.273	–5.3
>0.400	601	0.492	0.333	–8.3

Source: Benter (1994).

Table 3 categorizes horses by the public's estimate of their win probability (through the win odds) and compares that estimate to the actual win frequency. There is no obvious bias. Tables 4 and 5 calculate Harville's estimate of finishing second and of finishing third, respectively, and compare those with actual frequencies. The reverse favorite-longshot bias is clearly exhibited, and is more pronounced for third position than for

second. Benter's explanation of this bias is that Harville's formula does not recognize the increasing randomness of the contests for second and third place.

Lo (1994) provided a theoretical basis for this bias, demonstrating that Harville's formulas produce such a bias if the underlying distribution of running times is indeed independent gamma. HZR described a special type of late-charging horse that tends to either win or finish far back in the field, making a second or third place finish unlikely. For such horses, this "Silky Sullivan" phenomenon (named after a horse that displayed this racing pattern) provides another explanation of the probability bias for second and third. Stern (1994) described this as a type of information contained in the horse's failure to win that j is not used to adjust the probabilities, and related it to the memoryless property of the exponential distribution for running times. Stern (1990) analyzed 47 races and found that ordering probabilities estimated using $r = 1$ (Harville's model) were less accurate than those estimated using $r = 2$. Using a likelihood approach and Japanese wagering data, Lo (1994) found $r = 4$ to be best for Stern's model. He reported $r = \infty$ (Henery's model) was best on his Meadowlands and Hong Kong data, though. Thus, while one running-time distribution model does not appear to hold universally, there is limited empirical support for Harville's model. (Other empirical studies include Bacon-Shone et al., 1992a,b; Lo, 1994; and Lo and Bacon-Shone, 2008.)

While Henery's (1981) and Stern's (1987) ordering probabilities are superior to Harville's, the complex numerical calculations that they both require essentially precludes them from being used on-track (unless one determines one's own odds in advance of the race). Henery suggested using a first order Taylor series approximation, but Bacon-Shone et al. (1992b) demonstrated its inaccuracy. Stern (1994) also mentioned that this simplification does not seem to improve on Harville's model.

Thus, Harville's model still is useful; in particular for place and show probability at tracks where the favorite-longshot bias is exhibited in the win market. The probability of place, say, involves adding the probability of first, for which there is a favorite-longshot bias, and the probability of second, which has a reverse bias when calculated using Harville's formula. These biases tend to cancel each other, as HZR demonstrated. Lo and Bacon-Shone's (2008) discount model retains the simplicity of the Harville model but attempts to directly correct its bias for ordering probabilities. They define $LO(i, j|k) \equiv$ logarithm of the odds that horse i beats j for the k th position given that neither i nor j finish in the top $k - 1$ positions. For Harville this is simply $\log(q_i/q_j)$, independent of k . Lo and Bacon-Shone assumed instead that

$$LO(i, j|k) = \lambda_k LO(i, j|k),$$

and further assumed that λ_k is decreasing in k , presuming that the relative abilities of the horses matters less as the finish position worsens and the prize money decreases. The parameters λ_2 and λ_3 can be estimated to best fit with Henery's or Stern's models, or estimated to best fit data from a particular track, from which simple ordering probabilities q_{ijk} can be calculated. For example, Benter's (1994) Hong Kong data generated values of $\lambda_2 = 0.81$ and $\lambda_3 = 0.65$, and Lo et al. (1995) used $\lambda_2 = 0.88$ and $\lambda_3 = 0.8$

for their Japanese data. They reported an improved fit between actual and expected frequencies.

For any of these ordering probability models, the expected return on a bet on horse i to place and show can, respectively, be calculated using Equations (1) and (2) as:

$$\text{EXP}_i = \sum_{j \neq i} (q_{ij} + q_{ji}) \left(1 + \frac{QP - P_i - P_j}{2P_i} \right), \quad (7)$$

and

$$\text{EXS}_i = \sum_{k \neq j} \sum_{j \neq i} (q_{ijk} + q_{jik} + q_{jki}) \left(1 + \frac{QS - S_i - S_j - S_k}{3S_i} \right). \quad (8)$$

Using Harville's model and Equations (3) and (4), HZR studied Equations (5) and (6) with data from the 1978 summer season at Exhibition Park (1,065 races and 9,037 horses over 110 days) and the 1973–1974 winter season at Santa Anita Racetrack (627 races and 5,895 horses over 75 days). Table 6 shows the results of wagering on bets that, through Equations (5) and (6), are identified as having certain minimum levels of positive expected return.⁶ The results strongly suggest that inefficiencies can be identified in the place and show markets. Furthermore, HZR found that such inefficiencies appear about 2–4 times per racing day. Additional calculations along these lines appear in Harville (1973).

3.5. Optimal Capital Growth

The previous section showed evidence of inefficiencies in the place and show markets. The second aspect of HZR's model is to determine the bet size upon identifying a profitable wager. HZR employ the optimal capital growth model, or the Kelly criterion, which maximizes the expected logarithm of wealth on a race-by-race basis. It was developed by Kelly (1956) for information transmission independently developed by Latané (1959), extended by Breiman (1961) who provided rigorous proofs of the main results, and is described in detail by Hakansson and Ziemba (1995).⁷ Among its properties are: (1) it maximizes the asymptotic growth rate of capital; (2) asymptotically, it minimizes the expected time to reach a specified wealth goal; and (3) it outperforms in the long run any other essentially different strategy almost surely. While these are impressive asymptotic properties, Aucamp (1993) asked: how many periods are required to be reasonably confident that the Kelly criterion will be superior? Both theoretical and experimental evidence point to the need for only a moderate number of plays when the risk is low but

⁶See HZR for Equations (5) and (6) adjusted for breakage. Table 6 correctly accounts for the adverse effects of breakage.

⁷See also Algeot and Cover (1988) and Thorp (2006) for a more general mathematical treatment, and MacLean et al. (1992) and MacLean and Ziemba (2006) for fractional Kelly strategies and a comparison of its properties.

TABLE 6 Results of Betting \$1 at Exhibition Park and Santa Anita to Place and Show on Horses with a Theoretical Expected Profit of at Least a Specified Minimum Level

Minimum level	Place			Show		
	Number of bets	Total net profit (\$)	Rate of return (%)	Number of bets	Total net profit (\$)	Rate of return (%)
Exhibition Park						
1.04	225	5.10	2.3	612	33.20	5.4
1.08	126	−10.10	−8.0	386	53.50	13.9
1.12	69	11.10	16.1	223	40.80	18.3
1.16	40	5.10	12.8	143	26.30	18.4
1.20	18	5.30	29.4	95	21.70	22.8
1.25	11	−2.70	−24.5	44	11.20	25.5
1.30	3	−3.00	−100.0	27	10.80	40.0
1.50	0	—	—	3	6.00	200.0
Santa Anita						
1.04	103	12.30	−11.9	307	−18.00	−5.9
1.08	52	12.80	24.6	162	6.90	4.3
1.12	22	9.20	41.8	89	3.00	3.4
1.16	7	2.30	32.9	46	12.40	27.0
1.20	3	−1.30	−43.3	27	6.20	23.0
1.25	0	—	—	9	6.00	66.7
1.30	0	—	—	5	5.10	102.0
1.50	0	—	—	0	—	—

Source: Hausch et al. (1981).

a large number generally when the risk is high. See also Ziemba and Hausch (1986) for empirical calculations.

Using Equations (5) and (6), assuming the appropriateness of Harville’s formulas, introducing bets to place and show as decision variables, and accounting for one’s effect on the odds (but ignoring breakage), the optimal capital growth model solves optimization problem 1 (OP1).⁸ Notice that OP1 considers each possible 1-2-3 finish of the horses, determines the logarithm of final wealth in that event, and averages over all possible 1-2-3 finishes. For convenience, OP1 defines $P_{ij} \equiv P_i + P_j$ and $S_{ijk} \equiv S_i + S_j + S_k$. OP1 also involves a budget constraint with w_0 representing current wealth.

⁸Kallberg and Ziemba (1994) described OP1’s generalized concavity properties.

Optimization problem 1 is shown as follows with place and show betting with the optimal capital growth model.

$$\begin{aligned} \max_{p_i, s_i} & \sum_{i=1}^n \sum_{j=1, j \neq i}^n \sum_{k=1, k \neq i, j}^n \frac{q_i q_j q_k}{(1 - q_i)(1 - q_j - q_k)} \log \left[\begin{aligned} & \frac{Q(P + \sum_{l=1}^n p_l) - (p_i + p_j + p_{ij})}{2} \\ & \times \left[\frac{p_i}{p_i + p_j} + \frac{p_j}{p_j + p_i} \right] \\ & + \frac{Q(S + \sum_{l=1}^n s_l) - (s_i + s_j + s_k + s_{ijk})}{3} \\ & \times \left[\frac{s_i}{s_i + s_j} + \frac{s_j}{s_j + s_i} + \frac{s_k}{s_k + s_i} \right] \\ & + w_0 - \sum_{\substack{l=1 \\ l \neq i, j, k}}^n s_l + \sum_{\substack{l=1 \\ l \neq i, j}}^n p_l \end{aligned} \right] \\ \text{s.t.} & \sum_{l=1}^n (p_l + s_l) \leq w_0, \quad p_l \geq 0, \quad s_l \geq 0, \quad l = 1, \dots, n. \end{aligned}$$

3.6. Implementing the System and Empirical Results

HZR and Hausch and Ziemba (1985) provided tests of OP1 on data from three North American racetracks. Figures 3, 4, and 5 show, respectively, the wealth level histories for the 1978 season at Exhibition Park (Vancouver, BC), the 1973–1974 winter season at Santa Anita (Arcadia, CA), and the 1981–1982 winter season at Aqueduct (Jamaica, NY). Given the many approximations involved in the system, the authors established cutoffs on the expected return of a bet (see Equations (5) and (6)) that were necessary for wagering, cutoffs that decrease with the size of the track, and one's confidence with the accuracy of the public's win odds. Specifically, Figure 3 for Exhibition Park, a relatively small track, uses an expected return cutoff of 1.20. Figure 4 for Santa Anita, a much larger track, uses a 1.16 cutoff, and Figure 5 for Aqueduct uses 1.14. The track take is generally established by the state or province, and varied across these three tracks. For the seasons studied, the track takes were 18.9% for Exhibition Park, 17.5% for Santa Anita, and 15% for Aqueduct. To appreciate the dramatic long run effect of the take, Figure 5 also considers takes of 14% and 17%.

Figures 3–5 are based on OP1's use of final public wagers, that is, the presumption was that our bettor could wager last. In practice, that is not possible because of three time-consuming activities necessary for implementing OP1 in real time. First is the input of the required data—the public's win, place, and show bets on all horses in a race. This data is on the order of 3 four- or five-digit numbers for, perhaps, 10 horses. Second is the solution of the non-linear optimization problem OP1. And third is the time necessary to make one's bets before the end of the betting period. Since the public's wagering can and commonly does change over the betting period, these three activities mean that one must work with data that only imperfectly forecast the eventual payoffs. To reduce the time involved in the first two activities, HZR developed regression approximations to the solution of OP1 that, once a horse was identified as being

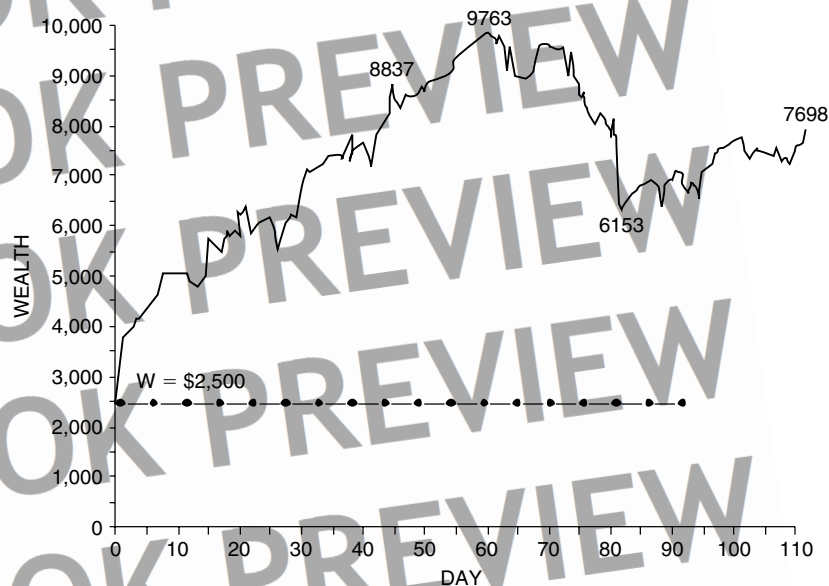


FIGURE 3 Wealth level history, Exhibition Park, 1978 season. Results from expected log betting to place and show when expected returns 1.20. Initial wealth is \$2,500, track payback is 81.9%, and breakage is accounted for. *Source:* Hausch et al. (1981).

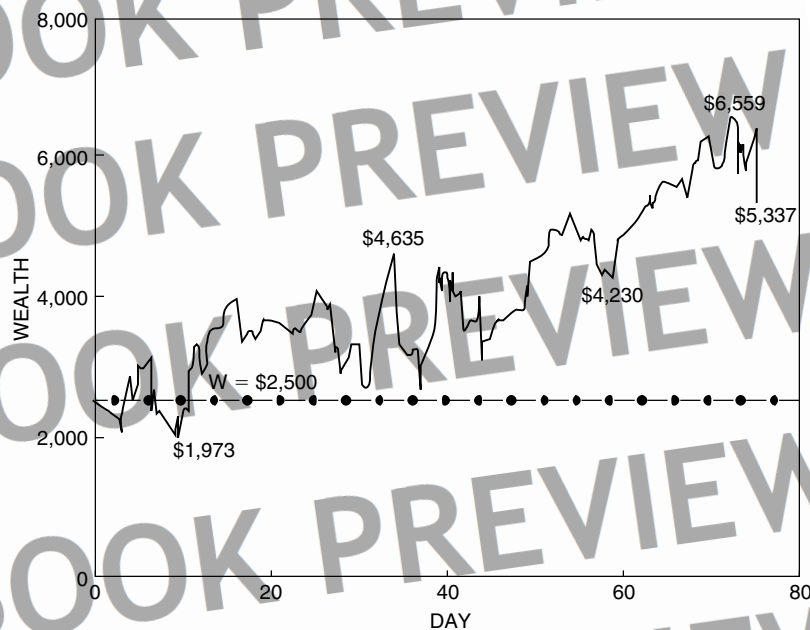


FIGURE 4 Wealth level history, Santa Anita, 1973–1974 season. Results from expected log betting to place and show when expected returns 1.16. Initial wealth is \$2,500, track payback is 82.5%, and breakage is accounted for. *Source:* Hausch et al. (1981).

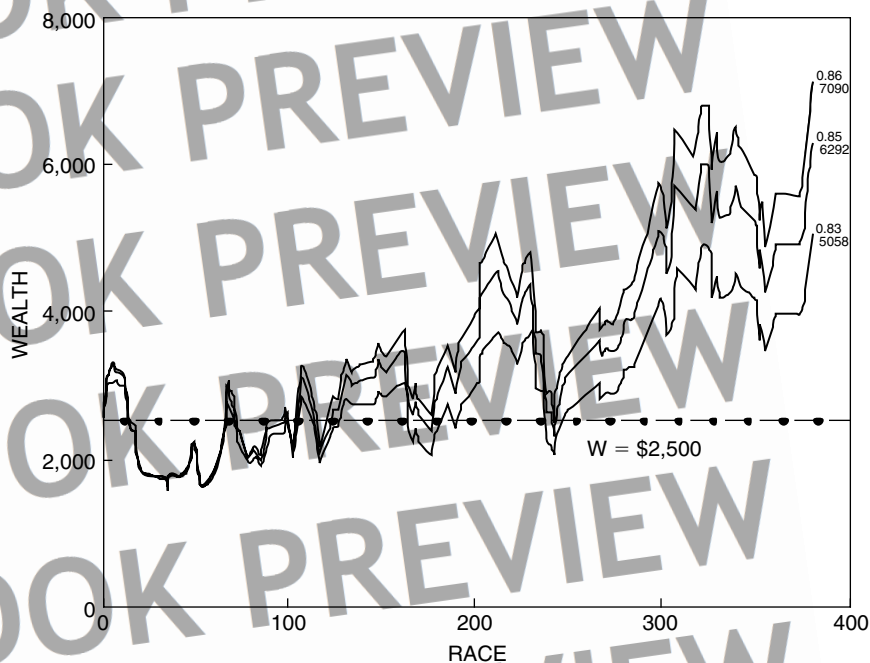


FIGURE 5 Wealth level history. Aqueduct, 1981–1982 season. Results from expected log betting to place and show when expected returns 1.14. Initial wealth is \$2,500, breakage is accounted for and track paybacks considered are 83%, 85% (actual), and 86%. *Source:* Hausch and Ziemba (1985).

a possible bet, required as inputs only the public's win and place (or show) wagers on that horse as well as the win and place (or show) pools. Analyzing a small sample of races with data 2 min prior to the end of betting, HZR found that the problem of using such data is limited—bets identified as profitable based on odds 2 min before the end of the betting period generally remained profitable based on the eventual final odds. Hausch and Ziemba (1985) and Ziemba and Hausch (1987) improved upon HZR's regression approximations and introduced additional approximations for multiple horse entries and multiple bets. Programmed into a speciality calculator, entry of the required data and calculations takes about 15 sec.

Hausch et al. (1995) tested these approximations on 62 runnings of the Kentucky Derby (1934–1995). The anomaly was present during this period and \$2,500 grew to \$8,002. With a breeding filter (based on dosage theory), wealth increased to \$12,508. See Bain et al. (2006) and Gramm and Ziemba (2008) in this volume for updates of this and applications to the Preakness and Belmont Stakes.

Ritter (1994), in a revision of work predating HZR, also considered place and show betting. Instead of computing expected returns, Ritter used various filter rules to select place and show bets. He achieved positive profits using final odds, but not when wagers were based on the odds 1.5 min from the end of betting.

Lo et al. (1995) demonstrated that OP1s performance can be improved by replacing Harville's model with Lo and Bacon-Shone's (2008) discount model that corrects for biases in Harville's model.

3.7. Does the System Still Provide Profits?

We have a final odds place and show wager simulation; see Ziemba (2004). The effect of bets on payoffs was not considered. For 4,879 races from major U.S. tracks, there were 752 wagers made including full, half, and third Kelly wagers. The initial bankroll was \$10,000 and yielded final bankrolls of \$55,757 with full, \$32,878 with half, and \$24,100 with third Kelly. The assumptions were

- No rebates.
- Expected value cutoff = 1.14 for large tracks and 1.18 for small tracks.
- For all tracks evaluated, no bets were made when the odds were greater than 8.00/1.
- All surfaces and track conditions were included.
- All classes of races were included.
- Coupled entries were used and optimized.
- Place and show bets on the same horse were optimized.
- Bets on multiple horses in the same race were not optimized.

Figure 6 shows the wealth path. The odds of the favorite, and often our place and/or show bet, tended to fall as the betting period ended. But more and more of the betting

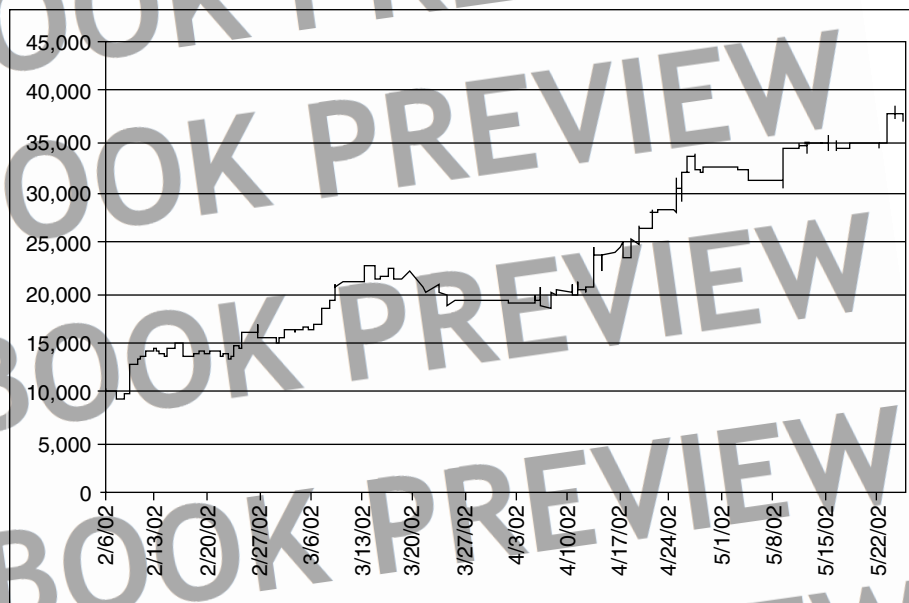


FIGURE 6 Wealth path of place and show bets on 752 wagers from 4,879 races in 2003.

actually is not posted until after the race has begun. By 2008 this was about 50% largely because about 87% of the total bets for a typical U.S. racetrack are made off track. This money is bet before the race begins, but the time lags in its reporting in the pools from the other track betting and from professional betting near the start of the race causes this effect and significant odds changes from the off-track prices. A typical application in 2006 with rebates provided a loss of about 7% over 80 U.S. racetracks with an average of about 9% rebate for a profit of about 2%. A test account with \$5,000 grew to \$30,000 with total betting of about \$1,500,000 over eight months.

3.8. Exotic Markets

For at least two reasons, exotic bets are typically more complicated to assess than bets to win, place or show: (1) they involve the outcome of two or more horses; and (2) one often has poor access to information about the public's exotic wagering (in fact, some tracks display no exotic wagering data). The reason for not fully displaying exotic data is entirely due to the large quantity of it. For example, in a 10 horse race, the public's win, place, and show bets involve 30 numbers. By comparison, there are 45 quinella numbers, 90 exacta numbers, and 720 trifecta numbers. Exotic wagers tend to be popular with the public. One attraction is their low probability and high payoff characteristics. In view of the favorite-longshot bias for win bets, it is not surprising that longer-odds events than win bets will be heavily wagered. To illustrate a second attraction, consider the daily double. Bettors can create for themselves a daily double by wagering to win on a horse and then, if successful, betting all the proceeds to win on a horse in the next race. This self-constructed daily double, called a parlay, differs from a daily double in an important way. The parlay is subjected to the track take twice while the daily double pays it only once. Thus, transaction costs are higher with the sequence of win bets. A successful handicapper needs skills significantly better than those of the average bettor. As transaction costs decrease, though, success demands less of a skill advantage or, for the same skill advantage, profits are greater.⁹ Most tracks appreciate this and charge a higher track take on exotic wagers, reducing their advantage.

While a parlay that is a self-constructed version of a daily double pays the track take twice, it does allow the bettor more information. Since the daily double must be wagered before either race, the bettor has little information about the public's view of the second race (other than any information that can be gleaned from the payoffs that are offered on daily double combinations, but that information is usually difficult to access). The parlay allows one to wager on the second race with a better sense of the public's impressions of the horses. Ali (1979) found that returns of parlays and double bets were not significantly different. Asch and Quandt (1987), however, found that doubles are statistically more profitable than parlays. When parlay payoffs are adjusted as if parlay bettors paid the track take just once, then returns on parlays and doubles are not significantly different. Lo and Busche's (1994) conclusions were similar for Hong Kong data.

⁹For more details on this effect and a numerical example, see Benter (1994).

Asch and Quandt (1987) found some support for the notion that the smart money is in the exotic pools. The basis for this notion is that the informational content of the smart money is more difficult for the public to discern in the exotic market than it would be if it were wagered in the win market. Their analysis ignores the systematic biases of the Harville (1973) model, though. Also, Dolbear (1993) described a further bias in their comparison of the theoretical and subjective probabilities of exacta outcomes. Bacon-Shone et al. (1992b) addressed these concerns and concluded that the public's exacta betting provides more accurate estimates of ordering probabilities (the probability that *i* wins and finishes second) than does the win market, and this is so using any of the probability models offered by Harville (1973), Henery (1981), or Stern (1990). Similarly, the trifecta market provides more accurate estimates of their ordering probabilities than does the win market.

Asch and Quandt (1988) analyzed the unbiasedness of the probabilities implied by exacta and trifecta bet fractions. Using simple linear regressions on the relationship between objective probabilities (estimated by win frequencies) and subjective probabilities (average bet fractions), they drew different conclusions for the two pools. Exacta pools implied probabilities that appeared to be unbiased while the trifecta bet fractions more weakly approximated the objective probabilities and some clear over/underbetting bias was exhibited.

Ziembra and Hausch (1986) adapted for exotic wagering the techniques developed by HZR for place and show wagering: win probabilities from the efficient win market, Harville's (1973) model to price other wagers and the Kelly criterion to determine wagers. Hausch et al. (1994b, 2008) developed general formulas for optimal wagering on exotic bets, allowing ordering probabilities based on Harville (1973), Henery (1981), or Stern (1990) with the help of approximations developed by Lo and Bacon-Shone (1993). Quinella data on 369 Hong Kong races was used to illustrate the system.

Kanto and Rosenqvist (1994) developed a betting system for quinella bets (called double bets in Finland) at a Finnish racetrack. Instead of using the win odds data directly, they used maximum likelihood estimation and Harville's (1973) model to estimate the win probabilities and the probabilities associated with a quinella bet by assuming that the quinella bet amounts for different combinations follow a multinomial distribution. Using the Kelly criterion for wagering and 111 races, they showed some evidence of positive profits.

3.9. Cross-Track Betting

Cross-track betting allows bettors to wager at their track (a cross track) on a race being run at another track (the home track). Since each track operates a separate pool, the payoffs at the tracks can differ. And, in fact, they do often differ, sometimes quite dramatically. For instance, a \$2.00 win ticket on Ferdinand, the winner of the 1986 Kentucky Derby, paid \$16.80 at Hollywood Park in California, \$79.90 at Woodbine in Toronto, and \$90.00 at Evangeline in Louisiana. Using data from Triple Crown races, Hausch and Ziembra (1990a) developed and tested optimal betting strategies for cross-track betting to exploit different odds across tracks. One strategy identified whether

a risk-free hedge could be developed by betting each horse at the track where its offered odds were longest, in relative amounts that guarantee a profit. Examples where the variance in odds across tracks was sufficient were provided. Also analyzed was the optimal capital growth strategy. This latter strategy was studied in two environments: (1) a single bettor at a cross track observing (perhaps by television) the home tracks odds; and (2) a syndicate of bettors, one at each track, communicating with each other.

Leong and Lim (1994) also found evidence of profits using cross-track betting that exists between races in Singapore and Malaysia. Both papers showed profits but neither had sufficient data for statistically significant profits. Thus, further work is needed to understand cross-track betting, particularly as it is becoming more popular. Tracks have been facing declining attendance, due in part to the increase in other forms of gambling, for example, lotteries. Cross-track betting helps attendance by offering bettors the opportunity to wager on prestigious horses in races at large tracks. Since the cross track does not need to stage an expensive race and the home track receives a portion of the cross track's revenues, both tracks can benefit from this arrangement.

4. THE FOOTBALL BETTING MARKET

Bettors on National Football League (NFL) games are offered a point spread. For example, suppose team A is a 10 point favorite over team B. Then a bet on A pays only if A wins by at least 11 points while a wager on B pays only if B either wins or loses by fewer than 10 points. If A wins by exactly 10 points, then wagers are usually refunded. Typically bettors pay \$11 for a \$10 profit when they win. This provides the bookmaker a commission and means that a bettor has to beat the spread 52.4% of the time to break even. When the actual point spread equals the offered point spread, the bookmaker receives no return. Otherwise, by perfectly balancing the wagers, a bookmaker can guarantee a profit of 4.55% (since \$21 is paid for each \$22 wagered). The Las Vegas sports books, which dominate the market, offer opening point spreads on the coming week's game. These spreads may change over the week, but bettors receive the spread offered at the time they placed their bet.¹⁰

The efficiency of NFL betting rests on the accuracy of the point spreads. An obvious and common approach to study their accuracy is to regress the actual point spread on the offered point spread. If, say, bettors tend to wager on underdogs then, to balance the books, the bookmaker has to offer point spreads lower than unbiased expectations about actual point spreads. Bettor biases should be reflected in the point spread offered and, if they are sufficiently large, should allow profitable betting opportunities, which would reject efficiency.

¹⁰These dynamics are also present in horse race wagering against bookies. Wagering on jai alai is similar (see Lane and Ziemba, 2004, 2008), too, but its odds change during the contest as points are scored rather than before the contest as with sports betting. Pari-mutuel betting is different, though; its odds change over the course of the betting period as betting patterns change, but payoffs to all bettors are based only on the final odds.

Let A_i be the actual point spread in game i and let P_i be the point spread offered. The following equation can be estimated:

$$A_i = \beta_1 + \beta_2 P_i + \epsilon_i, \quad (9)$$

where ϵ_i is the error term. The efficiency test is the joint hypothesis that $\beta_1 = 0$ and $\beta_2 = 1$. Pankoff (1968), Zuber et al. (1985), and Sauer et al. (1988) all found significant support for the hypothesis. Gandar et al. (1988) results are similar for both opening and closing point spreads (the point spread can change over the betting period as bookmakers attempt to balance their books). Their large t -statistic on β_2 and low R^2 (3.4% for closing data) suggest that while the point spread for any particular game is a poor predictor of the actual point spread, it is a good predictor of the average actual point spread for a group of games with this point spread.

While these results support market efficiency, they are not directly useful in answering whether there might be technical rules that are economically profitable. Vergin and Scriabin (1978) used NFL data from 1969 to 1972 to consider various rules, such as betting on the underdog when the point spread exceeds some specified level, and identified several profitable strategies. Using 1975–1981 data, Tryfos et al. (1984) demonstrated that most of these strategies were unprofitable or, if profitable, not at the 5% significance level. Those that were significantly profitable all required a syndicate taking advantage of different point spreads in different cities. Gandar et al. (1988) found similar negative results for these strategies using their 1980–1985 data.

Golec and Tamarkin (1991) discussed how a model such as Equation (7) can mask specific biases: consider that β_1 measures the average of the biases that do not change with the magnitude of the point spread. If half the observations in a data sample include a positive bias and the other half a negative bias of equal magnitude, then $\beta_1 = 0$ (p. 314). The problem is that Equation (7) can deal with only one bias. For instance, a bias in favor (or against) the home team can be considered by defining the data, P_i , relative to the home team. But if there is also a bias for (or against) the favored team, that can confound measuring the home team bias. To specifically test for possible biases for the favorite and the home teams, Golec and Tamarkin used the following model:

$$A_i = \beta_1 + \beta_2 P_i + \beta_3 H_i + \beta_4 F_i + \epsilon_i,$$

where H_i is a dummy variable that is one for home teams and zero otherwise, and F_i is another dummy variable that is one if the team is favored and zero otherwise. Here, the test of efficiency is that $\beta_1 = \beta_3 + \beta_4 = 0$ and $\beta_2 = 1$. Their empirical results for NFL games from 1973 to 1987 indicated that bettors tend to underestimate the home field advantage and overestimate the distinction of being the favorite. Interestingly, they showed that the home field bias is disappearing over time while the underdog bias is actually growing. Despite demonstrating these biases, profits are shown to be slim at best in the face of the bookmaker's commission. Neither bias is present in college football.

Sauer et al. (1988) considered explanatory variables beyond the point spread, such as the number of wins prior to this game, fumbles, interceptions, penalties, yards passed,

and so on. Regressing these variables on the difference between the offered and the actual point spreads, they were unable to reject the hypothesis that their coefficients are all jointly zero. They concluded that these variables add essentially no information beyond that already in the point spread. Dana and Knetter (1994) allowed two modifications. Since fumbles, interceptions, and penalties affect the game but are relatively uninformative about a team's ability, they accounted for these unsystematic sources of noise. Further, they used a non-linear function of past point spreads. There is scant support for any of their models achieving the minimum 52.4% winners needed for profitable wagering.

What is the probability that a team favored to win a football game by p points does win the game? Stern (1991) showed that the margin of victory for the favorite is approximately normally distributed with mean equal to the point spread and standard deviation estimated at 13.86. The probability of winning a game is then

$$Pr(F > U | P = p) = 1 - N\left(\frac{-p}{13.86}\right) = \left(\frac{p}{13.86}\right)$$

where F and U represent actual points scored by the favorite and the underdog, respectively, and $N(\cdot)$ is the standard normal cumulative distribution function. A linear approximation to the probability of winning is

$$Pr(F > U | P = p) = 0.50 = 0.03p.$$

This formula is accurate to within 0.0175 for $p < 6$ and is based on data from the 1981, 1983, and 1984 NFL seasons. Data from 1985 and 1986 indicate that the normal approximation is valid outside of the original dataset. This approximation is useful for a variety of applications, for example, estimating the probability distribution of games won by a team, the probability a team makes the playoffs, and the probability distribution of season or playoff outcomes for particular teams.

5. THE BASKETBALL BETTING MARKET

Do athletes have performance that runs in streaks? Gilovich et al. (1985) using data from the 1980–1981 season for the Philadelphia 76ers found that consecutive shots, if anything, were negatively autocorrelated. Hence there is no hot hand. They also let college players take shots while the players and other observers bet on the outcomes. Both players and observers made larger bets after players had just made shots, although bet size and actual performance were uncorrelated.¹¹ Camerer (1989, p. 1257) argued that belief in the hot hand is a mistake generated by the persistent misunderstanding of randomness. People usually expect more alternations and fewer long streaks than actually occur in random series.

¹¹Albright (1993) studied hitting streaks of baseball players and found no evidence of streaks beyond those expected by a statistical model of randomness. Comments on Albright's work and a rejoinder follow the Albright (1993) article in the same journal issue.

If the hot hand is believed to exist within a game then bettors might also believe in hot and cold streaks across games. And if point spreads reflect a mistaken belief in hot hands, then winning-streak teams should do worse than expected. For NBA regular season games from 1983 to 1986, Camerer (1989) found this effect to be very weak. The effect for losing streaks is slightly stronger, but in neither case is the bias sufficient to overcome the bookmaker's transactions costs. Camerer's test is premised on the myth of the hot hand. Using more data and a test that can also detect the presence of the hot hand, Brown and Sauer (1993b) demonstrated that the market believes in the hot hand. Neither the hypothesis that the hot hand is real nor that it is a myth could be rejected, though.

Brown and Sauer (1993a) examined the error term in a point spread pricing model. While the model's ex ante predictions explained 85% of the variation in point spreads, the error term has significant predictive power. Hence, the error term contains unobserved fundamentals, not just noise.

Sauer (1991) showed that the Las Vegas market point spreads offered at 5 PM Eastern time on the day of 5,636 NBA games are an unbiased estimate of the actual difference in scores. In a subsample of 700 games that involved injuries to star players, the teams with the injured stars performed more than a point worse than the point spread. Obviously this is a nonrepresentative sample, though, because it consists of the games in which the injured star did not play, but not the games where the injured player decided after 5 PM to play. Accounting for the likelihood that a star with a nagging injury will play, the point spreads provided unbiased estimates of actual outcomes.

6. LOTTERIES

6.1. Introduction to Lotteries

For thousands of years, choosing by lots has been used as one means of resolving disputes. The first lottery of a more traditional form, where one pays for a chance to win, dates at least to the Middle Ages in Italy, studied by Ziemba, Brumelle, Gautier, and Schwartz (Ziemba et al., 1986, hereafter ZBGS). Prior to this century, lotteries were successfully used in the U.S. for local and state governments, and to fund numerous causes, such as universities. Corruption, fraud, and moral opposition together with lottery restrictions imposed by Congress ended legalized lotteries by the end of the nineteenth century, with 35 states going so far as to explicitly prohibit them in their constitutions (Clotfelter and Cook, 1991, p. 38). State lotteries continued to be non-existent¹² until 1964 when New Hampshire introduced its lottery. Since then, the U.S. has seen an explosive resurgence of lotteries. In 1991, Washington, DC and 32 states offered lotteries. Furthermore, ticket sales across states exceeded \$19 billion. Of that

¹²Other lottery possibilities were available, such as charity raffles, foreign lotteries like the Irish Sweepstakes, and illegal lotteries.

amount, prizes were \$10.4 billion, administration including advertising was \$1.1 billion, and net revenue was \$7.6 billion.¹³ Clotfelter and Cook (1990) mention that in the course of a year, 60% of the adults who live in lottery states play the lottery at least once (p. 105). They also report that per capita sales in lottery states increased from (in 1989 dollars) \$22 in 1975 to \$108 in 1989 (p. 105). The present popularity of lotteries is more widespread than just the U.S.; in 1986, over 100 countries offered legalized lotteries (ZBGS, 1986, p. 2). See ZBGS for more on the history and the practice of lotteries.

Despite their popularity, with expected returns typically of 40–60%, lotteries are usually a poor investment.¹⁴ This range is even lower (10–20%) if prizes are not tax-free or if they are paid in installments over, say, 20 years, as they typically are in the U.S. Canadian and UK prizes are paid in cash and are tax-free. (See ZBGS for calculations on the effects of tax and payment in installments.)

Lotteries take several forms. A simple version has players buy prenumbered tickets followed by a random drawing. Instant scratch-off games allow one to determine immediately if a prize has been won. Another form is the numbers game that requires players to match a randomly generated three- or four-digit number. Lastly, players of lotto games attempt to match five to seven numbers (with six most common) drawn from 50 or so numbers (with 49 most common), with the actual choice of the parameters varying state by state. A feature distinguishing the numbers and lotto games from the other two forms is the player's act of choosing his or her numbers. For reasons not easily explained by traditional economics, the feature of choice is of tremendous importance. This was illustrated by Langer (1975) who conducted two lotteries where tickets cost \$1 and all the money collected was awarded to the winner, that is, the payback was 100%. Players in the first lottery were assigned their tickets while those in the second lottery chose theirs. As the winner was randomly drawn, subjects in both lotteries had the same chance of winning. However, Langer found that ticket holders in the two lotteries viewed their situations differently. When individually approached to sell their tickets before the drawing, those in the first lottery demanded a mean payment of \$1.96, while in the second lottery the mean was \$8.67. Langer referred to this phenomenon as the illusion of control, that choosing one's ticket improves in some way the likelihood one will win. States seem to appreciate this phenomenon and lotteries involving choices are very common.

The prenumbered and instant scratch-off games allow a state to establish winning payoffs that exactly conform to any payback percentage. For instance, if the instant scratch-off game has \$1 tickets and a \$100 prize, then a 40% payback can be guaranteed by printing 0.4% winning tickets. The numbers game can also involve fixed payoffs. For instance, if the game is to pick the three-digit number that is randomly drawn from the 1,000 possible three-digit numbers, then a prize of \$400 is a 40% payback. The difference here is that the state averages a 60% return, but this return is not guaranteed. If the winning number has disproportionately many bettors, then the state's return will

¹³See State Government Finances in 1991 (Washington, DC, Government Printing Office), Table 35.

¹⁴An exception was the inaugural offering of a new lottery in British Columbia. To create a keen interest in its game, participants received six tickets for the price of one, for an expected return of \$0.385 times 6 or \$2.31, a 131% edge. See Ziemba (1995).

be less than 60%, and the possibility exists that it could even be negative. Despite this difference to the state, the advice to bettors remains: no profitable betting scheme exists for lottery games of this sort and each bet's expected return equals the state's payback percentage.

For the numbers game, Clotfelter and Cook (1993) document a tendency for the public to choose numbers relatively less often immediately after they have been drawn. They describe this pattern as a form of the gambler's fallacy, the belief that if an event just occurred, then the likelihood that it will occur again falls.¹⁵

6.2. Inefficiencies with Unpopular Numbers

Fixed payoffs for lotteries are not the only possibility. Pari-mutuel payoffs are used by all states for lotto games and by Massachusetts for its numbers game. The pari-mutuel method allows a state to guarantee its percentage take by having the payoff to winners decreasing in the number of winners. Given that all numbers are equally likely,¹⁶ no system can be developed that will improve the likelihood of winning any of the lotteries that have been described. However, if a numbers or lotto game employs pari-mutuel payoffs, then by choosing unpopular numbers, upon winning, one is likely to share the given prize with fewer other winners. If some numbers are sufficiently unpopular, bets with positive expected return may exist, despite the lottery's low payout rate. Chernoff's (1980, 1981) study of the Massachusetts number game, where players pick a number from 0,000 to 9,999, found that numbers with 0s, 9s, and to a lesser extent 8s, tended to be unpopular. He showed that by concentrating on the unpopular numbers, bets with a positive expected return were possible. Clotfelter and Cook (1991) provided some evidence of this, too, with three days of 1986 data from Maryland's three-digit numbers game. The most popular three-digit choice was 333 which was 9.93 times more common than the average. The seven most popular choices were all triples—333, 777, 555, 444, 888, 666, and 999—and all were at least five times more popular than the average number. The least popular was 092, picked 0.23 times as often as the average number, and was followed in unpopularity by 086, 887, 884, and 968, all 0.25 times as popular as the average.

Lotto, with its possibility of prizes of tens of millions of dollars, is one of the most popular games, and it has received the most media attention. It involves matching six numbers drawn without replacement from 50 or so total possible numbers. If T is the total possible numbers and D is the number drawn, then the probability of matching is 1 in $T!/(D!(T-D)!)$. So, for example, the probability of winning when six numbers

¹⁵Metzger (1985) considered the gambler's fallacy at the racetrack, and found support for the hypothesis that betting on the favorite should be more attractive after a series of longshots have won than after a series of wins by favorites.

¹⁶Johnson and Klotz (1993), on the basis of 200 Lotto America winning combinations, suggest that each number may not be equally likely. They find that, roughly, small numbers are drawn more frequently than large numbers. They suggest that it may be a consequence of the mechanical mixing process, that small-numbered balls are dropped into the urn first.

are drawn from 49 is 1 in 13,983,816. Most games have prizes for matching fewer than all the drawn numbers, too, but it is common for half the prize money to go to the grand prize. The long odds mean that none of the perhaps millions of bettors might win in a given week (the usual period over which lotto is played). In this event, the grand prize jackpot is carried over to the next week. ZBGS studied whether unpopular numbers and the carryover can allow a profit. Using several methods, they determined that there were unpopular numbers, they were virtually the same ones year to year, and they tended to be high numbers (non-birthdays, etc.) and those ending in 0s, 9s, and 8s. For instance, a regression method based on actual payoffs generated the following as the 12 most unpopular numbers: 32, 29, 10, 30, 40, 39, 48, 12, 42, 41, 38, and 18. They were 1,530% less popular than average. The most popular number, 7, was selected nearly 50% more often than the average number. Using a maximum entropy distribution approach, Stern and Cover (1989) identified 20, 30, 38, 39, 40, 41, 42, 46, 48, and 49 as the 10 most unpopular numbers while 3, 7, 9, 11, 25, and 27 were the six most popular.¹⁷

ZBGS showed that expected returns of \$1.50 without carryover and up to \$2.25 with carryover per dollar bet are possible. Does this imply that lotto games can be profitable, though? To see that it may not, consider a hypothetical game where you pay \$1, choose a number between 1 and 1,000,000, and if your number matches the one that is randomly selected, then you win \$2,000,000. In spite of your edge, you are likely to go bankrupt before winning the jackpot. A reduced wager will increase the likelihood that you will eventually hit the correspondingly reduced jackpot before you go bankrupt, but your expected wealth will suffer. MacLean et al. (1992) analyzed this problem using a model contrasting the growth of wealth and the security of wealth and found that lotteries are an impractical way for modestly endowed investors to enhance their long-term wealth. For instance, by wagering an optimally small amount each round, one's initial stake can be increased tenfold before losing half the stake with a probability close to one. However, millions of years of wagering are required, on average. For example, consider the hypothesized data in Table 7 and the results in Figure 7. With a more attractive set of prizes, the probability is arbitrarily close to one for sufficiently small wagers (see MacLean et al., 1992).

Rather than make optimally small wagers in the face of small probability gambles, growth may be improved by increasing the probability of success. For lotteries, this can be accomplished by buying more than one combination of numbers. It may even be possible in the face of a substantial carryover to profitably purchase most, or perhaps all, of the combinations. There have been times when this would have been profitable. In practice, though, the transaction costs are enormous because tickets must be purchased one at a time. Furthermore, there is the worry that others might also be covering all the numbers, to your joint detriment.¹⁸

¹⁷ See also Joe (1987). Clotfelter and Cook (1991) provided another example of popular numbers from Maryland's lotto, which has 40 total possible numbers. On the particular day they analyzed, players picked the 1-2-3-4-5-6 combination over 2,000 times more frequently than the average pick. Had this been the winning combination (at a chance of one in 3,838,380), winners would have collected only \$193.50!

¹⁸ A related opportunity arises with horse racing pick-sixes (pick the winners of six consecutive races) if there are substantial carryovers. Covering all pick-six possibilities is easily accomplished at the track and may be profitable if few others behave likewise.

TABLE 7 Lotto 6/49 Data

Prizes	Probability	Value	Contribution
Jackpot	1/13,983,816	\$6,000,000	42.9
Bonus	1/2,330,636	\$800,000	34.3
5/6	1/55,492	\$1,000,000	9.0
4/6	1/1,032	\$5,000	14.5
3/6	1/57	\$150	17.6
Edge			18.1%
Kelly bet			0.00000011
Number of tickets with \$10,000,000 bankroll			11

Source: MacLean et al. (1992).

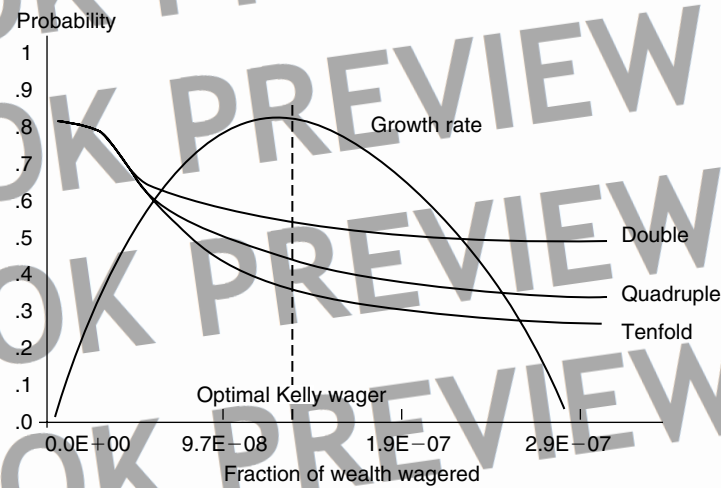


FIGURE 7 Lotto 6/49—Probability of multiplying before losing half of one’s fortune vs. bet size. Source: MacLean et al. (1992).

Lotto typically involves drawing six numbers. Different states have different total possible numbers, though, resulting in very different probabilities of winning. In 1990, the extremes were one chance in 974,000 (36 numbers and two picks per ticket) in Delaware and one chance in 22,957,480 (53 total numbers and one pick per ticket) in California (Cook and Clotfelter, 1993, p. 635). Cook and Clotfelter (1993) explain this as a trade-off that states must make between the size of the jackpot and a player’s estimate of the likelihood that he or she will win. The former is easily learned through advertisements and the media. The latter, according to Cook and Clotfelter, is generally not well understood but tends to be based on the frequency with which someone wins (p. 634). Thus, Delaware could increase its total possible numbers to 53 as in California but, given its population, on average there would be many weeks between

winners. This would lower the public's view of the likelihood of winning and the attractiveness of purchasing a ticket. On the other hand, given California's population, even with 53 total possible numbers there will usually be a winner each week. This nonrational means of probability assessment causes a scale effect whereby per capita expenditure increases with the population base of the lottery. Smaller states cannot exploit this scale effect themselves but can through forming consortia with other states, as happens with the Tri-State lottery (involving Maine, New Hampshire, and Vermont) and the states constituting Lotto America.

References

- Abbate, G. 1995. Gambling Afflicts Ontarians, Study Says, *The Globe and Mail*, August 11, A1.
- Akst, D. 1989. This Is Like Stealing, *Forbes*, November 13, 142–144.
- Albright, C. 1993. A Statistical Analysis of Hitting Streaks in Baseball, *Journal of the American Statistical Association* 88(424), 1175–1183.
- Algeot, P., and T. Cover. 1988. Asymptotic Optimality and Asymptotic Equipartition Properties of Log-Optimum Investment, *Annals of Probability* 16, 875–898.
- Ali, M. 1977. Probability and Utility Estimates for Racetrack Bettors, *Journal of Political Economy* 85, 803–815.
- Ali, M. 1979. Some Evidence of the Efficiency of a Speculative Market, *Econometrica* 47, 387–392.
- Arid R. 1987. A Monthly Effect in Stock Returns, *Journal of Financial Economy* 18, 161–174.
- Asch, P., B. Malkiel, and R. Quandt. 1982. Racetrack Betting and Informed Behavior, *Journal of Financial Economy* 10, 187–194.
- Asch, P., B. Malkiel, and R. Quandt. 1984. Market Efficiency in Racetrack Betting, *Journal of Business* 57, 65–75.
- Asch, P., B. Malkiel, and R. Quandt. 1986. Market Efficiency in Racetrack Betting: Further Evidence and a Correction, *Journal of Business* 59, 157–160.
- Asch, P., and R. Quandt. 1986. *Racetrack Betting: The Professors' Guide to Strategies*. Auburn House, Dover, MA.
- Asch, P., and R. Quandt. 1987. Efficiency and Profitability in Exotic Bets, *Economica* 59, 278–298.
- Asch, P., and R. Quandt. 1988. Betting Bias in Exotic Bets. *Economic Letters* 28, 215–219.
- Aucamp, D. 1993. On the extensive Number of Plays to Achieve Superior Performance with the Geometric Mean Strategy, *Management Science* 39(9), 1163–1172.
- Bacon-Shone, J., V. S. Lo, and K. Busche. 1992a. Modelling Winning Probability. Working Paper, University of Hong Kong.
- Bacon-Shone, J., V. S. Lo, and K. Busche. 1992b. Logistic Analyses for Complicated Bets. Working Paper. University of Hong Kong.
- Benter, W. 1994. Computer Based Horse Race Handicapping and Wagering Systems: A Report, in D. B. Hausch, V. S. Lo, and W. T. Ziemba (eds.), *Efficiency of Racetrack Betting Markets*. Academic Press, San Diego, CA, pp. 183–198.
- Bain, R., D. B. Hausch, and W. T. Ziemba. 2006. An Application of Expert Information to Win Betting on the Kentucky Derby, 1981–2005, *European Journal of Finance* 12, 283–301.
- Betton, S. 1994. Post Position Bias: An Econometric Analysis of the 1987 Season at Exhibition Park, in D. B. Hausch, V. S. Lo, and W. T. Ziemba (eds.), *Efficiency of Racetrack Betting Markets*. Academic Press, San Diego, CA, pp. 511–521.
- Beyer, A. 1978. *My \$50,000 Year at the Races*. Harcourt, Brace and Jovanovich, New York.
- Beyer, A. 1993. *Beyer on Speed*. Houghton Mifflin, Boston.
- Blough, S. 1994. Differences of Opinions at the Racetrack, in D. B. Hausch, V. S. Lo, and W. T. Ziemba (eds.), *Efficiency of Racetrack Betting Markets*. Academic Press, San Diego, CA, pp. 323–341.

- Bolton, R., and R. Chapman. 1986. Searching for Positive Returns at the Track: A Multinomial Logit Model for Handicapping Horse Races, *Management Science* 32, 1040–1060.
- Breiman, L. 1961. Optimal Gambling Systems for Favorable Games, in *Proceedings of the Fourth Berkeley Symposium*, University of California Press, Berkeley, CA, pp. 65–85.
- Brown, W., and R. Sauer. 1993a. Fundamentals or Noise? Evidence from the Point Spread Betting Market, *Journal of Finance* 48(4), 1193–1209.
- Brown, W., and R. Sauer. 1993b. Does the Baseball Market Believe in the Hot Hand? Comment, *American Economic Review* (December), 1377–1386.
- Busche, K. 1994. Efficient Market Results in an Asian Setting, in D. B. Hausch, V. S. Lo, and W. T. Ziemba (eds.), *Efficiency of Racetrack Betting Markets*, Academic Press, San Diego, CA, pp. 615–616.
- Busche, K., and C. Hall. 1988. An Exception to the Risk-Preference Anomaly, *Journal of Business* 61, 337–346.
- Cadsby, C. 1992. The CAPM and the Calendar: Empirical Anomalies and the Risk-Return Relationship, *Management Science*, 38(11), 1543–1561.
- Camerer, C. 1989. Does the Basketball Market Believe in the Hot Hand? *American Economic Review* (December), 1257–1261.
- Canfield, B., B. Fauman, and W. T. Ziemba 1987. Efficient Market Adjustment of Odds Prices to Reflect Track Biases, *Management Science* 33, 1428–1439.
- Chapman, R. 1994. Still Searching for Positive Returns at the Track: Empirical Results from 2,000 Hong Kong Races, in D. B. Hausch, V. S. Lo, and W. T. Ziemba (eds.), *Efficiency of Racetrack Betting Markets*, Academic Press, San Diego, CA, pp. 173–181.
- Chernoff, H. 1980. An Analysis of the Massachusetts Numbers Game. Tech. Report No. 23, Department of Mathematics, MIT, Cambridge, MA.
- Chernoff, H. 1981. How to Beat the Massachusetts Number Game: An Application of Some Basic Ideas in Probability and Statistics, *Mathematical Intelligencer* 3(4), 166–172.
- Clotfelter, C., and P. Cook. 1990. On the Economics of State Lotteries, *Journal of Economic Perspectives* 4(4), 101–119.
- Clotfelter, C., and P. Cook. 1991. *Selling Hope: State Lotteries in America*. Harvard University Press, Cambridge, MA.
- Clotfelter, C., and P. Cook. 1993. The Gamblers Fallacy in Lottery Play, *Management Science* 39(12), 1521–1525.
- Conlisk, J. 1993. The Utility of Gambling, *Journal of Risk Uncertainty* 6, 255–275.
- Cook, P., and C. Clotfelter. 1993. The Peculiar Scale Economies of Lotto, *American Economic Review* 83(3), 634–643.
- Dana, J. D., and M. Knetter. 1994. Learning and Efficiency in a Gambling Market, *Management Science* 40(10), 1317–1328.
- Dansie, B. 1983. A Note on Permutation Probabilities, *Journal of the Royal Statistical Association* 68, 312–316.
- Dolbear, F. 1993. Is Racetrack Betting on Exactas Efficient? *Economica* 60, 105–111.
- Epstein, R. 1977. *The Theory of Gambling and Statistical Logic*, Academic Press, San Diego, CA.
- Fabricand, B. 1965. *Horse Sense*. McKay, New York.
- Fama, E. 1970. Efficient Capital Markets: A Review of Theory and Empirical Work, *Journal of Finance* 25, 383–417.
- Fama, E. 1991. Efficient Capital Markets: II, *Journal of Finance* 46, 1575–1617.
- Figlewski, S. 1979. Subjective Information and Market Efficiency in a Betting Model, *Journal of Political Economy* 87, 75–88.
- Friedman, M., and L. Savage. 1948. The Utility Analysis of Choices Involving Risk, *Journal of Political Economy* 56, 279–304.
- Gandar, J., R. Zuber, T. O'Brien, and B. Russo. 1988. Testing Rationality in the Point Spread Betting Market, *Journal of Finance* (September), 995–1008.
- Garrett, T. A., and G. A. Wagner. 2003. State Government Finances: World War II to the Current Crisis. Working Paper 2003-035A, Federal Reserve Bank of St Louis. Available at <http://research.stlouisfed.org/wp/2003/2003-035.pdf> November 2003.

- Gilovich, T., R. Vallone, and A. Tversky. 1985. The Hot Hand in Basketball: On the Misperception of Random Sequences, *Cognitive Psychology* 17, 295–314.
- Golec, J., and S. Tamarkin. 1991. The Degree of Inefficiency in the Football Betting Market, *Journal of Financial Economy* 30, 311–323.
- Gramm, M., and W. T. Ziemba. 2008. The Dosage Breeding Theory for Horse Race Predictions, in D. B. Hausch and W. T. Ziemba (eds.), *Handbook of Sports and Lottery Markets*. North-Holland, Amsterdam, pp. 307–340.
- Griffith, R. 1949. Odds Adjustments by American Horse Race Bettors, *American Journal of Psychology* 62, 290–294.
- Hakansson, N., and W. T. Ziemba. 1995. Capital Growth Theory, in R. Jarrow, V. Maksimovic, and W. T. Ziemba (eds.), *Finance, Handbooks in Operations Research and Management Science*, Vol. 9. North-Holland, Amsterdam, pp. 123–144.
- Harville, D. 1973. Assigning Probabilities to the Outcome of Multi-Entry Competitions, *Journal of the American Statistical Association* 68, 312–316.
- Hausch, D. B., R. Bain, and W. T. Ziemba. 1995. Betting with the Dr. Z System at the Kentucky Derby, 1934–1995, Working Paper, University of British Columbia.
- Hausch, D. B. and W. T. Ziemba. 1995. Efficiency of Sports and Lottery Betting Markets, in R. Jarrow, V. Maksimovic, and W. T. Ziemba (eds.), *Handbooks in Operations Research and Management Science*, Vol. 9: Finance, Elsevier B. V., Amsterdam.
- Hausch, D. B., V. S. Lo, and W. T. Ziemba (eds.). 1994a. *Efficiency of Racetrack Betting Markets*, Academic Press, San Diego, CA, Second edition (2008). World Scientific, Singapore.
- Hausch, D. B., V. S. Lo, and W. T. Ziemba. 1994b. Pricing Exotic Racetrack Wagers, in D. B. Hausch, V. S. Lo, and W. T. Ziemba (eds.), *Efficiency of Racetrack Betting Markets*, Academic Press, San Diego, CA, pp. 469–483.
- Hausch, D. B., and W. T. Ziemba. 1985. Transactions Costs, Extent of Inefficiencies, Entries and Multiple Wagers in a Racetrack Betting Model, *Management Science* 31, 381–394.
- Hausch, D. B., and W. T. Ziemba. 1990a. Arbitrage Strategies for Cross-Track Betting on Major Horse Races, *Journal of Business* 63, 61–78.
- Hausch, D. B., and W. T. Ziemba. 1990b. Locks at the Racetrack, *Interfaces* 20, 41–48.
- Hausch, D. B., and W. T. Ziemba. 1992. Efficiency of Sports and Lottery Betting Markets, in P. Newman, M. Milgate, and J. Eatwell (eds.), *The New Palgrave Dictionary of Money and Banking*, Macmillan, London, pp. 735–739.
- Hausch, D. B., W. T. Ziemba, and M. Rubinstein. 1981. Efficiency of the Market for Racetrack Betting, *Management Science* 27, 1435–1452.
- Hawawini, G., and D. Keim. 1995. On the Predictability of Common Stock Returns: Worldwide Evidence, in R. Jarrow, V. Maksimovic, and W. T. Ziemba (eds.), *Finance, Handbooks in Operations Research and Management Science*, Vol. 9, North Holland, Amsterdam, pp. 497–544.
- Henery, R. 1981. Permutation Probabilities as Models for Horse Races, *Journal of the Royal Statistical Society* 43, 86–91.
- Henery, R. 1984. An Extreme-Value Model for Predicting the Results of Horse Races, *Applied Statistics* 33, 125–133.
- Hensel, C., G. Sick, and W. T. Ziemba. 1993. The Turn-of-the-Month Effect in the S&P 500 (1928–1993). Mimeo, Frank Russell Company.
- Isaacs, R. 1953. Optimal Horse Race Bets, *American Mathematical Monthly* 60, 310–315.
- Joe, H. 1987. An Ordering Dependence for Distribution of k -tuples, with Applications to Lotto Games, *Canadian Journal of Statistics* 15(3), 227–238.
- Johnson R., and J. Klotz. 1993. Estimating Hot Numbers and Testing Uniformity for the Lottery, *Journal of the American Statistical Association* 88(422), 662–668.
- Kahneman, D., and A. Tversky. 1979. Prospect Theory: An Analysis of Decision under Risk, *Econometrica* 47, 263–291.
- Kahneman, D., and A. Tversky. 1984. Choices, Values, and Frames, *American Psychologist* 39, 341–350.
- Kallberg, J., and W. T. Ziemba. 1994. Parimutuel Betting Models, in D. B. Hausch, V. S. Lo, and W. T. Ziemba (eds.), *Efficiency of Racetrack Betting Markets*. Academic Press, San Diego, CA, pp. 99–107.

- Kanto, A., and G. Rosenqvist. 1994. On the Efficiency of the Market for Double (Quinella) Bets at a Finnish Racetrack, in D. B. Hausch, V. S. Lo, and W. T. Ziemba (eds.), *Efficiency of Racetrack Betting Markets*. Academic Press, San Diego, CA, pp. 485–498.
- Kendall, M. 1953. The Analysis of Economic Time-Series, Part I: Prices, *Journal of the Royal Statistical Society* 96(1), 11–25.
- Kelly, J. 1956. A New Interpretation of the Information Rate, *Bell System Technical Journal* 1, 917–926.
- Kleinfeld, N. 1993. Legal Gambling Faces Higher Odds, *New York Times*, August 29, E3.
- Lakonishok, J., and S. Smidt. 1988. Are Seasonal Anomalies Real? A Ninety-Year Perspective, *The Review of Financial Studies* 1, 403–426.
- Lane, D., and W. T. Ziemba. 2004. Jai Alai Hedging Strategies, *European Journal of Finance*, 353–369.
- Lane, D., and W. T. Ziemba. 2008. Arbitrage and Risk Arbitrage in Team Jai Alai, in D. B. Hausch and W. T. Ziemba (eds.), *Handbook of Sports and Lottery Markets*. North-Holland, Amsterdam, pp. 253–271.
- Langer, E. 1975. The Illusion of Control, *Journal of Personality and Social Psychology* 32(2), 311–328.
- Latané, H. 1959. Criteria for Choice Among Risky Projects, *Journal of Political Economy* 67, 144–155.
- Leong, S., and K. Lim. 1994. Cross-Track Betting: Is the Grass Greener on the Other Side? in D. B. Hausch, V. S. Lo, and W. T. Ziemba (eds.), *Efficiency of Racetrack Betting Markets*. Academic Press, San Diego, CA, pp. 617–629.
- Levin, N. 1994. Optimal Bets in Pari-Mutuel Systems, in D. B. Hausch, V. S. Lo, and W. T. Ziemba (eds.), *Efficiency of Racetrack Betting Markets*, Academic Press, San Diego, CA, pp. 109–125.
- Lo, V. S. 1994. Application of Running Time Distribution Models in Japan, in D. B. Hausch, V. S. Lo, and W. T. Ziemba (eds.), *Efficiency of Racetrack Betting Markets*, Academic Press, San Diego, CA, pp. 237–247.
- Lo, V. S., and J. Bacon-Shone. 2008. Approximating the Ordering Probabilities of Multi-Entry Competitions, in D. B. Hausch and W. T. Ziemba (eds.), *Handbook of Sports and Lottery Markets*. North-Holland, Amsterdam, pp. 51–66.
- Lo, V. S., J. Bacon-Shone, and K. Busche. 1995. The Application of Ranking Probability Models to Racetrack Betting, *Management Science* 41, 1048–1059.
- Lo, V. S., and K. Busche. 1994. How Accurately Do Bettors Bet in Doubles? in D. B. Hausch, V. S. Lo, and W. T. Ziemba (eds.), *Efficiency of Racetrack Betting Markets*. Academic Press, San Diego, CA, pp. 465–468.
- Losey, R., and J. Talbott. 1980. Back on the Track with the Efficient Markets Hypothesis, *Journal of Finance* 35, 1039–1043.
- Ludlow, L. 1994. An Empirical Cross-Validation of Alternative Classification Strategies Applied to Harness Racing Data for Win Bets, in D. B. Hausch, V. S. Lo, and W. T. Ziemba (eds.), *Efficiency of Racetrack Betting Markets*, Academic Press, San Diego, CA, pp. 199–212.
- MacLean, L. C., and W. T. Ziemba. 2006. Capital Growth: Theory and Practice, in S. A. Zenios and W. T. Ziemba (eds.), *Handbook of Asset and Liability Management, Vol. 1: Theory and Methodology*, pp. 429–473.
- MacLean, L. C., W. T. Ziemba, and G. Blazenko. 1992. Growth Versus Security in Dynamic Investment Analysis, *Management Science* 38(11), 1562–1585.
- Markowitz, H. 1952. The Utility of Wealth, *Journal of Political Economy* 60, 151–158.
- McGlothlin, W. 1956. Stability of Choices Among Uncertain Alternatives, *American Journal of Psychology* 63, 604–615.
- Metzger, M. 1985. Biases in Betting: An Application of Laboratory Findings, *Psychological Reports* 56(3), 883–888.
- Mitchell, D. 1986. *Thoroughbred Handicapping as an Investment*. Cynthia Publishing, Los Angeles, CA.
- Pankoff, L. 1968. Market Efficiency and Football Betting, *The Journal of Business* 41, 203–214.
- Plackett, R. 1975. The Analysis of Permutations, *Applied Statistics* 24, 193–202.
- Quandt, R. 1986. Betting and Equilibrium, *Quarterly Journal of Economics* 101, 201–207.
- Quinn, J. 1986. *The Handicappers Condition Book*. William Morrow, New York.
- Quinn, J. 1987. *The Best of Thoroughbred Handicapping (1965–1986)*. William Morrow, New York.
- Quinn, J. 1992. *Figure Handicapping*. William Morrow, New York.

- Quirin, W. 1979. *Winning at the Races: Computer Discoveries in Thoroughbred Handicapping*. William Morrow, New York.
- Quirin, W. 1984. *Thoroughbred Handicapping: State of the Art*. William Morrow, New York.
- Ritter, J. 1994. Racetrack Betting—An Example of a Market with Efficient Arbitrage, in D. B. Hausch, V. S. Lo, and W. T. Ziemba (eds.), *Efficiency of Racetrack Betting Markets*. Academic Press, San Diego, CA, pp. 431–441.
- Ritter, J., and N. Chopra. 1989. Portfolio Rebalancing and the Turn-of-the-Year Effect, *Journal of Finance* 44, 149–166.
- Roberts, H. 1967. Statistical Versus Clinical Prediction of the Stock Market, Mimeo, University of Chicago.
- Roll, R. 1977. A Critique of the Asset Pricing Theory's Tests, *Journal of Financial Economics* 4, 129–176.
- Rosner, B. 1975. Optimal Allocation of Resources in a Pari-Mutuel Setting, *Management Science* 21(9), 997–1006.
- Rosett, R. 1965. Gambling and Rationality, *Journal of Political Economy* 73, 595–607.
- Sauer, R. 1991. An Injury Process Model of Forecast Bias in NBA Point Spreads, Working Paper, Clemson University.
- Sauer, R., V. Brajer, S. Ferris, and W. Marr. 1988. Hold Your Bets: Another Look at the Efficiency of the Gambling Market for National Football League Games, *Journal of Political Economy* (February), 206–213.
- Savage, I. 1957. Contributions to the Theory of Rank Order Statistics—The Trend Case, *Annals of Mathematical Statistics* 28, 968–977.
- Schnytzer, A., and Y. Shilony. 1995. Inside Information in a Betting Market, *Economic Journal* 105, 963–971.
- Shin, H. 1992. Prices of State Contingent Claims with Insider Traders, and the Favorite-Longshot Bias, *Economic Journal* 102, 426–435.
- Shin, H. 1993. Measuring the Incidence of Insider Trading in a Market for State Contingent Claims, *Economic Journal* 103, 1141–1153.
- Snyder, W. 1978. Horse Racing: Testing the Efficient Markets Model, *Journal of Finance* 33, 1109–1118.
- Stern, H. 1987. Gamma Processes, Paired Comparisons and Ranking. Ph.D. Thesis, Stanford University, Department of Statistics, Palo Alto, CA.
- Stern, H. 1990. Models for Distributions on Permutations, *Journal of the American Statistical Association* 85, 558–564.
- Stern, H. 1991. On the Probability of Winning a Football Game, *American Statistician* 45(3), 179–183.
- Stern, H. 1994. Estimating the Probabilities of Outcomes on a Horse Race (Alternatives to the Harville Formulas), in D. B. Hausch, V. S. Lo, and W. T. Ziemba (eds.), *Efficiency of Racetrack Betting Markets*, Academic Press, San Diego, CA, pp. 225–235.
- Stern, H., and T. Cover. 1989. Maximum Entropy and the Lottery, *Journal of the American Statistical Association* 84(408), 980–985.
- Sung, M. C. and J. Johnson. 2008. Semi-Strong form Efficiency in the Horse Race Betting Market, in D. B. Hausch and W. T. Ziemba (eds.), *Handbook of Sports and Lottery Markets*. North-Holland, Amsterdam, pp. 275–306.
- Thaler, R., and W. T. Ziemba. 1988. Parimutuel Betting Markets: Racetracks and Lotteries, *Journal of Economic Perspectives* 2, 161–174.
- Thalheimer, R., and M. Ali. 1995. The Demand for Parimutuel Horse Race Wagering and Attendance, *Management Science* 41(1), 129–143.
- Thorp, E. 1961. A Favorable Strategy for Twenty-One, *Proceedings of the National Academy of Science* 47(1), 110–112.
- Thorp, E. 1962. *Beat the Dealer*. Blaisdell Publishing, New York.
- Thorp, E. 2006. The Kelly Criterion in Blackjack, Sports Betting and the Stock Market, in S. A. Zenios and W. T. Ziemba (eds.), *Handbook of Asset and Liability Management, Vol. 1: Theory and Methodology*, North-Holland, Amsterdam, 385–476.
- Tryfos, P., S. Casey, S. Cook, G. Leger, and B. Pylypiak. 1984. The Profitability of Wagering on NFL Games, *Management Science* (January), 123–132.

- Vergin, R., and M. Scriabin. 1978. Winning Strategies for Wagering on National Football League Games, *Management Science* 1 (April), 809–818.
- Welles, C. 1989. America's Gambling Fever, *Business Week*, April 24, 112–120.
- Weitzman, M. 1965. Utility Analysis and Group Behavior: An Empirical Study, *Journal of Political Economy* 73, 18–26.
- Ziembra, W. T. 1994. World Wide Security Market Regularities, *European Journal of Operations Research* 74, 198–229.
- Ziembra, W. T. 1995. Collection of Dr. Z Columns on Racing, Lotteries, Sports and Casino Gambling, Mimeo, Dr. Z Investments, San Luis Obispo, CA.
- Ziembra, W. T. 2004. Behavioral Finance, Racetrack Betting and Options and Futures Trading, Mathematical Finance Seminar, Stanford University, January 30.
- Ziembra, W. T., S. Brumelle, A. Gautier, and S. Schwartz. 1986. *Dr Zs Lotto 6/49 Guidebook*. Dr. Z Investments, San Luis Obispo, CA.
- Ziembra, W. T., and D. B. Hausch. 1986. *Betting at the Racetrack*, Dr. Z Investments, San Luis Obispo, CA.
- Ziembra, W. T., and D. B. Hausch. 1987. *Dr Zs Beat the Racetrack*. William Morrow, New York.
- Zuber, R., J. Gandar, and B. Bowers. 1985. Beating the Spread: Testing the Efficiency of the Gambling Market for NFL Games, *Journal of Political Economy* 93(4), 800–806.

CHAPTER 11

Point Spread and Odds Betting: Baseball, Basketball, and American Football

Hal S. Stern

University of California, Irvine, CA, USA.

1. Introduction	224
2. Efficiency of Odds Betting Markets	224
2.1. <i>Horse Race Betting</i>	224
2.2. <i>Baseball</i>	226
3. Efficiency of Point Spread Betting Markets	227
3.1. <i>American Football</i>	227
3.2. <i>Basketball</i>	230
4. Relationship of Point Spread and Odds Betting	230
4.1. <i>Normal Distribution Result</i>	231
4.2. <i>Applications of the Normal Approximation</i>	233
5. The Normal Model and Mid-Event Wagering	234
6. Summary	236
References	237

Abstract

Sports betting is extremely popular and many sports offer a variety of wagering options. These include odds betting (also known as the money line) on the simple proposition of who will win the game, and point spread betting in which the proposition concerns the margin of victory. Studies of betting market efficiency in horse racing have been common, but betting markets for other sports are less well studied. This chapter considers the efficiency of the point spread and odds betting markets for a variety of sports; we find that the markets appear to be weakly efficient in the sense that betting based only on previous results is unlikely to yield profitable strategies.

1. INTRODUCTION

In some sports, for example, horse racing worldwide and baseball in the U.S., odds betting is the dominant (often only) form of wagering. In other U.S. sports, such as basketball and American football, there is also point spread wagering. There is much evidence, some of it reviewed in other chapters of this volume, that horse racing odds reflect a form of collective wisdom. Win probability estimates for the horses that are derived from the odds are quite accurate. There is anomalous behavior such as the favorite-longshot bias, but overall, the racetrack betting market achieves a level of efficiency beyond what many might expect initially. This raises a natural question as to whether odds betting in other sports and point spread betting achieve a similar level of market efficiency. For some U.S. sports events, both point spread and odds wagering are offered. In such cases, it is natural to ask how the point spread and odds relate. This chapter offers an empirical exploration of the efficiency of point spread and odds betting in a variety of U.S. sports and an analytic investigation of the relationship between point spread and odds betting. Finally, having a relationship between the point spread and the odds of winning a game allows for some interesting speculation about mid-game betting.

2. EFFICIENCY OF ODDS BETTING MARKETS

We begin with a review of odds and probabilities. Odds are used to represent probabilities in betting situations. If the odds against an event happening are quoted as b/a , then a bettor that believes the event will occur must risk a wager of a units to win a profit of b units. If the event occurs, then the bettor has a profit of b units; if the event does not occur, then the bettor has a loss of a units. This is a fair bet with zero expected profit/loss if the probability that the event occurs is $a/(a + b)$. Thus odds of b/a against an event imply a probability of $a/(a + b)$.

2.1. Horse Race Betting

The horse race betting market at the racetrack uses a pari-mutuel system in which the track takes a portion of all bets (15–18% in most U.S. states) and the remainder of

the pool is used to pay off winning bettors. With rebates now commonly offered to off-track bettors, the effective track take can be reduced to approximately 10% for large players and 12–13% for small players. The odds against each horse's winning the race are calculated based on the remainder of the win pool after the track take is removed. One key point in such a system is that (after the track takes its piece) racetrack bettors are essentially betting against the opinion of the collective group. The odds against each horse's winning the race are derived from the win betting pool and are displayed at the track. One can estimate the implied probability that each horse wins the race from the odds as described above, $O/1$ odds correspond to estimated probability $1/(O + 1)$. If you add up the values of $1/(O + 1)$ for all the horses in a given race, the sum will be greater than one, so these cannot possibly be valid probability estimates. This occurs because of the track take. It is easily remedied by taking the estimated probability for a horse to be proportional to $1/(O + 1)$ and then normalizing the probabilities to ensure that they sum to one.

Though other chapters have already made the case, we use one dataset here to demonstrate the accuracy of the win probabilities derived from race track odds. Table 1 provides results for all horses from a set of 3,785 Hong Kong horse races (38,047 horse race entrants in all) held between 1981 and 1987. The horses are placed in categories based on their estimated probability of winning. Table 1 gives the average estimated win probability for the horses in each category along with the observed proportion of winners. The observed proportions match up quite well with the estimates. As seen elsewhere, for example in Ziemba and Hausch (1987), horses with high estimated probabilities of winning (0.3 or above) actually win more often than expected.

TABLE 1 Probability of Winning Horse Race Derived from Racetrack Odds Compared to Actual Performance

Probability implied by the odds	Number of horses	Expected number of winners	Observed number of winners	Estimated probability of winning	Observed proportion of wins
[0.000, 0.010)	1,830	12.7	15	0.007	0.008
[0.010, 0.025)	5,096	87.0	78	0.017	0.015
[0.025, 0.050)	6,562	242.9	278	0.037	0.042
[0.050, 0.100)	9,718	713.9	695	0.073	0.072
[0.100, 0.150)	6,461	792.6	781	0.123	0.121
[0.150, 0.200)	3,742	643.4	657	0.172	0.176
[0.200, 0.250)	2,167	480.7	451	0.222	0.208
[0.250, 0.300)	1,227	334.6	325	0.273	0.265
[0.300, 0.400)	856	290.4	303	0.339	0.354
[0.400, 0.500)	267	117.7	121	0.441	0.453
[0.500, 1.00)	121	69.0	81	0.570	0.669

NOTE: Data from 3,785 Hong Kong races during the time period 1981–1987.

There is of course no guarantee that probability estimates derived from a collective betting effort such as is found at the racetrack will be accurate. Explanations for the efficiency of the betting market typically emphasize that if the collective effort consistently made large mistakes, then an intelligent observer (or a group of such people) could capitalize on these errors. As the knowledgeable bettors placed their wagers, they would correct the odds and fix the anomalous behavior. The theory holds that small errors like the favorite bias observed in Table 1 continue to exist because there is not enough financial reward for individuals to make the wagers required to correct them. Other chapters in this volume demonstrate that it is in fact possible to capitalize on inefficiencies that arise in the place and show pools at the race track and in various combination bets at the racetrack.

2.2. Baseball

The efficiency of the horse race betting market is well known and the subject of several chapters in this book. A major contribution of this chapter is to consider the efficiency of betting markets in other, primarily U.S., sports. Odds are the dominant form of betting for baseball games in the U.S. Baseball odds are usually quoted in the form “NY – 150/+140 Cleveland.” This expression contains the odds for two bets. The odds against New York winning are 100/150 (i.e., New York is favored) so that a New York supporter must bet \$150 to earn a profit of \$100. The odds against Cleveland winning are 140/100 so that a Cleveland supporter must bet \$100 to earn a profit of \$140. The spread between the two numbers, that is, the lack of symmetry in the odds, leaves the betting shops with their profit. Assuming they can balance the volume of bets appropriately, they can generate a guaranteed profit. For example, if \$1,000 were bet on New York and \$700 on Cleveland, then the game will end up in one of two ways: (1) If New York wins, then New York bettors are owed \$667; (2) If Cleveland wins, then Cleveland bettors are owed \$980. In either case, the losing team’s bets will pay off the winners and leave a profit for the betting shop. There is risk here in that the bookmaker is guaranteed a profit only if the amount bet on Cleveland is between $100/150 = 0.667$ and $100/140 = 0.714$ times the amount bet on New York.

As with horse racing odds, one can derive the implied probability that a team wins a baseball game from the odds that are offered. The 100:150 odds against New York imply that their probability of winning is $150/250 = 0.600$ while the 140/100 odds against Cleveland imply a probability of $100/240 = 0.417$. The two implied probabilities add up to more than one but can be easily renormalized so that, for example, the probability that New York wins is estimated as $0.600/(0.417 + 0.600) = 0.590$. The accuracy of baseball odds is assessed from a modest dataset comprised of games from the 1986 National League season. A total of 969 games were played and odds were obtained from a historical record (Jack Painter’s 1986 Baseball Money Line and Totals Guide) for 937 games. The games are collected in Table 2 grouped according to the odds of the favored team. Table 2 shows that, once again, the actual outcomes are consistent with the probabilities implied by the odds.

TABLE 2 Implied Probability from Baseball Odds Compared to Actual Performance

Favorite's odds	Estimated probability favorite wins	Games played	Games won by favorite	Observed proportion won by favorite
-110	0.51	89	41	0.46
-115	0.52	98	56	0.57
-120	0.53	71	32	0.45
-125	0.54	82	47	0.57
-130	0.55	70	32	0.46
-135	0.56	80	41	0.51
-140	0.57	75	43	0.57
-145	0.58	79	41	0.52
-150	0.59	54	34	0.63
-155	0.60	58	29	0.50
-160 to -165	0.61	67	41	0.61
-170 to -185	0.63	51	36	0.71
-190 to -300	0.67	63	43	0.68

NOTE: Data are from 969 National League baseball games from 1986.

3. EFFICIENCY OF POINT SPREAD BETTING MARKETS

Point spread betting is used in sports like football and basketball. A betting line is established for each game. For example, in the 2008 American football championship (known as the Super Bowl) the New England Patriots were established as a 12 point favorite to defeat the New York Giants. Bettors can bet on New England to win by more than 12 points or they can bet on New York to lose by less than 12 points (or win outright). Bettors bet \$11 to win \$10. If New England wins by exactly 12 points then the bet is cancelled and wagers are returned. The betting line is ideally set at a point that attracts equal amounts of money to each team. If that happens, then the bookmakers are assured a profit since the losers pay the winners with bookmakers keeping \$1 from each \$11 bet on the losing team.

3.1. American Football

The betting line for American football games is regularly reported in newspaper sports sections. During the period from 1981 through 1997 data were recorded by the author for each National Football League game (except the years 1982 and 1986 that were affected by labor actions). For each game, the point spread (recorded from a local newspaper that varied over the period) and the game outcome were recorded. There is some variability in the published point spreads (from day to day and from newspaper

to newspaper), however, that variability is small (typically less than one point) and should not have a large impact on the results described here. An attempt was made to use point spreads from late in the week since these incorporate more information (e.g., regarding injured players) than point spreads from early in the week. For reasons of convenience, the day on which the data was collected varied. Tables 3 and 4 provide some basic evidence regarding the efficiency of the point spread betting line. Table 3 presents results related to the effectiveness of the betting line as a predictor of game outcomes. If F denotes the favored team's score, U denotes the underdog's score, and P denotes the point spread, then the margin of victory over the point spread (M) is defined by $M = F - U - P$. This variable measures the difference between the point spread and the actual game outcome. Table 3 presents the mean and standard deviation of M for each season. The means are quite small and not significantly different from zero according to standard statistical tests. There is a tendency for the means to be negative but this does vary over the time period. As with the horse racing odds, the collective wisdom reflected in the point spread is reasonably accurate.

Table 4 records the outcomes of games separately for each betting line. The point spread seems to be a reliable measure of the relative ability of the teams in that the proportion of games won increases as the point spread increases. The relationship is not perfectly monotone but the observed deviations may be due to the relatively small

TABLE 3 Mean and Standard Deviation of the Differences Between Football Game Point Spreads and Actual Game Outcomes

Year	Number of games	Mean point spread error	SD point spread error
1981	224	-0.7	14.1
1982	224	-0.4	13.8
1984	224	1.3	13.6
1985	224	1.1	13.5
1986	224	-0.5	13.8
1988	224	-0.3	14.3
1989	224	0.7	13.6
1990	224	0.9	13.6
1991	224	0.8	12.6
1992	224	-0.7	13.9
1993	224	-1.0	13.1
1994	224	-1.4	12.3
1995	240	-1.1	12.5
1996	240	-0.5	12.9
1997	240	-1.6	13.3
Total	3,408	-0.24	13.42

NOTE: Data from 3,408 National Football League games over the period 1981–1997.

TABLE 4 Proportion of Games Won by the Favorite for Each Betting Line and the Proportion of Games in Which the Favorite Won by More Than the Betting Line

Point spread P	Games played	Proportion won by favorite	Proportion favorite beat spread
0.0	85	—	—
1.0	141	0.52	0.51
1.5	128	0.47	0.46
2.0	201	0.55	0.51
2.5	231	0.51	0.44
3.0	386	0.62	0.52
3.5	284	0.57	0.47
4.0	200	0.63	0.48
4.5	110	0.67	0.50
5.0	145	0.69	0.50
5.5	111	0.78	0.55
6.0	186	0.65	0.44
6.5	198	0.71	0.50
7.0	209	0.73	0.50
7.5	83	0.78	0.48
8.0	75	0.80	0.49
8.5	52	0.85	0.52
9.0	106	0.75	0.41
9.5	79	0.78	0.52
10.0	82	0.74	0.43
> 10	307	0.83	0.48

NOTE: Games ending in a tie are ignored when computing the proportion won and games ending on the betting line are ignored when computing the proportion won by more than the betting line. (Data from 3,408 NFL games, 1981–1997 [except 1982, 1986].)

number of games at each point spread. The final column of Table 4 provides additional evidence of the efficiency of the point spread betting market. Notice that at each level of the betting line, approximately half of the bets are won by the favorite and half by the underdog. None of the deviations from 50% are statistically significant. The consistency is really quite remarkable. Given the 11/10 odds for point spread wagers, a successful betting strategy must win 52.4% (11 of 21) of wagers (assuming a flat or constant bet size). For almost every row of the table, the proportion of games in which the favorite beats the spread is between 47.6% and 52.4% so that simple strategies based on past outcomes would not be successful. The few exceptions are so close to the cutoff that there is little reason to believe the patterns persist. Some online bookmakers

are reducing the 11/10 odds, which reduces the required win rate, but to this point the effect is a small one. Related evidence of market efficiency in football point spreads was earlier reported by Pankoff (1968), Vergin and Scriabin (1978), Tryfos et al. (1984), and Zuber et al. (1985). Cultural evidence of the efficiency of the football betting market abounds. One online sports columnist picks the games each week against the spread. As an experiment, he began comparing his picks to those of his wife who does not watch any football. Her selections have performed better than his over the two years of the experiment.

3.2. Basketball

A similar analysis was done for professional and college basketball in the U.S. Data from the 1985–1986 and 1986–1987 National Basketball Association seasons was obtained in the same manner as described above for football. Point spreads and game outcomes were recorded from newspaper reports. During the two seasons, a total of 1,886 games were played; point spreads were obtained for 1,880 games. Across the 1,880 games, the mean margin of victory over the point spread is 0.42 and the standard deviation is 11.6. The mean is positive here (it was negative in the football data) but not significantly different from zero. Once again, the point spread appears to be an unbiased estimate of the difference in ability of the competing teams. An analysis of the games according to the point spread indicates once again that for each possible betting line, approximately half of the games are won by more than the point spread and half are not.

It is also possible to wager on college basketball in the U.S. Data from the 1995–1996 regular college basketball season was obtained from the Website of The Gold Sheet, a professional gambling information service (<http://www.goldsheet.com>). The data are results for 152 Division I (the most competitive division) teams. These are not all of the Division I basketball teams, but it includes all of the major conferences. In all, there are game results for 2,656 games; point spreads are available for 2,109 games. The mean value of the 2,109 differences is -0.2 and the standard deviation is 10.9. The mean is not significantly different from zero, which suggests the point spread can again be considered an unbiased estimate of the difference between the two teams. The standard deviation tells us about the variability of the outcomes—there is less variability in college basketball than in professional basketball. This is not a surprise in that the average team score in college basketball during this era is about 70 points, whereas the average professional team score during the 1985–1987 period is about 110 points.

4. RELATIONSHIP OF POINT SPREAD AND ODDS BETTING

The previous sections demonstrate that both odds and point spreads appear to be accurate assessments of the relative strengths of two teams. It is natural to wonder if there is a relationship between the two forms of wagering. There does appear to be a relationship and this relationship is most easily seen by first determining the relationship between the point spread and the probability of winning a game.

4.1. Normal Distribution Result

Though the point spread is not intended as a predictor of the outcome of a game, the results of the previous section indicate that it appears to be a fairly accurate predictor. For example, Table 3 indicates that the mean difference between the point spread and the game outcome is zero. The relationship between the point spread and the game outcome is explored further here, initially in the context of professional football games but then subsequently with basketball data as well.

What can one say about the probability that a P -point favorite wins a football game? A natural estimate is obtained by looking at the proportion of P -point favorites in the sample that have won their game as reported in Table 4. This empirical approach works well but leads to estimates with large standard errors due to the small number of games with any particular point spread. An alternative approach builds on an interesting distributional result for the victory over the point spread, $M = F - U - P$. Figure 1 presents a graphical display of the distribution of M (the difference between the point spread and the game outcome) for 3,000+ professional football games. The figure suggests that the outcomes of National Football League (NFL) games can be approximated as normal random variables. This observation is quite similar to the result in Stern (1991) which was based on a much smaller sample. Table 3 reports a sample mean for M of -0.24 and a standard deviation of 13.42 . For the remainder of this chapter we take the normal distribution with mean 0 and standard deviation 13.5 as the

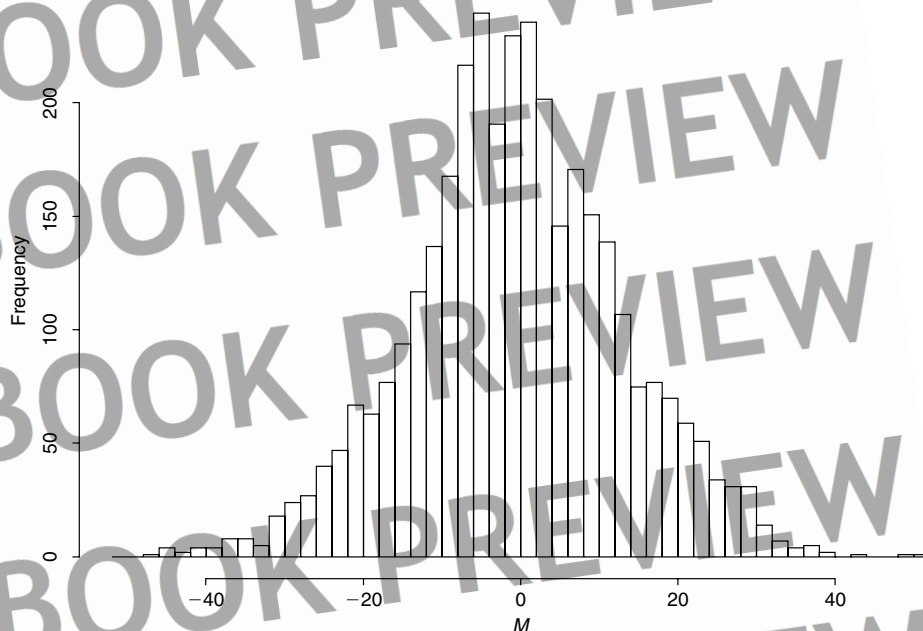


FIGURE 1 Histogram of margin of victory over the point spread supports a normal approximation.

approximating distribution of M . The normal approximation can perhaps be justified by an argument that takes the game outcome as the “sum” of the contributions of a large number of plays/events and thus subject to the Central Limit Theorem. Clearly, though, the normal distribution is just an approximation. For one thing, football scores vary in size (field goals worth 3; touchdowns worth 7), and for another, the number of scores in a game is not always variable M is concentrated on multiples of one-half, and integer values occur twice as often as noninteger values. This would not be the case if normality provided a more exact fit. Indeed with this large sample size the normal approximation is rejected by some statistical tests. Nonetheless the approximation turns out to be useful in a number of contexts, including as a useful way to translate point spreads into probabilities.

The normal approximation can be used to provide improved estimates of the probability that a P -point favorite wins a game. The probability that a team favored by P points wins the game is

$$\Pr(F > U|P) = \Pr(F - U - P > -P|P) = \Pr(M > -P|P).$$

The discussion above shows that $M = F - U - P$ is approximately normal (accumulated over all point spreads). A more detailed analysis indicates that normality appears to be a valid approximation for $M = F - U - P$ conditioned on each value of P . Note that the conditional distribution result is more difficult to demonstrate since there are not many games with any particular value of P . A series of statistical tests were performed for games with similar point spreads, and these tests seem to indicate that normality is an adequate approximation for each range of point spreads. If we apply normality for a particular point spread, P , then $F - U$ is approximately normal with mean P and standard deviation 13.5. The probability of winning a game is then computed as

$$\Pr(F > U|P) = \Pr(M > -P|P) = 1 - \Phi\left(-\frac{P}{13.5}\right) = \Phi\left(\frac{P}{13.5}\right)$$

where $\Phi(\cdot)$ is the cumulative distribution function of the standard normal random variable.

The normal approximation for the probability of victory is given for some sample point spreads in Table 5 along with the observed proportion of P -point favorites that won their games (repeated from Table 4). The estimates from the normal formula are consistent with the estimates made directly from the data. In addition, they are monotone increasing in the point spread. This is consistent with the interpretation of the point spread as a measure of the difference between two teams.

A similar distributional result appears to hold for the professional and college basketball data. A normal approximation again appears to be a useful approximation. The mean of the normal approximation is the point spread; the standard deviation of the outcome appears to vary from sport to sport. In professional football the standard deviation is approximately 13.5; for professional basketball the estimated standard deviation is 11.6, and for college basketball the estimated standard deviation is 11.0. The standard

TABLE 5 Probability that a P Point Favorite Wins a Professional Football Game Based on Normal Approximation Along with Empirical Results

Point spread P	Normal approximation to win probability	Proportion won by favorite
1	0.53	0.52
2	0.56	0.55
3	0.59	0.62
4	0.62	0.63
5	0.64	0.69
6	0.67	0.65
7	0.70	0.73
8	0.72	0.80
9	0.75	0.75
10	0.77	0.74
12	0.81	—
14	0.85	—
17	0.90	—
21	0.94	—

NOTE: Data from 3,408 NFL games, 1981–1997 (except 1982, 1986).

deviation is a key parameter in translating the point spread to the probability of winning and thence to the odds for a game. One can think of the standard deviation of M in the same way that one thinks of volatility in valuing financial assets.

4.2. Applications of the Normal Approximation

The normal approximation formula for the probability of winning a game as a function of the point spread has a number of applications. Most simply, it can be used as a form of calibration. For example, if two professional football teams are to play and we believe the stronger team might win seven of 10 times in such games, then we implicitly are assigning a point spread of $13.5 * \Phi^{-1}(0.7) = 7$ points to the game. This can be compared to the published point spread as a way of calibrating our opinion against the betting line. As another example, the home-field advantage in college basketball is generally assumed to be worth approximately five points. Assuming a normal approximation with standard deviation of 11 for college basketball, the probability that a five point favorite wins is 0.68; this is exactly the observed winning percentage for home teams in the 1995–1996 college basketball data.

Beyond its uses as a method for calibrating probabilistic assessments and point spread assessments, the normal approximation also provides an opportunity for

examining the relationship between odds and point spread wagering. The 2008 Super Bowl provides an instance where this may have been useful knowledge. Recall that the New England Patriots were regarded as a 12 point favorite to defeat the New York Giants. For significant events such as the Super Bowl, oddsmakers frequently offer both point spread and odds wagering. One oddsmaker offered the New England Patriots as $-450/+350$ favorites to win the Super Bowl. Note that the Patriots' portion of the odds appears well calibrated; the probability implied by odds of $100/450$ is 0.818 which corresponds to a point spread of 12.3 points. There is good agreement between the odds wager and the point spread wager for a Patriots bettor. However, the odds received by a Giants bettor, $350/100$ imply the Giants probability of winning is 0.222 which corresponds to a point spread of 10.3 points or so. For a Giants fan, the point spread wager offers the better value!

5. THE NORMAL MODEL AND MID-EVENT WAGERING

The normal approximation derived in the previous section pertains to end of game results. Stern (1994) examines professional basketball data and finds that the normal approximation appears to work well at shorter time scales also. That work examines results for games between a generic "home" and "visiting" team but did not use point spread data. The results of the previous section suggest that the point spread can serve as a measure of the expected difference in ability between two teams in generalizing the results of Stern (1994). Assume that Team A is a P point favorite to defeat Team B and that the standard deviation for the normal approximation is σ . We follow the argument in Stern (1994) by initially transforming the time scale of a sporting event to the unit interval. Then a time $t \in (0, 1)$ refers to the point in a sports contest at which a fraction t of the contest has been completed. Define $X(t)$ as a random variable giving the advantage at time t of the favored team. Then assuming a Brownian motion model (i.e., a normal model) is a reasonable approximation for all t in the unit interval one can take $X(t) \sim N(Pt, \sigma^2 t)$. Of course, the model is not correct in that $X(t)$ is integer-valued, but it turns out to be a useful approximation. We return to the limitations of the model below. Another consequence of assuming a Brownian motion model for $X(t)$ is that $X(s) - X(t)$, $s > t$, the change in the score from time t to time s , is independent of $X(t)$, the score difference at time t , with

$$X(s) - X(t) \sim N(P(s - t), \sigma^2(s - t)).$$

At $t = 1$ (the end of the game) the normal model indicates the probability that the favorite wins as $\Pr(X(1) > 0) = \Phi(P/\sigma)$, which matches our earlier expression. Under the Gaussian process model, however, we can do much more. The probability that the

avored team wins (i.e., $X(1) > 0$) given that they have an l point advantage (or deficit) at time t (i.e., $X(t) = l$) is

$$\begin{aligned}\Pi_{P,\sigma}(l,t) &= \Pr(X(1) > 0 \mid X(t) = l) = \Pr(X(1) - X(t) > -l) \\ &= \Phi\left(\frac{l + (1-t)P}{\sqrt{(1-t)\sigma^2}}\right)\end{aligned}$$

where Φ is the cumulative distribution function of the standard normal distribution. The formula is quite intuitive; the outcome of the game from time t to the end is approximated by a normal random variable with mean equal to the current score difference plus the remaining fraction of the presumed pre-game advantage P and variance equal to the remaining fraction of the presumed pre-game variance σ^2 . Of course, as $t \rightarrow 1$ for fixed $l \neq 0$, the probability tends to either zero or one, indicating that any lead is critically important very late in a game. For fixed t , the lead l must be relatively large compared to the remaining variability in the contest in order for the probability of winning to be substantial.

There are several important limitations that must be recognized before applying the normal approximation to mid-game probability calculations. One limitation arises from the fact that the normal approximation assumes a continuous distribution on scores when of course the distribution is discrete. This can be addressed via a continuity correction as in Stern (1994). A more important limitation is that the computed probability of winning does not account for which team is in possession of the ball at time t and therefore has the next chance to score. This is crucial information in the last few minutes of a game (e.g., $t > 0.96$ in a 48 min basketball game). There is a very large difference in the last minute of a basketball game between being a team with a one-point lead and having possession of the ball and being a team with a one-point lead that does not have the ball. Thus, the probability of winning should change radically with possession of the ball, yet this factor is not incorporated in the above. The results from the formula should thus be viewed with great caution late in the game. The point at which the approximation would be expected to fail depends on various characteristics of the sport including the length of a typical possession and the number of points that can be scored on a single possession. The approximation is likely best for basketball in which each possession lasts a short time (less than 24 sec in the professional game) and each team can score at most three points on a possession.

Table 6 gives the probability of winning for several values of l, t for two different point spreads. The top half of the table displays results for a five point favorite while the bottom half displays results for a 10 point favorite. Results for five or 10 point underdogs ($P = -5$ or -10) can be easily obtained by noting that $\Pi_{P,\sigma}(l,t) = 1 - \Pi_{P,\sigma}(-l,t)$. At the start of the game the five point favorite has a better than 50% chance of winning as should be true. Note that even if it is behind by two points at halftime ($t = 0.50$) the favored team is still given a better than 50% chance to win the game. Under the Brownian motion model it is not possible to obtain a tie at $t = 1$ so this cell is blank; we

TABLE 6 Probability that a Favorite Wins a Professional Basketball Game Given a Lead l After a Fraction t of the Game Assuming a Standard Deviation $\sigma = 11.6$

Elapsed time t	Lead l						
	$l = -10$	$l = -5$	$l = -2$	$l = 0$	$l = 2$	$l = 5$	$l = 10$
Probability $P = 5$ point favorite wins game:							
0.00				.67			
0.25	.27	.45	.57	.65	.72	.81	.91
0.50	.18	.38	.52	.62	.71	.82	.94
0.75	.07	.26	.45	.55	.75	.86	.97
0.90	.005	.11	.34	.55	.75	.93	.998
1.00	.00	.00	.00		1.00	1.00	1.00
Probability $P = 10$ point favorite wins game:							
0.00				.81			
0.25	.40	.60	.71	.77	.83	.89	.96
0.50	.27	.50	.64	.73	.80	.89	.97
0.75	.10	.33	.53	.67	.78	.90	.98
0.90	.01	.14	.39	.61	.79	.95	.99
1.00	.00	.00	.00		1.00	1.00	1.00

NOTE: Top half of table assumes a five point favorite; bottom half assumes a 10 point favorite.

might think of the value there as being approximately 0.50. In professional basketball $t = 0.9$ corresponds roughly to 5 min remaining in the game. Notice that having the favorite comeback from five points in the final 5 min is not terribly unusual (about 1 in 10 chance) while a 10 point comeback is considerably more unusual (about 1 in 200 chance). This latter value seems somewhat low. Stern (1994), using a different dataset, finds a much greater value of $\sigma = 15.8$. It may be that the larger value adjusts for some of the limitations mentioned earlier and yields more realistic probabilities near the end of a game.

6. SUMMARY

Data collected from a variety of U.S. sports leagues argues for weak-form efficiency of the betting markets in these sports. Extensive data from professional football shows that the betting point spread is an approximately unbiased estimate of the expected game outcome. Similar results hold for college and professional basketball. These results argue for efficiency in point spread betting markets. In addition, win probabilities derived from

the odds of professional baseball games appear to match the empirical win probabilities. Thus odds wagering markets also appear to satisfy weak efficiency.

Additional data analysis for those sports in which point spread betting is popular suggests a useful normal approximation for game outcomes. This approximation allows one to relate point spreads to win probabilities and to betting odds. Moreover, with some additional assumptions, the normal model can be used to make probabilistic assessments of game outcomes during the game, which has implications for mid-game betting.

References

- Pankoff, L. D. 1968. Market Efficiency and Football Betting, *Journal of Business* 41, 203–214.
- Stern, H. S. 1991. On the Probability of Winning a Football Game, *The American Statistician* 45, 179–183.
- Stern, H. S. 1994. A Brownian Motion Model for the Progress of Sports Scores, *Journal of the American Statistical Association* 89, 1128–1134.
- Stern, H. S. 1997. How Accurate Are the Posted Odds? *Chance* 11(4), 17–21.
- Tryfos, P., S. Casey, S. Cook, G. Leger, and B. Pylypiak. 1984. The Profitability of Wagering on NFL Games, *Management Science* 30, 123–132.
- Vergin, R. C., and M. Scriabin. 1978. Winning Strategies for Wagering on Football Games, *Management Science* 24, 809–818.
- Ziembra, W. T., and D. B. Hausch. 1987. *Dr. Z's Beat the Racetrack*, William Morrow, New York.
- Zuber, R. A., J. M. Gandar, and B. D. Bowers. 1985. Beating the Spread: Testing the Efficiency of the Gambling Market for National Football League Games, *Journal of Political Economy* 93, 800–806.

DOWNLOAD FULL VERSION HERE: <https://bettingbooks.org/12>

This page intentionally left blank

Download More Sports Betting Books at: www.bettingbooks.org

CHAPTER 12

Comparing Efficiency of the Over/Under-Bets on NFL and NBA Games

Joseph Golec

Associate Professor, Department of Finance, School of Business, University of Connecticut,
Storrs, CT, USA.

Maurry Tamarkin

Professor of Finance, Graduate School of Management, Clark University, Worcester, MA, USA.

1. Introduction	240
2. The Sports Betting Market: Setting Point Spreads and Over/Unders	240
3. NFL and NBA Betting Market Efficiency	242
4. Conclusion	249
<i>References</i>	251

1. INTRODUCTION

Sports betting markets are attractive research candidates because they are similar to traditional financial markets in many aspects but are often much simpler to study. Both expected values and outcomes are more clearly observable over fixed betting (investment) horizons. Earlier studies typically focus on the efficiency of a single type of bet in a particular sports market. For example, Brown and Abraham (2002) study over/under betting on professional baseball, and Paul and Weinbach (2002) do the same for football. Our study of over/under-bets in the U.S. professional football league (NFL) and the U.S. professional basketball association (NBA) compares the efficiency of a single bet across two sports.

In Golec and Tamarkin (2002), we identified some inefficiencies in over/under betting on NFL games, applying the usual regression and profitability tests. We started by identifying inefficiencies in the aggregate sample, but also showed that they tended to be stronger in certain years and nonexistent or reversed in others. Like most earlier gambling studies, the aggregate inefficiencies were statistically significant but not particularly large. Hence, one might question whether these inefficiencies represent a systematic gambling tendency by bettors, or just a chance result, perhaps due to some quirk of the NFL market data in particular years. For example, football point scoring is lumpy, usually in 7s and 3s, so some runs of unusual outcomes in a particular year could drive the results. If predictions by line-setters are off by one score, they can be relatively far away from outcomes. Indeed, the NFL sample size is relatively small because teams play 16 or fewer games per regular season. Hence, a relatively small number of unusual outcomes could drive results.

We now propose to run the same tests for NBA games, where each team plays 82 regular season games and scoring is less lumpy (in 1s, 2s, or 3s). The idea is to see if the same types of inefficiencies arise in the same years as in the NFL market. If they do, then we can be more confident that the inefficiencies observed in the two markets are legitimate. The underlying assumption is that the behavior of the marginal NFL gambler is similar to that of the marginal NBA gambler. Indeed, the two pools of bettors probably overlap. Nevertheless, the difference in the statistical properties of point scoring could produce different results. For example, predictions that deviate from outcomes by one score will represent a smaller error as a proportion of the prediction for NBA games. Hence, inefficiencies in the NBA market may be smaller or less significant than in the NFL market.

The chapter is organized as follows: Section 2 briefly describes the football betting market and how point spreads are set; Section 3 describes the data and presents the test results. The results are summarized in the conclusion.

2. THE SPORTS BETTING MARKET: SETTING POINT SPREADS AND OVER/UNDERS

Although we will focus on the over/under market, we examine some of the related statistical properties of point spreads to see if information in that market can be used to

win over/under-bets. Like stocks and options, or spot commodities and futures, spread and over/under-bet pricing could be related. NFL and NBA betting markets are prediction markets like securities markets. Both spreads and over/unders are “line” bets, with winners and losers defined by whether the outcome ends up above or below the quoted line. For example, gamblers “invest” through a bookie (market-maker) at a point spread (price), which is the market’s prediction of the number of points by which the (stronger) favored team will outscore the (weaker) underdog team. Those who bet on the favorite speculate that their team is underpriced; that is, that the favorite will defeat the underdog by more than the point spread. Those who bet on the underdog expect their team to win the game or lose by less than the point spread.

Sports bookies also offer bets on the total points scored by the two teams. The bettor must predict whether the total number of points scored in a game will be over or under a total quoted by the bookie, that is, the over/under line. The over/under total depends on the two teams’ relative offensive and defensive strengths, and possibly, the playing surface, weather, and so on. Spread betting is more established and common, but over/under betting has grown in popularity. For example, in the 1994 season, spread bets were available for all NBA games, but over/under-bets were not offered on about 5% of them. By the year 2002, both were offered on all games.

Licensed Las Vegas sports books dominate the organized sports betting markets. Betting starts on games at the opening point spreads (the line), which typically reflect the expert opinions of a small group of professional spread forecasters. When new information on the relative strengths of opposing teams (e.g., a player injury) becomes available before game time, bookies often adjust the line. In addition, because gamblers’ identities are known, bookies sometimes change the line if professional gamblers place bets disproportionately on one side of the line. Of course, bets placed early at a specific line remain fixed regardless of future adjustments. The last line available shortly before game time when the bookie stops taking bets is called the closing line.

Efficient market theory implies that, like securities prices at the end of trading, closing lines unbiasedly aggregate all relevant information up to that point. For line bets, perfect efficiency implies that bets on either opponent in a game have a 50% chance of winning. However, if the marginal market participant is irrational in some way, closing lines may not properly reflect all relevant information, and bets on one of the opponents may have a better than 50% chance of winning.

In Las Vegas and other markets for large bettors, winners receive two dollars for each dollar bet; losers forfeit their bets plus an additional 10% of the bet. The 10% is the bookie’s commission or “vigorish.” Consequently, a profitable betting strategy must have at least a 52.4% win percentage. If the line perfectly predicts the outcome (a tie), all bets are usually canceled (a push).

Many researchers have assumed that bookies adjust the line to even out the betting on each game, essentially to hedge their positions in each game. But because the bookie manages a portfolio of mutually uncorrelated gambles, the risk can be diversified away. Therefore, assuming that lines are a good forecast of the outcomes and that each game’s bets represent a small component of a bookie’s portfolio, bookies can expect to earn a return approximately equal to the vigorish, regardless of how disproportionate the betting might be on any particular game.

One reasonable strategy for a bookie is to maximize the total dollar amount of bets placed subject to the constraint that the line is set to produce an even money gamble for each game. This is no different than how a casino treats even money bets in craps or roulette. They care more about the total amount bet over time as opposed to how much is bet on one side in any particular game. Of course, over many games, one would expect the amounts bet on either side of a gamble to be equal, although Levitt (2002) argues that bookmakers may intentionally offer biased lines if they expect to make enough vigorish on the additional amount wagered to offset any expected losses due to the bias.

3. NFL AND NBA BETTING MARKET EFFICIENCY

Our data is provided by *The Gold Sheet*, a well-known sports information publication that has been in business since 1957. The sample consists of all National Football League (NFL) games from 1993 through 2000 and NBA games from 1994 through 2002. The data consists of the point spreads and over/under lines, game dates, scores of opposing teams, and an overtime designation. There are a total of 2,008 NFL games and 10,819 NBA games during their respective sample periods.

Table 1 reports summary statistics for the point spread (PS) and the over/under (OU) lines for both the NFL and NBA. The PS and the margin of victory (MV) are defined relative to the favorite, hence, PS is the number of points by which the favorite is expected to win, and MV is the actual difference in score between the favorite and the underdog. Surprisingly, NFL and NBA average PS and MV do not differ by much, and the average PS error, $(MV - PS)$, is about minus one half point for both. That is, favorites tend to score about one half point less than predicted by the PS line. This implies that favorites could be less attractive bets than underdogs. Considering the medians, the median NFL error is still minus one half point, but it is zero for NBA games. This suggests that the favorite bias could be stronger for NFL games. The systematic favorite bias for both could support Levitt's (2002) profit-maximizing bookie model.

As one would expect, the average NFL OU and total points scored (TP) is much smaller than those of the NBA. The nature of the NFL game is that fewer total points are scored. The average NFL OU error $(TP - OU)$, however, is larger than that of the NBA. The average NFL error is about a point compared to a quarter point in the NBA. The positive average errors imply that betting the over could be profitable on average. As a proportion of the TP, the average NFL error is much larger than the average NBA error. This could reflect the lumpy NFL scoring increments when compared to NBA scoring increments. Although the standard deviation of the NFL errors is smaller than that of the NBA, this could reflect the effect of overtime. NFL overtime is sudden death, so only one score is required to end the game. The NBA has an overtime period, in which many points can be scored. Overtime could explain the average positive OU error. The zero median error, however, implies that OU bets are still even bets.

One feature of the PS and OU lines in both sports is that they exhibit skewness and kurtosis, with PS negatively skewed and OU positively skewed. PS also exhibits positive

TABLE 1 Summary Statistics for NFL and NBA Point Spread and Over/Under-Bets

Variable	Mean	Median	<i>SD</i>	Skewness	Kurtosis
Panel A: NFL statistics					
Point Spread (PS)	5.64	5.00	3.58	-0.98	0.75
Margin of victory (MV)	5.17	4.00	13.50	0.02	0.30
[MV - PS]	-0.47	-0.50	12.91	0.02	0.24
Over/under (OU)	40.15	40.00	4.13	0.72	1.20
Total points scored (TP)	41.22	41.00	14.22	0.35	-0.04
[TP - OU]	1.07	0.00	13.76	0.36	0.06
Panel B: NBA statistics					
Point Spread (PS)	6.37	6.00	3.72	-0.64	-0.10
Margin of victory (MV)	5.94	6.00	12.16	-0.01	0.43
[MV - PS]	-0.43	0.00	11.53	0.01	0.49
Over/under (OU)	193.82	193.00	11.145	0.16	-0.21
Total points scored (TP)	194.10	194.00	20.31	0.17	0.10
[TP - OU]	0.28	0.00	17.11	0.19	0.38

NOTE: For both NFL and NBA games, the data are defined relative to the favorite, that is in the NBA the favorite is expected to win by 6.37 points on average. NFL data cover 2,008 observations of point spreads and over/unders for the 1993–2000 seasons. The NBA data cover 10,815 point spreads and 10,819 over/unders for the 1994–2002 seasons.

kurtosis. Margins of victory are basically normally distributed, with little skewness or kurtosis. This is surprising because if the goal of the bookie is to set PS to mirror MV, one might expect them to have similar distributions. Of course, the bottom line for bookies is to end up with an even odds bet, which would be reflected in a zero median. Except for the NFL PS, the median errors are zero.

Market efficiency requires the closing OU line to be an unbiased measure of TP. That is, the closing OU should not be systematically higher or lower than the actual TP. We test market efficiency using the following ordinary least squares regression of TP on OU.

$$TP = \beta_1 + \beta_2(OU) + \varepsilon,$$

where TP is a vector of total points scored, OU is a vector of over/under lines, β_1 and β_2 are regression coefficients, and ε is an error term. The market efficiency test is an *F*-test of the joint hypothesis, $\beta_1 = 0$ and $\beta_2 = 1$.

Table 2, Panels A and B, present the regression tests for the NFL and the NBA, respectively, and for samples of all years and separate years. For the full sample periods, we find that the *F*-statistics reject OU market efficiency for both the NFL and the NBA. At the top of each panel, the full sample results show that the positive intercept estimates (β_1) are significantly larger than zero, and the slopes are smaller than one for both the NFL and NBA games.

TABLE 2 Regression Estimates for Tests of Market Efficiency for NFL and NBA Over/Under-Bets

Sample period	β_1	β_2	SER	R^2	F -statistic ($\beta_1 = \beta_2 = 0$)	F -statistic ($\beta_1 = 0, \beta_2 = 1$)	Obs.
Panel A: NFL regression tests							
1993–2000	5.96* (3.00)	0.88* (0.07)	13.75	0.065	139.34*	7.39*	2,008
1993	13.18 (9.57)	0.66* (0.25)	13.68	0.027	6.71*	1.14	242
1994	8.87 (8.66)	0.81* (0.21)	13.97	0.055	13.68*	1.53	236
1995	29.65* (11.02)	0.34 (0.27)	14.54	0.006	1.52	7.01*	252
1996	13.32 (11.77)	0.68* (0.29)	12.64	0.022	5.56*	0.85	252
1997	13.75 (11.90)	0.67* (0.29)	13.68	0.021	5.38*	0.78	252
1998	–3.85 (8.56)	1.13* (0.20)	13.16	0.105	29.66*	1.87	254
1999	6.41 (8.34)	0.87* (0.20)	14.08	0.065	18.08*	1.64	259
2000	–3.73 (6.31)	1.08* (0.15)	14.06	0.165	51.24*	0.18	261
Panel B: NBA regression tests							
1994–2002	8.73* (2.78)	0.96* (0.01)	17.10	0.291	4,435.41*	6.11*	10,819
1994	19.75* (9.08)	0.90* (0.25)	16.58	0.255	402.92*	2.38	1,178
1995	22.59* (9.15)	0.89* (0.04)	18.02	0.240	390.90*	3.34*	1,241
1996	19.91* (9.79)	0.90* (0.05)	16.96	0.222	333.07*	2.08	1,169
1997	1.33 (8.65)	1.00* (0.04)	17.15	0.288	496.97*	1.94	1,230
1998	–3.72 (10.72)	1.02* (0.06)	17.46	0.205	328.67*	0.06	1,280
1999	–5.70 (11.82)	1.03* (0.06)	16.75	0.238	252.85*	0.16	812
2000	14.95 (9.42)	0.92* (0.05)	16.75	0.223	362.85*	1.28	1,270
2001	18.49 (10.2)	0.91* (0.05)	17.35	0.173	276.36*	2.24	1,328
2002	2.24 (10.19)	0.99* (0.05)	16.68	0.208	343.15*	0.20	1,311

NOTE: This table presents regression tests for the over/under betting market for combined seasons and for individual seasons separately. The regression is $TP = \beta_1 + \beta_2(OU) + \varepsilon$, where TP is total points scored, OU is the over/under line, β_1 and β_2 are regression coefficients, and ε is an error term. The test of efficiency is an F -test of the joint hypothesis that $\beta_1 = 0$ and $\beta_2 = 1$. Standard errors in parentheses appear below the estimates. SER is the standard error of the regression. *denotes statistical significance at least at the 5% significance level.

Notice that the NFL regression R^2 is only about one quarter of the NBA R^2 . This implies that the NBA OU explains a larger proportion of the variation in the TP. This could reflect the fact that NFL point scoring is comparatively lumpy. A one-score NFL error is three or seven points, whereas it is three points or less in the NBA. Consequently, one would expect more noise in the NFL regressions.

Further evidence on the consistency of the bias can be found in the regressions for the data subset by year. Panel A shows that in six of the eight years, the NFL had positive intercept estimates, although only the 1995 estimate is statistically significant. Furthermore, 1995 is the only year where the F -statistic rejects market efficiency. During the last three years of the sample period, the NFL market appears to be relatively more efficient in the sense that the intercepts are closer to zero and the slopes are closer to one. The regression R^2 are also considerably larger. This means that the OU line explains a larger portion of the variation in TP.

In Panel B, the NBA regressions show seven of nine years with positive intercepts, with three being statistically significant. Like the NFL, the early years appear to drive the full sample results, particularly 1995. And again, like the NFL, 1995 is the only year where the F -statistic rejects market efficiency. Therefore, whatever was driving the OU market inefficiency could perhaps explain both the NFL and NBA bias because the bias appears mostly around 1995.

Table 1 results showed that the average total points scored exceeded the average over/under line, and Table 2 results reject efficiency for both the NFL and NBA markets. Hence, betting the over would have been more profitable than betting the under, particularly in 1995.

As mentioned above, overtime rules for NFL and NBA games differ, hence, our results could change if we account for the effect of overtime on TP. Surprisingly, about 5.4% of games end up in overtime for both the NFL and NBA. But with the NFL playing sudden death and the NBA playing extra periods, the effects of overtime on TP could differ substantially. If overtime can be predicted, then the effects should be included in the OU. To consider this possibility, we first examined whether the probability of overtime could be predicted from the PS or OU. PS could be related to overtime probability, because if team scores are expected to be close, one might expect games to end regular play in a tie. We ran probit regressions of overtime on the PS separately, OU separately, and then PS and OU together. The results showed no significant relations for NFL games. For NBA games, however, PS is negatively related and OU is positively related to the probability of overtime, beyond the 1% level of significance. This was true in separate regressions or when PS and OU appeared in the same regression. Therefore, overtime appears to be at least partially predictable in the NBA, but PS does not account for all of the predictability.

Overtime games tend to result in larger point totals. Table 3 reports market efficiency tests after adding an overtime dummy variable and the spread. For the entire NFL sample period, we find that overtime games increase TP by a statistically significant average of 3.73 points on average. This is reasonable because NFL sudden death overtime typically ends on a three-point field goal, which is easier to score than a touchdown. Indeed, the actual difference between TP and OU in overtime games is 3.75, which shows that

TABLE 3 Market Efficiency Tests for NFL and NBA Over/Under-Bets Adjusted for Overtime Games and Point Spread

Sample period	β_1	β_2	β_3	β_4	R^2	F-statistic ($\beta_1 = \beta_4 = 0, \beta_2 = 1$)	Obs.
Panel A: NFL regression tests							
1993–2000	5.64* (3.00)	0.87* (0.07)	3.73* (1.35)	0.06 (0.08)	0.069	3.52*	2,008
1993	11.66 (9.56)	0.73* (0.26)	8.69 (4.63)	-0.21 (0.24)	0.044	0.85	242
1994	10.40 (8.74)	0.71* (0.23)	-0.25 (3.63)	0.42 (0.26)	0.066	1.90	236
1995	27.15* (11.11)	0.41 (0.28)	5.95 (3.31)	-0.17 (0.26)	0.021	3.58*	252
1996	12.94 (11.79)	0.68* (0.29)	4.81 (3.48)	0.02 (0.24)	0.029	0.44	252
1997	12.96 (11.97)	0.67* (0.29)	-1.37 (3.55)	0.23 (0.26)	0.025	0.82	252
1998	0.53 (8.98)	0.97* (0.23)	2.10 (4.75)	0.40 (0.26)	0.115	1.95	254
1999	5.34 (8.31)	0.91* (0.21)	10.82* (4.53)	-0.12 (0.26)	0.086	0.66	259
2000	-4.13 (6.39)	1.08* (0.15)	1.40 (3.90)	0.09 (0.25)	0.166	0.18	261
Panel B: NBA regression tests							
1994–2002	9.32* (2.72)	0.94* (0.01)	16.92* (0.71)	0.07 (0.04)	0.327	10.04*	10,819
1994	19.66* (9.00)	0.89* (0.05)	9.68* (1.71)	0.25* (0.12)	0.277	3.74*	1,178
1995	24.62* (8.93)	0.97* (0.04)	17.74* (2.00)	0.16 (0.13)	0.286	3.61*	1,241
1996	16.28 (9.49)	0.91* (0.05)	18.65* (2.10)	0.09 (0.13)	0.272	2.76*	1,169
1997	0.60 (8.41)	1.00* (0.04)	17.24* (2.04)	-0.14 (0.24)	0.328	0.46	1,230
1998	-5.19 (10.72)	1.03* (0.06)	11.86* (3.21)	-0.05 (0.12)	0.213	0.23	1,280
1999	-3.14 (11.64)	1.02* (0.06)	16.90* (2.93)	-0.04 (0.17)	0.268	0.30	812
2000	11.03 (9.26)	0.94* (0.05)	18.88* (4.53)	0.05 (0.13)	0.268	1.57	1,270
2001	13.94 (9.82)	0.91* (0.05)	22.31* (1.92)	0.26 (0.14)	0.250	2.84*	1,328
2002	3.96 (9.88)	0.98* (0.05)	18.50* (1.96)	-0.02 (0.13)	0.258	0.94	1,311

NOTE: This table presents regression tests for the NFL and NBA over/under-betting market efficiency after controlling for the effects of overtime and point spread information for combined seasons and for individual seasons. The regression is $TP = \beta_1 + \beta_2(OU) + \beta_3(OT) + \beta_4(PS) + \varepsilon$, where TP is total points scored, OU is the over/under line, OT equals 1 for an overtime game and zero otherwise, PS is point spread, β_1 , β_2 , β_3 , and β_4 are regression coefficients, and ε is an error term. The test of efficiency is an F-test of the joint hypothesis, $\beta_1 = \beta_4 = 0$ and $\beta_2 = 1$. Standard errors appear in parentheses below the estimates. SER is the standard error of the regression. *denotes statistical significance at least at the 5% significance level.

the OU line explains none of the overtime effect on TP. Adding an overtime dummy has almost no effect on R^2 (compare Tables 2 and 3). But the intercept decreases, which means that the over bias is reduced after accounting for overtime effects. The F -test of market efficiency is still significant, but the level of significance falls. But this means that it could be important to account for overtime in other cases.

For the NBA, where the overtime is not sudden death, Panel B shows that overtime games increase the total points scored by a statistically significant 16.92 points on average. Furthermore, the actual difference between TP and OU for these games is 16.1 points. Therefore, the OU accounts for about 0.82 points of the overtime effect.¹ In this case, the intercept increases, which means that the over bias is greater after controlling for the predictability of NBA overtime. This is reinforced by the F -test, which is now more significant than it was when overtime was ignored in Table 2. It also reinforces the notion that OU predicts TP more precisely for the NBA than the NFL, perhaps because of the NFL's lumpy scoring.

Compared to Table 2, the overall results in Table 3 are little changed. Market efficiency is still rejected and the years around 1995 again drive the rejection. Adding the PS to the regression tests shows no significant relation between TP and PS.

Table 4 presents outcomes for betting strategies of both over and under. Panel A shows that betting the over for the full sample of NFL games is only marginally better than betting the under. Furthermore, the 50.1% winning percentage in the NFL is nowhere near the 52.4% required to cover the vigorish paid to the bookie. For the NBA in Panel B, the number of over wins is exactly equal to the number of under wins. These full sample results illustrate how the statistically significant biases found in Tables 2 and 3 do not necessarily translate into profitable betting strategies.

Of course, variations in efficiency in particular years could make a strategy profitable in some years. One should not be surprised to find one profitable strategy in the 17 years of combined NFL and NBA data, especially because they represent 34 trials as we are considering both sides of the bet. The regression tests imply that 1995 could be a profitable year to bet the over. Indeed, for the NFL this is true; the winning percentage of 54.3% significantly exceeds the 52.4% breakeven point. But the over strategy for NBA games in 1995 does not exceed breakeven. In 1997, the over looks profitable for the NBA, but the 52.5% winning percentage is not statistically larger than 52.4%.

Because there is no way of knowing beforehand that the over-bet would be profitable in the NFL in 1995, our results are of little value to a bettor. Even in 1993 through 1997 in the NFL, where the regression tests appear to show relatively large biases, betting the over would not have yielded a profit. Furthermore, in three of the five years, betting the under would have been as good as or better than betting the over.

The results in Table 3 show that information impounded in the PS cannot be used systematically to predict the TP after controlling for the OU and OT. Nevertheless, we considered two ways to use PS and OU in a non-linear fashion that might generate

¹The correlation between OU and OT for the NBA is a statistically significant 0.023 ($p < 0.02$). The same correlation for the NFL is insignificant. If we run the regressions in Table 3 after excluding OT, we find little change in the NFL intercept and OU estimates (5.64 vs. 5.98 and 0.87 vs. 0.87) but significant changes for the NBA estimates (9.32 vs. 8.67 and 0.94 vs. 0.96).

TABLE 4 NFL and NBA Over/Under Betting Strategies' Winning Percentages

Sample period	Betting strategy	Number of bets	Bets won	Ties	Winning percentages
Panel A: NFL bet results					
1993–2000	Over	2,008	988	36	0.501
	Under	2,008	984	36	0.499
1993	Over	242	119	4	0.500
	Under	242	119	4	0.500
1994	Over	236	115	4	0.496
	Under	236	117	4	0.504
1995	Over	252	133	7	0.543*
	Under	252	112	7	0.457
1996	Over	252	125	5	0.506
	Under	252	122	5	0.494
1997	Over	252	122	2	0.488
	Under	252	128	2	0.512
1998	Over	254	126	5	0.506
	Under	254	123	5	0.494
1999	Over	259	125	3	0.488
	Under	259	131	3	0.512
2000	Over	261	123	6	0.482
	Under	261	132	6	0.518
Panel B: NBA bet results					
1994–2002	Over	10,819	5,290	239	0.500
	Under	10,819	5,290	239	0.500
1994	Over	1,178	565	29	0.492
	Under	1,178	584	29	0.508
1995	Over	1,241	625	25	0.514
	Under	1,241	591	25	0.486
1996	Over	1,169	563	29	0.494
	Under	1,169	577	29	0.506
1997	Over	1,230	631	29	0.525
	Under	1,230	570	29	0.475
1998	Over	1,280	620	32	0.497
	Under	1,280	628	32	0.503
1999	Over	812	409	16	0.514
	Under	812	387	16	0.486
2000	Over	1,270	603	32	0.487
	Under	1,270	635	32	0.513
2001	Over	1,328	640	25	0.491
	Under	1,328	663	25	0.509
2002	Over	1,311	635	24	0.493
	Under	1,311	652	24	0.507

NOTE: The profitability of over/under betting strategies for NFL and NBA games for combined seasons and for individual years. *denotes a statistically significant winning percentage based on a binomial distribution. Winning percentages are calculated assuming that ties push. A push means that all bets are returned when the over/under betting line equals the total points scored in the corresponding game (a tie).

profitable bets. First, if PS is larger than average and OU is smaller than average, it might be profitable to bet the underdog. The idea is that if OU is small, the market is expecting fewer points to be scored, which could make it more difficult for a favorite to outscore an underdog by many points. Second, if PS is larger than average and OU is also larger than average, it might be profitable to bet on the favorite. In this case, because the market expects a large number of points to be scored, favorites might be more likely to outscore underdogs by a large number of points and beat a large spread.

Table 5 reports results for these strategies. Panel A starts by examining the NFL over the full sample without any filter. Overall, betting the underdog was nearly a profitable strategy, with a 52.1% winning percentage. But in Panel B, the NBA underdog was only marginally better than an even bet (50.5% winning probability). Next, we filtered by choosing only games where the PS was above average ($PS > 5.5$ for NFL and $PS > 6.5$ for NBA) and the OU was below average ($OU < 41$ for NFL and $OU < 195$ for NBA). Betting on the underdog was more profitable for the NFL (55.8%) but still only an even bet for the NBA (50.1%). The NFL bets were profitable in six of the eight sample years, with a winning percentage greater than 55% in four of those years.

For the second filtered bet, we chose games where the point spread was above average ($PS > 5.5$ for NFL and $PS > 6.5$ for NBA), but the over/under was above average ($OU > 40$ for NFL and $OU > 194$ for NBA). Our proposed strategy of betting on the favorite after this filter was actually worse than betting the underdog for both the NFL and NBA.

Table 5 also reports more extreme filters for the two filtered betting strategies. We selected games with PS and OU at least one standard deviation away from their means. For the NFL, the first strategy includes bets where $PS > 8$ and $OU < 38$. The more extreme filter increases the underdog winning percentage from 55.8% to 58.3%. For the second strategy, we selected games where $PS > 8$ and $OU > 43$. Betting the favorite becomes more profitable at 52.4%, but is still not significantly better than breakeven. For the NBA, the first strategy includes bets where $PS > 9$ and $OU > 185$. The more extreme filter increases the underdog winning percentage slightly from 50.1% to 50.7%. For the second strategy, we selected games where $PS > 9$ and $OU > 210$. The favorite winning percentage increases slightly from 46.9% to 47.6%, but the bet is still unprofitable. Overall, the restrictive filters increase the winning probabilities, but the improvement is only significant for the NFL bets.

4. CONCLUSION

We have studied the efficiency of the NFL and NBA betting markets with respect to over/under-bets. Regression tests of over/under-bets reject efficiency in both markets. Those tests imply that there is a bias that makes the over a better bet than the under on average. Furthermore, in both markets, the inefficiency is largest in the early years of the sample, with 1995 exhibiting the largest bias. Therefore, the NBA bias confirms the NFL bias, and vice versa. Because it appears in both data samples in the same years,

TABLE 5 NFL and NBA Favorite/Underdog Point Spread Betting Strategies Using the Over/Under Line

Sample period	Betting strategy	Number of bets	Bets won	Ties	Winning percentages
Panel A: NFL bet results					
1993–2000	Favorite	2,008	920	86	0.479
	Underdog	2,008	1,002	86	0.521
	PS > 5.5, OU < 41				
	Favorite	468	202	11	0.442
	Underdog	468	255	11	0.558*
	PS > 5.5, OU > 40				
	Favorite	438	205	15	0.485
	Underdog	438	218	15	0.515
	PS > 8, OU < 38				
	Favorite	99	40	3	0.417
	Underdog	99	56	3	0.583*
	PS > 8, OU > 43				
	Favorite	131	66	5	0.524
	Underdog	131	60	5	0.476
Panel B: NBA bet results					
1994–2002	Favorite	10,819	5,215	292	0.495
	Underdog	10,819	5,312	292	0.505
1994–2002	PS > 6.5, OU < 195				
	Favorite	2,109	1,033	38	0.499
	Underdog	2,109	1,038	38	0.501
1994–2002	PS > 6.5, OU > 194				
	Favorite	2,311	1,125	30	0.493
	Underdog	2,311	1,156	30	0.507
1994–2002	PS > 9, OU < 185				
	Favorite	325	151	3	0.469
	Underdog	325	171	3	0.531
1994–2002	PS > 9, OU > 210				
	Favorite	778	368	5	0.476
	Underdog	778	405	5	0.524

NOTE: The profitability of over/under betting strategies for NFL and NBA games for combined seasons and for individual years. *denotes a statistically significant winning percentage based on a binomial distribution. Winning percentages are calculated assuming that ties push. A push means that all bets are returned when the over/under betting line equals the total points scored in the corresponding game (a tie).

it likely reflects some systematic feature of the over/under market as opposed to an anomaly in the data for either NFL or NBA data.

Although the bias is statistically significant, we find that it is not economically significant. On average over the full samples, betting the over is not significantly profitable. Bets based on filter rules are sometimes profitable for the NFL sample, but not for the NBA sample, hence, profitability in one market is not confirmed in the other. This dichotomy could reflect the fact that the OU line tends to better predict the TP in the NBA market. For example, the average R^2 of the NBA regressions of TP on OU are four times that of the NFL. Furthermore, the NBA OU line accounts for some of the effects of overtime on TP, while the NFL OU line accounts for none of it. This could mean that the economic magnitude of the NFL bias exceeds that of the NBA bias. We suggest that these results could be due to the lumpiness in scoring in the NFL compared to the NBA.

References

- Brown, K. H., and F. J. Abraham. 2002. Testing Market Efficiency in the Major League Baseball Over-Under Betting Market, *Journal of Sports Economics* 3, 311–319.
- Golec, J., and M. Tamarkin. 2003. Efficiency of the Over/Under Betting Market for National Football League Games, in Vaughan-Williams (ed.), *The Economics of Gambling*. Routledge, London, pp. 135–144.
- Levitt, S. D. 2002. How Do Markets Function? An Empirical Analysis of Gambling on the National Football League. NBER Working Paper Series, Working Paper 9422, National Bureau of Economic Research, Cambridge, MA.
- Paul, R. J., and A. P. Weinbach. 2002. Market Efficiency and a Profitable Betting Rule: Evidence from Totals on Professional Football, *Journal of Sports Economics* 3, 256–263.

DOWNLOAD FULL VERSION HERE: <https://bettingbooks.org/12>

This page intentionally left blank

Download More Sports Betting Books at: www.bettingbooks.org

CHAPTER 13

Arbitrage and Risk Arbitrage in Team Jai Alai

Daniel Lane

Telfer School of Management, University of Ottawa, Ottawa, ON, Canada.

William T. Ziemba

Sauder School of Business, UBC, Vancouver, Canada, Mathematical Institute, Oxford University, and
ICMA Centre, University of Reading, UK.

1. Introduction	254
2. The Arbitrage	258
3. Risk Arbitrages	261
4. Final Remarks	269
<i>References</i>	271

Acknowledgments

This chapter is based on and updated from Lane and Ziemba (2004). This research was partially supported by the Natural Sciences and Engineering Research and the Social Science and Humanities Research Councils of Canada. Thanks are due to Ian Howard and friends for taking WTZ to the Mexico City fronton, in 1982, where the idea for this chapter was born, and to Donald Hausch and Jerry Kallberg for helpful discussions.

Abstract

We present arbitrage and risk arbitrage betting strategies for team jai alai. Most of the results generalize to other sports betting situations and some financial market applications. The arbitrage conditions are utility free. The risk arbitrage wagers use the Kelly expected log criterion.

Keywords: arbitrage, risk arbitrage, hedging, sequential investing

1. INTRODUCTION

This chapter discusses arbitrage and risk arbitrage strategies for betting on team jai alai.¹ The game originated in the Basque region of Spain and is played in Mexico City, Connecticut, Florida, Nevada, Rhode Island, and other locales (see Hollander and Schultz, 1978). It is played in a large enclosed rectangular court called a fronton between two opposing teams, each having two players. Players serve each point in turn and single points are scored by one team winning a rally off the serve, as in squash, racquetball, or tennis. Opposing teams alternate hurling and catching a ball (the pelota made of goatskin and hard rubber which must be recovered every 15 min of play) with an enlarged basket (cesta) against a wall (granite or concrete). When one team misses, the other team scores a point. The game is fast and exciting. Games are usually played to 30 points. At the fronton, bets may be placed on either team to win the game before every point is played at fixed locked-in odds, until the outcome of the game. Payoffs on bets made during the game are settled at the end of the game based on the quoted house odds at each betting point. We construct arbitrage and risk arbitrage bets with zero or little risk while at the same time yielding a positive return.

Arbitrage occurs in strategies when the net gain of all bets is always non-negative and sometimes positive and involves no risk of losing. Conditions that lead to arbitrage in various circumstances are studied in Kallio and Ziemba (2007). Risk arbitrages may yield losses, but occur more frequently and have higher mean returns. We develop these arbitrages for team jai alai. Section 2 provides conditions for arbitrage. Risk arbitrage is discussed in Section 3. Final remarks and applications to other areas are discussed in Section 4.

Assume that

1. The jai alai fronton bet payout rate is the constant $Q \in (1, 0)$.
2. The two teams' relative ability is known and defined by the probability of winning a single point—team A wins with the score invariant probability p , and B with $q = 1 - p$.

The probability of A reaching K points before B, given that A currently has $0 \leq m < K$ points and B has $0 \leq n < K$ points, is (according to Montmort, see

¹Goodfriend and Friedman (1975, 1977) and Skiena (1988) have analyzed the game of individual jai alai.

Epstein, 1977, p. 109):

$$\begin{aligned} P_a &= P_a(m, n) \\ &= p^{K-m} \left[1 + q(k-m) + q^2 \frac{(k-m)(K-m+1)}{2} \right. \\ &\quad \left. + \dots + q^{K-n-1} \frac{(2K-m-n-2)!}{(K-m-1)(K-n-1)!} \right] \\ &= p^{K-m} \left[\sum_{i=0}^{K-n-1} q^i \binom{K-m+i-1}{i} \right]. \end{aligned}$$

For $K = 30$, P_a is the probability that team A will win the game given the current score is m to n . A schedule of P_a and P_b for all values of m and n over the 30 point game in abbreviated form appears as Table 1 for the case of $p = 0.5$.

For given fixed Q , a schedule, as shown in Table 2, can be computed from the consistent house odds over the game since

$$\begin{aligned} P_a &= Q/(O_a + 1) \text{ and } P_b = Q/(O_b + 1) \text{ or} \\ O_a &= \frac{Q}{P_a} - 1 \text{ and } O_b = \frac{Q}{P_b} - 1 = \frac{Q - P_b}{P_b}. \end{aligned} \quad (1)$$

In Equation (1) O_a and O_b are the consistent house odds for teams A and B respectively, when the score is (m, n) . Odds of 1.5 to 1 means 1.5 profit plus the original bet or 2.5, is returned for each 1 bet, and so on. Consistent odds are those that return $1 - Q$ for the house's profit regardless of which team, A or B, wins. Then the expected return per dollar bet on each team is Q . Since $P_a + P_b = 1$ independent of Q , the odds then reflect the actual value of Q through Equation (1). Hence

$$O_a O_b = \frac{(Q - P_a)(Q - P_b)}{P_a P_b} = 1 - \frac{Q(1 - Q)}{P_a P_b}. \quad (2)$$

For given Q and odds on A of O_a to 1, consistent odds on B may fail to exist that guarantee the house advantage $1 - Q$. In that case, if there is a minimum payout of 1, that is, you just get your money back, then there is a *minus pool* and the house's actual take is less than $1 - Q$. A minus pool is defined to be this situation where the house take is $1 - Q^* < 1 - Q$, where $Q^* > Q$. So the effective Q , namely Q^* , is higher and the odds on B always exist since at their lowest the odds on O_b are the reciprocal of the odds on A. That is the case where the payback is all the money wagered and the house makes no profit since $Q^* = 1$. In general, these odds are possibly as low as the minimum guarantee, that is, $O_b = 1/O_a$. In a typical minus pool, the odds on B are higher than $1/O_a$. For example, if the minimum payout is 1.05, as is typical, then consistent odds on O_b may fail to exist. In this case, the odds on A are too large given the $1 - Q^*$ demanded by the minimum guarantee so that the odds on B do not exist. Let O_b be $Q = 1$ consistent odds (i.e., odds that give $Q = 1$) and O_{bh} are related by

TABLE 1 Probability that Team A Wins When the Score Is $A = m$ and $B = n$ and the Single Point Probability is $p = .5$

m/n	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29
0	.50	.55	.60	.66	.70	.75	.79	.83	.87	.90	.92	.94	.96	.97	.98	.99	.99													
1	.45	.50	.55	.61	.66	.71	.76	.80	.84	.87	.90	.93	.95	.96	.98	.98	.99	.99												
2	.40	.45	.50	.55	.61	.66	.71	.76	.80	.84	.88	.91	.93	.95	.97	.98	.99	.99												
3	.34	.39	.45	.50	.56	.61	.66	.72	.76	.81	.85	.88	.91	.94	.96	.97	.98	.99	.99											
4	.30	.34	.39	.44	.50	.56	.61	.67	.72	.77	.81	.85	.89	.92	.94	.96	.97	.98	.99	.99										
5	.25	.29	.34	.39	.44	.50	.56	.61	.67	.72	.77	.82	.86	.89	.92	.95	.96	.98	.99	.99										
6	.21	.24	.29	.34	.39	.44	.50	.56	.62	.67	.73	.78	.83	.87	.90	.93	.95	.97	.98	.99	.99									
7	.17	.20	.24	.28	.33	.39	.44	.50	.56	.62	.68	.73	.79	.83	.87	.91	.93	.96	.97	.98	.99	.99								
8	.13	.16	.20	.24	.28	.33	.38	.44	.50	.56	.62	.68	.74	.79	.84	.88	.91	.94	.96	.97	.99	.99								
9	.10	.13	.16	.19	.23	.28	.33	.38	.44	.50	.56	.63	.69	.74	.80	.84	.89	.92	.94	.96	.98	.99	.99							
10	.08	.10	.12	.15	.19	.23	.27	.32	.38	.44	.50	.56	.63	.69	.75	.80	.85	.89	.93	.95	.97	.98	.99							
11	.06	.07	.09	.12	.15	.18	.22	.27	.32	.37	.44	.50	.57	.63	.70	.76	.81	.86	.90	.93	.96	.97	.99	.99						
12	.04	.05	.07	.09	.11	.14	.17	.21	.26	.31	.37	.43	.50	.57	.64	.70	.76	.82	.87	.91	.94	.96	.98	.99	.99					
13	.03	.04	.05	.06	.08	.11	.13	.17	.21	.26	.31	.37	.43	.50	.57	.64	.71	.77	.83	.88	.92	.95	.97	.98	.99					
14	.02	.02	.03	.04	.06	.08	.10	.13	.16	.20	.25	.30	.36	.43	.50	.57	.64	.71	.78	.84	.89	.92	.95	.97	.99	.99				
15	.01	.02	.02	.03	.04	.05	.07	.09	.12	.16	.20	.24	.30	.36	.43	.50	.57	.65	.72	.79	.85	.89	.93	.96	.98	.99				
16	.01	.01	.01	.02	.03	.04	.05	.07	.09	.11	.15	.19	.24	.29	.36	.43	.50	.58	.65	.73	.80	.86	.91	.94	.97	.98	.99			
17		.01	.01	.02	.02	.03	.04	.06	.08	.11	.14	.18	.23	.29	.35	.42	.50	.58	.66	.74	.81	.87	.92	.95	.98	.99				
18			.01	.01	.01	.02	.03	.04	.06	.07	.10	.13	.17	.22	.28	.35	.42	.50	.58	.67	.75	.82	.88	.93	.96	.98	.99			
19				.01	.01	.01	.02	.03	.04	.05	.07	.09	.12	.16	.21	.27	.34	.42	.50	.59	.68	.76	.83	.89	.94	.97	.99			
20					.01	.01	.01	.02	.03	.04	.06	.08	.11	.15	.20	.26	.33	.41	.50	.59	.69	.77	.85	.91	.95	.98	.99			
21						.01	.01	.01	.02	.03	.04	.05	.08	.11	.14	.19	.25	.32	.41	.50	.60	.70	.79	.87	.93	.97	.99			
22							.01	.01	.01	.02	.03	.05	.07	.09	.13	.18	.24	.31	.40	.50	.60	.71	.81	.89	.95	.98				
23								.01	.02	.03	.04	.06	.08	.12	.17	.23	.30	.40	.50	.61	.73	.83	.91	.96	.99					
24									.01	.01	.01	.02	.03	.05	.07	.11	.15	.21	.29	.39	.50	.62	.75	.86	.94	.98				
25										.01	.01	.02	.02	.04	.06	.09	.13	.19	.27	.38	.50	.64	.77	.89	.97					
26											.01	.02	.03	.05	.07	.11	.17	.25	.36	.50	.66	.81	.94							
27												.01	.02	.03	.05	.09	.14	.23	.34	.50	.69	.88								
28													.01	.01	.02	.04	.06	.11	.19	.31	.50	.75								
29														.01	.02	.03	.06	.13	.25	.50										

NOTE: This table is symmetric in the sense that $\text{Prob}(A \text{ wins})$ with score m, n equals $\text{Prob}(B \text{ wins})$ with score n, m . This occurs if and only if $p = .5$.
 Source: Lane and Ziemba (2004).

TABLE 2 Consistent Odds for Teams A and B for Given Q

House odds, $O_a/1$	Payback rate, Q						
	1.00	0.950	0.900	0.850	0.800	0.750	0.700
0.100	10.000	9.025	8.100	7.225	6.400	5.625	4.900
0.111	9.000	8.123	7.290	6.503	5.760	5.063	4.410
0.125	8.000	7.220	6.480	5.780	5.120	4.500	3.920
0.143	7.000	6.318	5.670	5.057	4.480	3.938	3.430
0.167	6.000	5.415	4.860	4.335	3.840	3.375	2.940
0.200	5.000	4.512	4.050	3.613	3.200	2.813	2.450
0.250	4.000	3.610	3.240	2.890	2.560	2.250	1.960
0.333	3.000	2.707	2.430	2.167	1.920	1.688	1.470
0.500	2.000	1.805	1.620	1.445	1.280	1.125	0.980
1.000	1.000	0.903	0.810	0.723	0.640	0.563	0.490
1.500	0.667	0.602	0.540	0.482	0.427	0.375	0.327
2.000	0.500	0.451	0.405	0.361	0.320	0.281	0.245
2.500	0.400	0.361	0.324	0.289	0.256	0.225	0.196
3.000	0.333	0.301	0.270	0.241	0.213	0.188	0.163
3.500	0.286	0.258	0.231	0.206	0.183	0.161	0.140
4.000	0.250	0.226	0.202	0.181	0.160	0.141	0.123
4.500	0.222	0.201	0.180	0.161	0.142	0.125	0.109
5.000	0.200	0.180	0.162	0.144	0.128	0.113	0.098
5.500	0.182	0.164	0.147	0.131	0.116	0.102	0.089
6.000	0.167	0.150	0.135	0.120	0.107	0.094	0.082
6.500	0.154	0.139	0.125	0.111	0.098	0.087	0.075
7.000	0.143	0.129	0.116	0.103	0.091	0.080	0.070
7.500	0.133	0.120	0.108	0.096	0.085	0.075	0.065
8.000	0.125	0.113	0.101	0.090	0.080	0.070	0.061
8.500	0.118	0.106	0.095	0.085	0.075	0.066	0.058
9.000	0.111	0.100	0.090	0.080	0.071	0.063	0.054
9.500	0.105	0.095	0.085	0.076	0.067	0.059	0.052
10.000	0.100	0.090	0.081	0.072	0.064	0.056	0.049

Source: Lane and Ziemba (2004).

$O_{bh} = Q^*O_b - (1 - Q^*)$. Actual house odds, O_{ah} may differ from the consistent odds O_a and O_b for a number of reasons, including the desire of the oddsmakers to balance their books, competition among individual bookies for larger shares of the total pool, or additional information about teams' performance. This then adjusts the actual $1 - Q^*$ that the house receives.

For team A, or respectively B, the house odds are: $O_{ah} = O_a$ (consistent), $O_{ah} > O_a$ (favorable), and $O_{ah} < O_a$ (unfavorable).

2. THE ARBITRAGE

Arbitrage is the proverbial sure bet. For each betting point of a K point game with a utility function U and betting wealth W , optimal arbitrage bets can be found by solving

$$\max_{B_a > 0, B_b > 0} E[U(B_a, B_b)] \quad (3)$$

$$\begin{aligned} \text{s.t.} \quad & B_a + B_b \leq W \\ & B_a O_{ah} + B_a \geq B_a + B_b \\ & B_b O_{bh} + B_b \geq B_a + B_b \end{aligned}$$

where B_a and B_b are the amounts bet on A and B, respectively, and E is the expectation operator. Besides the budget constraint, the arbitrage constraints indicate that the return if either A or B wins is never less than the total bet on A and B. This reduces to the arbitrage betting condition at every point of the game. These constraints yield

$$1/O_{ah} \leq B_a/B_b \leq O_{bh}, \quad B_b \neq 0, \quad (4)$$

which demonstrates:

Theorem 1. The arbitrage exists if

$$O_{ah} O_{bh} \geq 1. \quad (5)$$

The arbitrage condition in Equation (5) is utility free and holds for all U . Both B_a and B_b must be positive or both zero.

The constraints in Equation (3) imply that

$$B_a/B_b \leq W/B_b - 1.$$

By Equation (4),

$$1/O_{ah} \leq W/B_b - 1.$$

Hence

$$B_b \leq O_{ah} W / (1 + O_{ah}).$$

Similarly,

$$B_a \leq W / (1 + O_{ah}).$$

We consider betting strategies using the strategy variable $f \geq 1$, using

$$B_a = fW/(1 + O_{ah}) \text{ and } B_b = W(1 - f + O_{ah})/(1 + O_{ah}),$$

when

$$O_{ah}O_{bh} = 1, B_a = W/(1 + O_{ah}), B_b = O_{ah}W/(1 + O_{ah}) \text{ and } f = 1.$$

To guarantee the arbitrage, the profit must be non-negative regardless of which team wins:

$$\begin{aligned} \text{profit if A wins} &= W(O_{ah}f - (1 - f) - O_{ah})/(1 + O_{ah}) \\ &= W(f - 1) \geq 0, \end{aligned} \quad (6)$$

$$\begin{aligned} \text{profit if B wins} &= W(O_{bh}(1 - f) + O_{bh}O_{ah} - f)/(1 + O_{ah}) \\ &= W(O_{bh}(1 + O_{ah}) - f(1 + O_{bh}))/ (1 + O_{ah}) > 0. \end{aligned} \quad (7)$$

To satisfy Equations (5–7)

$$f_{\min} \equiv 1 \leq f \leq O_{bh}(1 + O_{ah})/(1 + O_{bh}) \equiv f_{\max}. \quad (8)$$

Figure 1 illustrates how the net payoffs vary with f . For

$$f^* = (O_{ah} + 1)(O_{bh} + 1)/(1 + O_{ah} + O_{bh} + 2), \quad (9)$$

one maximizes the minimum arbitrage profit

$$W(O_{ah}O_{bh} - 1)/(O_{ah} + O_{bh} + 2).$$

More insight about the arbitrage betting condition in Equation (5) may be obtained by comparing it to consistent odds at each betting point. They require that

$$O_aO_b = Q^2. \quad (10)$$

The house odds favorability factors $r_a > 0$ and $r_b > 0$ for favorable odds are defined by

$$O_{ah} \equiv O_a(1 + r_a) \text{ and } O_{bh} \equiv O_b(1 + r_b). \quad (11)$$

Then Equation (5) implies that

$$Q^2(1 + r_a)(1 + r_b) > 1. \quad (12)$$

Note that $r_A > 0$ and $r_B > 0$, although typical, is not required for Equation (12) see also Figure 2.

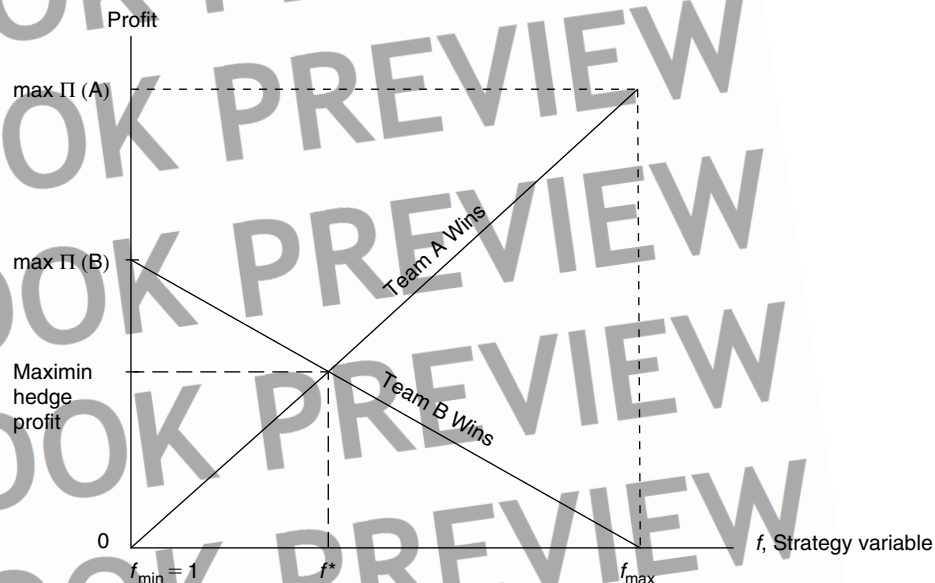


FIGURE 1 Profits for Teams A and B for initial betting wealth of 1 unit vs. strategy variable f , when $O_{ah} > O_{bh}$ and $O_{ah}O_{bh} > 1$. Source: Lane and Ziemba (2004).

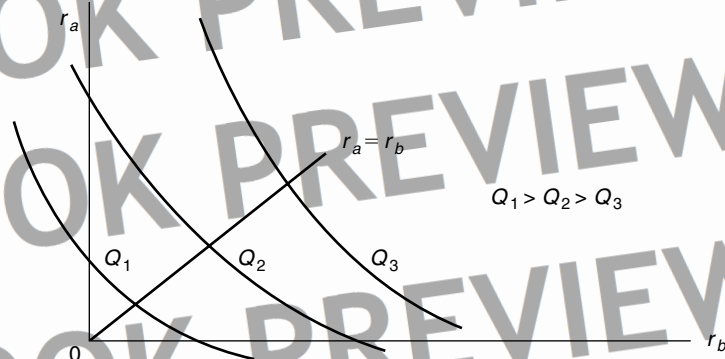


FIGURE 2 House odds favorability regions. Source: Lane and Ziemba (2004).

If $r_a = r_b = r$, the schedule of Q versus r and the region of betting under the perfect hedge is as shown in Figure 3.

For typical values for Q of about 0.85 the required favorability for house odds quoted on both teams is nearly 20%. Such discrepancies occur occasionally in actual betting. More frequently, however, one needs to take added risk to get good bets so we now turn to the construction of risk arbitrages.

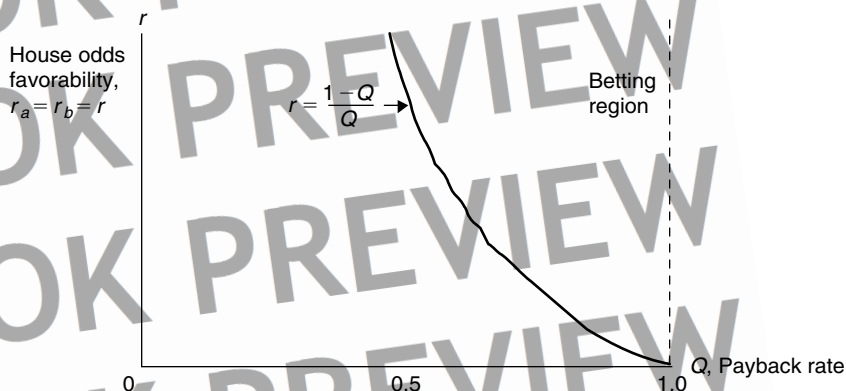


FIGURE 3 Betting region of Q vs. r . Source: Lane and Ziemba (2004).

3. RISK ARBITRAGES

Two approaches are considered for constructing risk arbitrage positions. They exploit the observed house odds favorability conditions to find *good* arbitrage strategies.

A maximal capital growth model encompassing these approaches is

$$\begin{aligned} & \max_{B_a(i) \geq 0, B_b(i) \geq 0} E_g[(\ln W(g))] \\ \text{s.t.} \quad & \sum_{i \in g} (B_a(i) + B_b(i)) \leq W_0 \\ & \sum_{i \in g} B_a(i) O_{ah}(i) \geq \alpha \sum_{i \in g} B_b(i) \\ & \sum_{i \in g} B_b(i) O_{bh}(i) \geq \alpha \sum_{i \in g} B_a(i), \forall g \in G \end{aligned} \quad (13)$$

where E_g represents mathematical expectation with respect to the game path g , G is the set of scenario game paths from $(0, 0)$ to the final outcome, $W(g)$ is the wealth with g , and the constant $\alpha > 0$ is the relative degree of risk of the bettor in the arbitrage. For example, if $\alpha = 3/4$, the bettor requires that total returns must cover at least 75% of total bets in any game. With $\alpha = 1$, the goal is to find an arbitrage to cover all bets and a premium is required for betting in any game if $\alpha > 1$. The log utility function corresponds to the Kelly (1956) system of betting which maximizes the asymptotic long run rate of growth of the bettor's fortune. See MacLean and Ziemba (2006), Thorp (2006), and Ziemba and Ziemba (2007), for a summary of results concerning such betting strategies.

The first approach to the risk arbitrage problem is a model that analyzes the objective function over all feasible paths of the game and computes the bets B_a, B_b which

TABLE 3 Cardinality of G is Large

Points in the game	Total betting points	Game betting points		Number of game paths
		Min	Max	
1	1	1	1	2
2	4	2	3	6
3	9	3	5	20
4	16	4	7	70
5	25	5	9	252
..
..
..
30	900	30	59	5.91×10^{16}

maximize Equation (13). This presupposes that information is known in advance about the odds O_{ah} and O_{bh} , actually set by the house throughout the game. This information may be a probability distribution or a function for O_{ah} and O_{bh} , over the scores of the game. The drawback is that the cardinality of G is very large as shown in Table 3. Calculations for two and three point games using $a = 1$, $Q = 0.85$, and $p = 0.5$ appear in Table 4. These calculations utilize the following probability distributions for the house odds favorability factors $r_a = r_b = r$:

- 1. $r \sim N(0, \sigma^2)$ is iid for all scores of the game.
- 2. $r \sim N[r(i), \sigma^2]$, depends upon the size of the lead and the number of points the leader is away from winning the game.

This relation is given by

$$r(i) = \begin{cases} \exp(-M/D) - 1 & \text{if team } i \text{ is leading by } M \text{ points} \\ \exp(pM/D) - 1 & \text{if team } i \text{ is trailing by } M \text{ points,} \end{cases}$$

D is the number of points the leader is away from winning the game, namely $K - S$, where K are the points needed to win the game and S is the score of the leader.

The results yield insights that may be useful in the construction of good heuristic strategies for 30 point games. For the normally distributed favorability factors, there is an intrinsic threshold value for the odds favorability below which no initial bet is placed. This threshold is about a 20% odds favorability. Secondly, betting almost always takes place in natural-arbitrage pairs, where a bet on one team at one point of a particular game path is paired with a compensating bet on the other team later in the game along the same game path. Where perfect arbitrages could be constructed, these dominated all other possible bet points. As variance increases, the number of lucrative bets also increases both in number and in size of bet.

TABLE 4 Sample Results for Two and Three Point Games

House odds type	W	σ	Games bet per games played	Average percent of bet compared to game bet	Average value of funds per game played
Two point games					
(1)	1	0	0/1	0	0
	1	.1	1/10	100	0.0005
	1	.2	5/10	84	0.0073
(2)	1	0	0/1	0	0
	1	.1	0/10	0	0
	1	.2	2/10	100	0.0119
Three point games					
(1)	1	0	0/1	0	0
	1	.1	0/10	0	0
	1	.2	7/10	80	0.0193
(2)	1	0	0/1	0	0
	1	.1	3/10	75	0.0098
	1	.2	5/10	60	0.0059

Source: Lane and Ziemba (2004).

Many betting points occurred in pairs of tied scores and trailing team bets, where the higher odds for the trailing team boost the combined bet pair over the arbitrage condition requirements. Under these situations, gains could be realized on a team that came from behind to win, and strategies can concentrate on this possible event occurring while hedging a priori that the leading team wins.

In the exponential function distribution, the high relative favorability of the trailing team contributes to still greater emphasis on betting on the trailer to come from behind to win while hedging (usually early in the game, or at a tied point) on the leader to win. As in the normal case, the initial bet favorability threshold value of 20% continues to manifest itself in the results.

Additional calculations showed modest expected total gains in the range of 1–2% using this method for sets of 25 games. The gains in the three point game are larger than the two point game for the same level of variance, which suggests that higher gains could be anticipated as the game size increases and more betting opportunities arise.

We now utilize these insights in a second approach to risk arbitrage by constructing single arbitrage bets for the 30 point game. A single arbitrage is a bet on one team and subsequently a bet on the opposing team later on in the game such that the constraints of Equation (13) hold. Unlike arbitrage, which required betting both teams at the same time, this risk arbitrage does not require concurrent bets. The idea is to exploit the favorability of quoted odds on one team at some point in the game and take the risk that the house odds will become attractive enough on the opposing team later on so that an arbitrage may be constructed. However, the second half of this bet will not

always materialize. Volatility and prediction models are useful here. This hedge may be formulated by examining the constraints on the bets B and

$$\begin{aligned} B_a(S_a)O_{ah}(S_a) &\geq \alpha B_b(S_a), \quad \alpha > 0 \\ B_b(S_b)O_{bh}(S_b) &\geq \alpha B_a(S_b), \end{aligned} \quad (14)$$

where S_a and S_b are the scores of the game at the time a bet is made on team A and B, respectively.

The approach taken in the construction of the single hedge is to simulate the passage of the 30 point game using the same assumptions about odds favorability as with the two and three point game model.

The simulation model of the 30 point game generates at each point an odds favorability for each team, and the winner of the point using a uniform distribution on p . Bets were generated by first initializing a threshold favorability value which is used as the basis for placing the initial bet. Once the initial bet is placed whose amount is determined by criteria discussed below, the game continues until a new betting point for the opposing team is found such that the arbitrage condition in Equation (5) is satisfied. At the end of the game the bets are settled and the results recorded. Each simulation run is a set of 30 point games.

To find better strategies, and study the sensitivity of the model, different sets of games were simulated and compared in order to find the threshold favorability values which gave consistent positive expected net gains over the entire set of games. The results are highly sensitive to the value of this parameter. Low threshold values typically mean early betting points and more likely completion of the single hedge. However, early bets typically mean low odds and thus a small hedge margin (the amount by which $O_{ah}O_{bh}$ actually exceeds 1) and consequently smaller net gains. Higher values mean delayed betting points with a greater chance of not completing the second half of the hedge. However, potential loss due to unhedged or single bet games is compensated for by higher hedge margins (due to later game scores, larger point spreads and higher odds) and hence more potential net gains when the risk arbitrage is successfully completed. Typical results for these single arbitrage betting pairs appear in Table 5. In two games, the risk arbitrage was completed. One leads to a positive gain, the other to a break even situation. In the other game, the risk arbitrage was not completed and leads to a loss. If a risk arbitrage is not completed, it invariably leads to a loss because the team that you would like to bet on to complete the quasi-hedge remains ahead throughout the game.

Figure 4 gives a profile of three different sets of games under various initial bet threshold favorability values for the normally distributed case.

In the cases illustrated and described above, the implicit assumption in the construction of the hedge and the determination of the amounts bet was that the arbitrage condition in Equation (5) is satisfied as an equality, that is, $O_{ah}O_{bh} = 1$. Thus the arbitrage margin was assumed to be zero in the construction of the arbitrage bet pair. However, if a margin, $m > 0$ exists, such that

$$O_{ah}O_{bh} = 1 + m$$

TABLE 5 Sample Results for Single Risk Arbitrage Betting Pairs

Game number	Bet	Team	Amount bet	Score		Actual odds	House odds	Favorability
				A	B			
10	1.	B	0.8581	18	23	0.1288	0.1654	0.2838
	2.	A	0.1419			6.2963	7.6719	0.2185
		Final score	Bets	Payoff				
	Team A	31	0.1419	1.0886				
	Team B	29	0.8581	−0.8531				
Net betting payoff = 0.2304								
11*	1.	A	0.0306	14	23	25.7480	31.6751	0.2302
		Final score	Bets	Payoff				
	Team A	15	0.0306	−0.0306				
	Team B	31	0.0	0.0				
	Net betting payoff = −0.0306							
21	1.	B	0.6493	0	3	0.4466	0.5401	0.2094
	2.	A	0.3507	1	5	2.0653	1.9782	−0.0422
		Final score	Bets	Payoff				
	Team A	17	0.3507	−0.3507				
	Team B	30	0.6493	0.3507				
Net betting payoff = 0.0000								

*Arbitrage condition not realized, only one bet placed. *Source:* Lane and Ziemba (2004).

then the effect is to permit bets to place more emphasis on one team or the other or neither within the limits of satisfying the constraint set of Equation (13) (see Figure 1 for the arbitrage case). The perceived existence of a risk arbitrage margin allows the bettor to express the bet as a function of the score at the time of the bet placement, or as merely a prediction about the eventual outcome of the game. The actual margin cannot be known until the arbitrage condition is satisfied at the second betting point. However, because the value of the margin affects the initial bet, it must be anticipated by some estimate.

If a margin exists, then the arbitrage, if it has been successfully constructed, satisfies all the conditions required for a member of the infinite set of betting pairs defined by

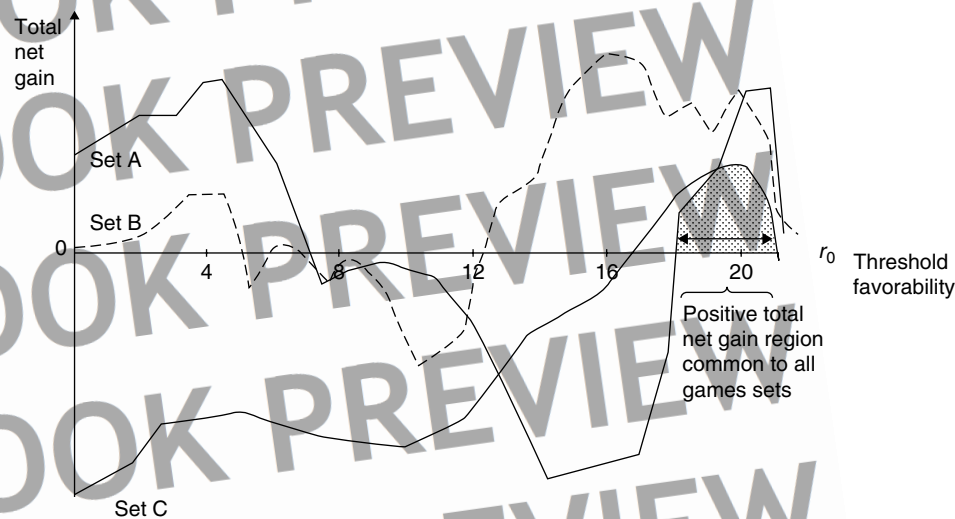


FIGURE 4 Thirty point game, single risk arbitrage simulation r_0 search. *Source:* Lane and Ziemba (2004).

TABLE 6 Total Gain Values for a 100 Game Set

m/r_0	f_{\min}				f^*				f_{\max}			
	0	0.05	0.10	0.20	0	0.05	0.10	0.20	0	0.05	0.10	0.20
0	1.41	0.47	-0.22	2.22	1.41	0.47	-0.22	2.22	1.41	0.47	-0.22	2.22
0.05	2.22	1.09	-0.02	3.35	2.01	0.93	-0.05	3.50	1.92	0.86	-0.10	3.89
0.10	2.67	2.43	-0.02	3.18	2.32	2.25	-0.08	3.62	2.07	1.92	-0.15	4.27
0.20	1.19	0.08	-1.47	3.09	0.78	-0.03	-1.50	4.09	0.17	-0.68	-1.57	5.17

NOTE: Exponential function favorability, $\sigma = 0.2$, constant strategy. *Source:* Lane and Ziemba (2004).

the single betting strategy variable, f of Figure 1. However, if the margin should not materialize, then the risk is that the risk arbitrage will not be completed and the game may end with an unpaired bet that may be lost. If on the other hand the anticipated margin understates the actual margin, then there is an opportunity loss due to the wrongly specified betting split. Hence even though the bet is not lost, we could have done better had we known the exact margin value.

This model was used in this second way to search for best values for the anticipated margin factor, m and the threshold favorability factor, r_0 values for the different sets of games described above. In the first instance, a constant betting strategy was employed independent of the score at the time of the initial bet. The strategy variable used here corresponds to the upper and lower limits of f , as well as f^* .

Typical results for various m values are in Table 6.

When the anticipated margin is zero, the results of the three strategies are identical since the zero margin assumption uniquely defines the bets for each team (given by $f = 1$).

For some (m, r_0) pairs, the f_{\min} strategy yields better results while for other pairs f_{\max} is superior. The f^* results always take intermediate positions between f_{\min} and f_{\max} for all (m, r) . The results for all three strategies however are not significantly different for any (m, r_0) . This performance is not unexpected and for larger game sets it is expected that there would be less difference between the results of these three strategies.

As m increases for constant r_0 , the total gain rises and then falls off as the second half of the bet pair is more difficult to complete. Low r_0 returns are, in general, positive with low variance since bets are placed early and the pairs are completed early, generally with low odds. High r_0 values yield better mean returns with larger variances. Intermediate r_0 values reduce the returns. The results of this strategy for the same set of games used in Table 6 appear in Table 7.

Comparing the automatic versus constant variable strategies shows that the automatic strategy dampens the extreme results of the f_{\min} and f_{\max} strategies of Table 6 while improving on the more conservative results of f^* . While these results may not be considered as being significantly different, the trend of the automatic strategy is toward a more stable and profitable outcome. More importantly, the variance of the expected gains is reduced by about a third from six to four over the comparable results in Table 6. Finally, the automatic strategy, being score dependent, is more intuitive and appealing and is the preferred policy for the single hedge construction problem in the 30 point game.

While single risk arbitrage jai alai results are encouraging, the gains that occur for the simulated game sets are modest. There is greater potential for larger gains in longer games as there is more opportunity for inefficiency to manifest itself through volatility.

Combining the results of the two risk arbitrages, our final analysis examines the policy of betting over the 30 point game through the construction of a series of arbitrage bet pairs.

The simulation model discussed above was modified to accommodate a series of single arbitrage bets. The same amount was assumed available for each arbitrage bet pair with a maximum total bet availability requirement of 60 betting units.

TABLE 7 Single Hedge Construction; Total Gain Over 100 Games Sets Using an Automatic Strategy Variable

m	Threshold favorability, r_0			
	0	0.05	0.10	0.20
0	1.41	0.47	-0.22	2.22
0.05	2.11	1.03	-0.02	3.55
0.10	2.49	2.33	0.02	3.61
0.20	0.94	0.02	-1.24	3.12

Source: Lane and Ziemba (2004).

TABLE 8 Betting Summary: Multiple Quasi-Hedge Bets

Bet	Team	Amount bet	Score		Actual odds	House odds	Favorability
			A	B			
1	A	0.5459	1	0	0.6894	0.8733	0.2668
1	B	0.4541	1	0	1.048	1.2047	0.1495
2	A	0.523	1	1	0.85	0.9356	0.1007
2	B	0.477	4	2	1.3188	1.4157	0.0735
3	A	0.535	2	1	0.6869	0.9127	0.3288
3	B	0.465	2	1	1.0519	1.16	0.1028
4	A	0.6342	4	2	0.5479	0.6057	0.1055
4	B	0.3658	9	6	1.759	1.7204	-0.0219
5	B	0.3139	10	7	1.7899	2.1853	0.2209
5	A	0.6861	10	9	0.6606	0.7306	0.1059
6	B	0.3855	11	9	1.4189	1.5938	0.1233
6	A	0.6145	12	11	0.6517	0.6392	-0.0192
7	B	0.3167	12	10	1.4385	2.1578	0.5
7	A	0.6833	12	10	0.5022	0.5551	0.1052
8	A	0.4782	12	12	0.85	1.1168	0.3139
8	B	0.5218	25	24	1.4049	1.0277	-0.2685
9	A	0.2919	19	21	1.7751	2.4262	0.3668
9	B	0.7081	19	21	0.407	0.4617	0.1344
10	A	0.2301	19	22	2.6866	3.3467	0.2457
10	B	0.7699	21	23	0.3706	0.338	-0.088
11	A	0.4152	22	23	1.3005	1.4085	0.0831
11	B	0.5848	23	23	0.85	0.7379	-0.1319
12	A	0.4123	23	24	1.3452	1.4254	0.0596
12	B	0.5877	25	25	0.85	0.8935	0.0512
13	B	0.4613	24	24	0.85	1.1942	0.4049
13	A	0.5387	24	24	0.85	0.9764	0.1487
14	A	0.4969	25	25	0.85	1.0372	0.2203
14	B	0.5031	26	25	1.4898	1.5992	0.0734
15	B	0.5303	26	26	0.85	0.9089	0.0693
15	A	0.4697	28	29	2.55	3.3276	0.3049
16	B	0.508	29	29	0.85	0.9926	0.1678
			Final score	Bets	Payoff		
			Team A	29	7.5549	-7.5549	
			Team B	30	7.9532	8.7968	

NOTE: Net betting payoff = 1.2419. Source: Lane and Ziemba (2004).

TABLE 9 Exponential Function Favorability,
 $\sigma = 0.3$; 100 Game Set Hedge Series Construction

m	Threshold favorability, r_0			
	0	0.1	0.2	
0	-1.96	2.12	11.29	17.21
0.05	-0.76	4.51	13.27	19.12
0.1	-0.79	3.77	10.34	17.26
0.2	2.83	4.74	8.19	14.17

Source: Lane and Ziemba (2004).

The algorithm proceeds by requiring that the preset threshold favorability factor value is satisfied before the initial bet of any pair is made. Priority over initializing the first half of a new risk arbitrage is given to matching unmatched hedges. The automatic strategy policy is used for the construction of all pairs.

Table 8 describes the bets in a typical game where 24% of a betting unit is the total profit. A summary of the results of this simulation appears in Table 9. The results are encouraging. Depending on the anticipated margin and the threshold value chosen, the number of hedges completed in the simulated games can range from 1 to 20 with never more than three uncompleted bet pairs among this series. Losses may be incurred on any particular game under this imperfect hedge strategy due to the possibility of uncompleted bet pairs. Such losses never exceeded one betting unit for any of the simulated games, whereas the single game gain ranged as high as three units. Game sets are divided approximately 60/40 in terms of winning to losing games for the sets examined. The total gain over the entire set of games is increased relative to the single hedge strategy and the instance of loss is reduced. As the odds favorability variance increases, the potential for more profitable bets occurs and the expected gains and variance rise accordingly, positive net gains occur regularly for similar values of the standard deviation of the odds favorability distribution. The variance of the expected gains are larger than for the single hedge case (40 versus 6) as expected.

4. FINAL REMARKS

It is possible to construct profitable arbitrage strategies for the 30 point jai alai game. Modest returns may be realized under strategies of arbitrage and risk arbitrage single bet pair constructions. Mathematical programming results imply that series of bet pairs may be optimal for games of this kind. Simulation results suggest that improved gains may be obtained under such strategies where bet placements are dependent on the favorability of quoted odds and the score. Further analysis of this situation might concentrate on more detailed investigation into the actual distribution of quoted house odds during the game. This will involve more intensive data collection at jai alai frontons.

Data collection in Mexico City indicates that there are substantial inefficiencies. The formulation here has assumed score invariant single point win probabilities. A more refined but possibly unmanageable analysis might consider score variant probabilities possibly using Markov chains. More study should also be undertaken with respect to the reaction of the oddsmakers to shifts in score and an examination of their individual objective functions. Parayre (1986) has some results along these lines for win and perfecta bets based on player strength and post positions. These results follow ideas in Ziemba and Hausch (1986).

The methodology and insights found for team jai alai also have potential applications in other situations where one has non-marketable financial instruments. These include certain horse racing (especially on betting exchanges), currency exposure, and production situations. Risk arbitrage in traded options and warrants markets is an additional example. See Shaw et al. (1995) for one such application related to the Japanese Nikkei put warrant in 1989–1990.

In England and other European and Commonwealth countries, legalized bookies set odds that various horses will win a given race both on-course and off-course. These odds may differ across bookies at a particular moment in time. The odds change during the 20 or so minutes before a race is run as opinions are altered in light of new information such as the horses' appearances and because the bookies would like to simultaneously balance their books to guarantee a profit no matter what horse wins, and maximize the number of tickets sold. This situation, from the bettor's perspective, mirrors the jai alai situation, once extended to multiple outcomes, assuming that he has an independent estimate of the probability that each horse will win obtained by a handicapping or statistical procedure.

A key feature of the team jai alai and racing situations is that the tickets once purchased are not marketable except perhaps at a substantial discount. Other situations share these features and we will describe two of them briefly here.

Consider a company with substantial foreign accounts receivable at a future date. The standard way to hedge against possible devaluations is through a futures contract in the country's currency. However, in many cases this is not possible because the currency does not have an active futures market or the time horizon is too long. The currencies of Italy, Thailand, and Turkey are examples of the former. Even for established heavily traded currencies such as the Euro and the Mexican Peso, such contracts will not cover a multiple year exposure. Negotiations with a bank might produce a special forward contract for part of the exposure. Such a contract would be difficult to sell except at a substantial discount. As time goes on, the company may add additional contracts to cover more of the exposure with the original or other banks. In terms of the jai alai formulation, one may think of the original exposure and any subsequent accounts receivable as bets on A and the covering as bets on B.

Farmers often have fixed contracts for delivery of the crops from their acreage at a specified time. Both the price he or she will receive and quantity he or she will have available are likely uncertain. In a publicly traded commodity such as corn or wheat he or she could hedge against these uncertainties. However, active futures markets are not available for most commodities. Lettuce and raspberries are two such examples.

The farmer can consider his or her crop as bets on A and contracts he or she makes with other farmers of specific quantities at fixed prices as bets on B.

Some analyses of problems similar to these two examples using hedging arguments for static problems appear in Anderson and Danthine (1981), Feiger and Jacquillat (1979), McKinnon (1967), and Rolfo (1980), and for a two period problem in which additional information becomes available, see Baesel and Grant (1982).

References

- Anderson, R. A., and J. P. Danthine. 1981. Cross Hedging, *Journal of Political Economy* 89, 1182–1196.
- Baesel, J., and D. Grant. 1982. Optimal Sequential Futures Trading, *Journal of Financial and Quantitative Analysis* 17, 683–695.
- Epstein, R. A. 1977. *The Theory of Gambling and Statistical Logic*, 2nd ed. Academic Press, San Diego, CA.
- Feiger, G., and B. Jacquillat. 1979. Currency Option Bonds, Puts and Calls on Spot Exchange and the Hedging of Contingent Foreign Earnings, *Journal of Finance* 34, 1124–1140.
- Goodfriend, M. J., and J. H. Friedman. 1975. A Monte-Carlo analysis of Jai Alai, Paper presented at the ORSA/TIMS Joint National Meeting.
- Goodfriend, M. J., and J. H. Friedman. 1977. Some Further Analyses of Jai Alai, Paper presented at the ORSA/TIMS Joint National Meeting.
- Hollander, Z., and D. Schultz. 1978. *The Jai Alai Handbook*. Pinnacle Books, Los Angeles.
- Kallio, M., and W. T. Ziemba. 2007. Arbitrage Pricing Simplified, *Journal of Banking and Finance* 31, 2281–2302.
- Kelly, J. 1956. A New Interpretation of Information Rate, *The Bell System Technical Journal* 35, 917–926.
- Lane, D., and W. T. Ziemba. 2004. Jai Alai Arbitrage Strategies, *European Journal of Finance* 10, 353–369.
- MacLean, L. C., and W. T. Ziemba. 2006. Capital Growth Theory and Practice, in S. A. Zenios and W. T. Ziemba (eds.), *Handbook of Asset and Liability Management, Vol. 1: Theory and Methodology*, North-Holland, Amsterdam, pp. 429–473.
- McKinnon, R. I. 1967. Futures Markets, Buffer Stocks and Income Stability for Primary Producers, *Journal of Political Economy* 75, 844–861.
- Parayre, R. 1986. A Study of Jai-Alai Betting. Term paper in UBC Commerce 612.
- Rolfo, J. 1980. Optimal Hedging Under Price and Quantity Uncertainty: The Case of a Cocoa Producer, *Journal of Political Economy* 88, 100–116.
- Shaw, J., E. O. Thorp, and W. T. Ziemba. 1995. Convergence to Efficiency of the Nikkei Put Warrant Market of 1989–90, *Applied Mathematical Finance* 2, 243–271.
- Skiena, S. S. 1988. A Fairer Scoring System for Jai Alai, *Interfaces* 18(6), 35–41.
- Thorp, E. O. 1975. Portfolio Choice and the Kelly Criterion, in W. T. Ziemba and R. G. Vickson (eds.), *Stochastic Optimization Models in Finance*. Academic Press, San Diego, CA, pp. 599–619.
- Thorp, E. O. 2006. The Kelly Criterion in Blackjack, Sports Betting and the Stock Market, in S. A. Zenios and W. T. Ziemba (eds.), *Handbook of Asset and Liability Management, Vol. 1: Theory and Methodology*, North-Holland, Amsterdam, pp. 385–428.
- Ziemba, R. E. S., and W. T. Ziemba. 2007. *Scenarios for Risk Control and Global Asset Allocation*, Wiley, New York.
- Ziemba, W. T., and D. B. Hausch. 1986. *Betting at the Racetrack*. Dr. Z Investments, Inc., San Luis Obispo, CA.

DOWNLOAD FULL VERSION HERE: <https://bettingbooks.org/12>

This page intentionally left blank

Download More Sports Betting Books at: www.bettingbooks.org

DOWNLOAD FULL VERSION HERE: <https://bettingbooks.org/12>

PART V: Semi-Strong Form Efficiency

Download More Sports Betting Books at: www.bettingbooks.org

DOWNLOAD FULL VERSION HERE: <https://bettingbooks.org/12>

This page intentionally left blank

Download More Sports Betting Books at: www.bettingbooks.org

CHAPTER 14

Semi-Strong Form Information Efficiency in Horse Race Betting Markets

M. Sung

Centre for Risk Research, School of Management, University of Southampton,
Southampton, United Kingdom.

J. E. V. Johnson

Centre for Risk Research, School of Management, University of Southampton,
Southampton, United Kingdom.

1. Introduction	276
2. Semi-Strong Form Efficiency in Horse Race Betting Markets: Single Variable Models	277
2.1. <i>Arbitrage Between Parallel Markets</i>	278
2.2. <i>Professional Predictions</i>	285
2.3. <i>Betting Volume</i>	289
2.4. <i>Post Position</i>	290
2.5. <i>Pedigree</i>	291
2.6. <i>Distance Preference</i>	292
2.7. <i>Single Variable Models: Overview</i>	292
3. Semi-Strong Form Efficiency in Horse Race Betting Markets: Multiple Variable Models	293
3.1. <i>Distribution-Based Methods</i>	294
3.2. <i>Distribution-Free Methods</i>	298
3.3. <i>Multiple Variable Models: Overview</i>	300
4. Semi-Strong Form Efficiency in Horse Race Betting Markets: Conclusion	301
References	302

Abstract

This chapter reviews the literature addressing the degree to which abnormal returns can be earned in horse race betting markets using publicly available information, other than odds alone. Consequently, the chapter examines the extent to which horse race betting markets are semi-strong form efficient. The conclusion to emerge is that horse race betting markets incorporate a significant amount of publicly available information. Bettors appear good at discounting simple, single variable information. However, they do not efficiently discount more complex information. Studies that examine the simultaneous effect of several variables on race outcome demonstrate that bettors often do not react to the subtle relationships that exist between variables.

1. INTRODUCTION

The degree and manner in which financial markets incorporate information provides important clues to the manner in which these markets operate. Wagering markets in general and horse race betting markets in particular have attracted considerable scrutiny in this regard since, as Sauer (1998, p. 2,021) observes, “wagering markets are especially simple financial markets, in which the scope of the pricing problem is reduced. As a result, wagering markets can provide a clear view of pricing issues which are complicated elsewhere.”

Horse race betting markets, in common with other financial markets, can be regarded as markets in information. “In its simplest formulation, the market for bets in an n -horse race corresponds to a market for contingent claims with n states in which the i th state corresponds to the outcome in which the i th horse wins the race” (Shin, 1992, p. 1,142). Investors in horse race betting markets purchase assets (place bets), returns to which depend on the result of the horse race to which the particular market relates. In state contingent claims terms, the purchase price of a claim on horse i in race j which pays £1 if horse i wins and nothing if it loses, is given by $1/(1 + O_{ij})$, where O_{ij} represents horse i 's odds. It is argued that bettors will continue to place money on a horse i in race j until the purchase price of a claim on this horse ($1/(1 + O_{ij})$) accurately reflects the market's best estimate of the horse's chance of winning the race (Figlewski, 1979).

Betting markets share many fundamental characteristics with wider financial markets including: large numbers of participants (market-makers, the holders of private information, institutional/professional buyers, and noise traders), ease of entry to the market, the availability of extensive information allowing both fundamental and technical analysis, the presence of complex and interdependent factors that influence an asset's value/bet's prospects, extensive market knowledge, and some similar behavioral biases (Law and Peel, 2002; Snyder, 1978a). In addition, with the advent of betting exchanges, bettors are now able to offer odds to other bettors; providing a close link with short selling in financial markets. In fact, decisions in betting and wider financial markets share many features in common. These include the collection, interpretation, and analysis of

quantitative and qualitative information, including that associated with expert opinion; the consideration of precedent, the formation of expectations, and the commitment of resources. In the light of these similarities, the value of betting markets in providing a window on decision-making in wider financial markets is well established (e.g., Snyder, 1978a; Hong and Chiu, 1988; Johnson and Bruce, 2001). Their value in this regard is enhanced because betting markets also possess a distinctive feature which makes it possible to discern behavior more clearly than in other financial markets: the generation of an unequivocal outcome (a winner) within a finite time frame. This provides a clear, objective benchmark against which to measure the appropriateness of the asset's price (revealed as odds), which helps in the assessment and understanding of factors influencing market efficiency. The finite nature of racetrack betting markets also means that there is a large pool of markets (races) of essentially similar type available for analysis.

In summary, racetrack betting markets share many features in common with wider financial markets and they also possess features which enable insights into the manner and degree to which information is used in markets to be clearly discerned. Consequently, "it is now appreciated that betting markets are well suited to testing market efficiency" (Law and Peel, 2002, p. 327). With this in mind, the aim of this chapter is to explore the literature associated with semi-strong form efficiency in horse race betting markets.

A betting market which is semi-strong form efficient incorporates all publicly available information into odds. Consequently, returns should be equal across identical betting opportunities and abnormal returns should not be available to those who use publicly available information. Clearly, publicly available information includes the odds themselves, and many studies have explored the extent to which abnormal returns can be made in horse race betting markets simply using information contained in odds. The results of these weak-form efficiency studies are reviewed elsewhere in this volume. Consequently, the studies examined here are restricted to those that investigate the extent to which abnormal returns are available using publicly available information other than odds alone. Consequently, the studies discussed here explore the extent to which information contained in odds, *together* with some supplementary information (e.g., in which market, pari-mutuel or bookmaker, the odds were formed) is employed by bettors. The chapter is structured around the two main classes of semi-strong form study: those which employ a *single* variable associated with publicly available information (in addition to odds) (see Section 2), and those which simultaneously employ *multiple* variables (section 3).

2. SEMI-STRONG FORM EFFICIENCY IN HORSE RACE BETTING MARKETS: SINGLE VARIABLE MODELS

A number of authors have examined the extent to which horse race betting markets incorporate various, *single* types of publicly available information into odds. These studies are classified here into those in which odds information in one market is effectively incorporated into odds in a parallel market (Section 2.1) and those which examine the

degree to which professional predictions (Section 2.2), betting volume (Section 2.3), post-position (Section 2.4), pedigree (Section 2.5), and distance preference (Section 2.6) are incorporated into odds. A summary of the conclusions to emerge from these studies is given in Section 2.7.

2.1. Arbitrage Between Parallel Markets

In certain contexts in horse race betting markets, odds on the same race may be formed in two or more parallel markets. If the odds, *together* with a knowledge of which market formed these odds, can be used to make abnormal returns in a parallel market, then the horse race betting market is not semi-strong form efficient. Research exploring this form of efficiency falls into three categories, namely, those which examine (i) cross-track betting, whereby betting pools on the same race are formed in two different locations and information on odds obtained in one market is used to place bets in the other market, (ii) arbitrage between betting pools for different bet types; for example, using the odds information from the win pool to bet in the place and show pools, and (iii) differences in odds formed in two parallel betting mediums; for example, comparing pari-mutuel odds with fixed odds formed in a parallel bookmaker market.

2.1.1. Cross-Track Betting

Cross-track betting markets operate where independent betting pools on the same race are formed at different racetracks, depending on the amounts bet at those tracks. These are distinguished from inter-track wagering markets, where all amounts bet at different racetracks on a given race are used to form one common betting pool. Clearly, in cross-track betting markets, the odds on a given horse may differ from one track to another. However, if there were no arbitrage costs involved, the odds at the various racetracks should be identical if the market is efficient. Bettors can generally only access publicly available information on the odds at the track at which they are located. In addition, they are only able to bet at the racetrack at which they are located except through establishing a syndicate to act across tracks. A syndicate is composed of a decision-making center, individuals at the various tracks, and communication networks. Some researchers have examined the extent to which markets in these contexts are efficient.

Hausch and Ziemba (1990) explored the efficiency of the North American cross-track wagering market. They obtained final win odds for each horse from various racetracks in popular Triple Crown races. They investigated two wagering strategies. In the first they selected the highest odds available for each horse in a race and employed these to bet enough on each horse to return \$1. Obviously, if a total of less than \$1 is needed to be wagered on a given race, then a risk-free return is possible. In fact, they discovered that this strategy generated an average risk-free profit of 4.7% on Triple Crown races between 1982 and 1985. The second wagering strategy involved the Optimal Capital Growth Model (Kelly, 1956), which maximizes the expected logarithm of one's final wealth. This approach involved adjusting the odds at the home track for the favorite-longshot bias, as Snyder (1978a) proposed, and normalizing the resulting

probabilities based on win odds. The expected return for each horse based on these normalized probabilities was determined and horses with expected returns exceeding a certain level (\$1.10 to a \$1 stake) were bet (assuming a maximum fixed, not accumulated, level of wealth available to bet on each race). This produced an overall return of 15% although this was not significant at any conventional level. Hausch and Ziemba (1990) argued that differences in final odds at different racetracks which make profitable arbitrage possible might be explained by regional differences in the familiarity of bettors with local horses' performances. However, while this study demonstrated that abnormal returns are possible from cross-track betting, there are a number of reasons why these returns may not be available in practice. For example, implementation costs, such as communication costs and administrative expenses of a syndicate, were not accounted for in the model. In addition, the time required to execute the arbitrage strategy may be insufficient for the syndicate to communicate and make the betting decision, but sufficient for the odds to change adversely. In light of these limitations, Sauer (1998, p. 2,048) concluded that Hausch and Ziemba's results represent "another crease in what is predominantly a smooth pattern of efficiency in the racetrack betting markets."

Leong and Lim (1994) replicated Hausch and Ziemba's (1990) study in the cross-track betting market between Singapore and Malaysia. They used a relatively large sample, involving 9,839 horses running in 867 races. Leong and Lim (1994) discovered that in 1.5% of the races in their sample a risk-free return could be obtained. In addition, when they restricted bets using the Kelly criterion to those horses offering expected returns of 2.0 and 2.5 (to a one unit stake) returns of 37% and 65%, respectively, were obtained; these were significantly different from zero at the 15% level. However, these profits may not be available in practice, since communication devices are not permitted at racetracks in either Singapore or Malaysia. In addition, odds information is not available to bettors in the country where the race is not taking place and police are active in shutting down illegal communication centers.

In summary, the cross-track betting papers suggest market inefficiency but physical and cost limitations may make profitable arbitrage difficult in practice.

2.1.2. Arbitrage Between Betting Pools

In pari-mutuel markets, a number of alternative bet types are offered. Consequently, a horse may be bet to win, to place (to finish first or second), to show (to finish first, second, or third), or to finish in the first two or three (exacta and trifecta, respectively; where the order of the first two/three horses must be specified). Equally, there are bets which span more than one race. For example, the "daily double" is a pari-mutuel bet paying a return only if a bettor selects the winners of two specified races, the bet being placed before the first of the two races is run. Bets on a given horse on these various bet types enter a separate pool. Within the definition of semi-strong form efficiency used in this chapter, it should not be possible to use the information concerning the horse's odds in one of these pools (together with the knowledge of which pool these odds relate to) to bet profitably on that horse in another pool.

A number of studies have examined this aspect of market efficiency. For example, Ali (1979) compared the odds on horses in the daily double pool with the odds in the win pools of the two races (i.e., assuming that the two horses are bet in succession, the total winnings from the first horse being placed on the second horse—a parlay). Ali's (1979) analysis employed data from 34 racetracks in North America and Canada in 1975, and confirmed that the win pool odds provide efficient estimates of the daily double pool odds; a result consistent with the hypothesis that the betting market is semi-strong form efficient. The daily double was also explored by Asch and Quandt (1987) using data from the Meadowlands racetrack in the U.S. during 1984, and by Lo and Busche (1994) on data from the same racetrack in 1984 and on data from Hong Kong between 1981 and 1989. Both these studies concluded that the daily double bet is significantly more profitable than the equivalent parlay. However, when differential track takes on the two bet types are taken into account, the differences in expected return disappear.

Hausch et al. (1981) used probabilities revealed by win pool betting to construct probabilities of horses finishing second and third. They then constructed profitable wagering strategies based on differences between these probabilities and the odds available in the place and show pools taking account of investor preferences and the effect of the bet size on the pari-mutuel pool. One practical barrier to operating this arbitrage strategy is that it relies on the final odds being known when the bet is struck. However, Hausch et al. (1981) demonstrated that a profitable arbitrage strategy can still be constructed using odds available in the pools 2 min before the race starts. In addition, they developed profitable regression-based solutions with limited data requirements that approximated their full complex model and demonstrated that the necessary calculations could be constructed in the time available at the racetrack. The approach adopted by Hausch et al. (1981) was developed further in Hausch and Ziemba (1985, 1990, 1991) and in Ziemba and Hausch (1984, 1987), including betting strategies that involve a combination of win and place/show bets which guarantee a positive profit.

The anomaly between place and show pools was also developed into a practical system (the Dr. Z system) for employing information on win odds to trade profitably in the place and show markets in North America (e.g., Hausch and Ziemba, 1985). A similar approach was employed by McCulloch and van Zijl (1986) in the New Zealand horse race betting market (where, similar to North America, only pari-mutuel operators are permitted to offer odds, on and off-track). They found a small and variable underestimation of show odds when these were predicted using win odds; suggesting that the model could not be employed profitably in New Zealand. Ziemba and Hausch (1994) also adapted the Dr. Z system for betting on races in the UK where place bets are normally restricted to those horses finishing in the first three. However, it was suggested that the significantly larger deductions from the pari-mutuel pools in the UK (cf. North America) may reduce the profitable opportunities available.

An approach based on that adopted by Hausch et al. (1981) was also employed by Swidler and Shaw (1995) at Trinity Meadows Raceway, a small racetrack in Texas. By betting in the show pool on horses that were first or second favorite in the win pool and those whose expected payoff was at least 1.1, expected returns of 18.5% were realized

which were statistically different from zero at the 9% level. However, only very modest gains could be made using this approach since (a) only about two or three bets would be identified in a typical 10 race card, (b) even relatively small bets would reduce the returns significantly due to the small pool size at the track, and (c) odds close to the market close were often not a good proxy for final odds. To help overcome these limitations, they explored a strategy which required less information: backing (in the place pool) any favorite in the win pool which was ranked 2–12 in the place pool. This produced a return of 2.9%. Ziemba (2008) updates the results on U.S. Dr. Z betting to the current situation in 2008; with rebates there are still profits.

Anomalies between win and place odds have also been explored in Australia, where bookmaker and pari-mutuel markets operate alongside each other at racetracks. Tuckwell (1981) used win probabilities derived from the bookmaker markets at Sydney and Melbourne racetracks to determine the “true” odds for horses to finish in the first three (to be “placed”). Tuckwell (1981) demonstrated that a strategy of betting on horses where the true place odds exceeded the place odds available in the pari-mutuel pool would have produced a return on outlay of 20%. However, two barriers exist to effective practical application of this approach. First, the place odds are based on bookmaker starting prices (i.e., odds available at the start of the race) and to operate the strategy approximations to these values (which may vary substantially from the correct values) would need to be obtained several minutes before the close of the market. Second, the strategy did not allow for the effect of system bets depressing the place odds available in the pari-mutuel pool. Despite these barriers, Tuckwell (1981) concluded that significant inconsistencies do exist between the win and place racetrack betting markets in Australia. Edelman and O’Brian (2004) developed a game theoretic approach to identifying opportunities for combining various forms of bet (e.g., win, place, exacta [first two in correct order], trifecta [first three in correct order]) in such a way as to produce a guaranteed profit. The strategy was tested on 2,667 races run in Australia in 2000, and they identified a number of races in which guaranteed profits could be earned. However, Edelman and O’Brian (2004) pointed out that the practical application of such an approach would be hampered, since the results were based on final dividend returns (which would not be known at the time the bet was placed) and large bets into the pools (which would be required to earn sizeable returns) would substantially reduce the returns themselves.

Jackson and Waldron (2003) demonstrated that anomalies in the place betting market can also be created by the pari-mutuel operator’s method of calculating place dividends. In the late 1970s in the UK and in 1995 in Ireland, a different method of calculating these returns was introduced (including a minimum guaranteed return) which differed from that used elsewhere in the world. Under these conditions, Jackson and Waldron (2003) demonstrated that in certain races, where particular betting patterns pertain (e.g., the fraction bet on the favorite is large), it is possible to construct a betting strategy that results in a positive expected return to the bettor and where the pool operator can expect to lose up to 50% of the pool. This occurs because instead of returning the initial stake, as occurs in standard pari-mutuel markets, the net pool is simply split between those that collect. Consequently, the minimum guaranteed return is easily reached (see Ziemba

and Hausch, 1987). Subsequently, the pari-mutuel operators changed their practices to close this opportunity. Their approach yielded large profits before the rules were changed. In their *Locks* paper, which motivates this application, Hausch and Ziemba (1991) found only infrequent arbitrages with small profits. But these opportunities still exist in the U.S. and Canada.

In horse race betting markets, in North America, the United Kingdom, and Australia, anomalies remain between different bet types on the same horse in a given race. Knowledge of which market produces the most accurate probabilities (or what practices the pari-mutuel operator employs) can be used to construct profitable betting strategies. Since knowledge of both the odds *and* the market in which they are formed is required to exploit these anomalies, this is evidence of semi-strong form inefficiency.

2.1.3. *Parallel Betting Mediums*

Within the UK horse race betting market, it is permissible to bet on a race with either the monopoly pari-mutuel operator, the Tote, or with fixed odds bookmakers. If these markets are semi-strong form efficient, then profitable arbitrage between the betting mediums should not be possible.

Gabriel and Marsden (1990, 1991) were the first to directly investigate the efficiency of these parallel markets. They explored the odds of the winners of 1,427 flat races run in the UK in 1978. To avoid identifying differences in average odds in the two markets resulting simply from a few very large pari-mutuel odds, they truncated the dataset by removing horses with odds greater than a pre-defined level (i.e., six subsets, where this level was set at 20/1, 15/1, and 10/1 in (a) the bookmaker, and (b) the pari-mutuel market). Their corrected results (Gabriel and Marsden, 1991) showed that the pari-mutuel odds were consistently higher than starting prices, whether for favorites or longshots, where the dataset truncation was based on the bookmaker market odds. Furthermore, an ordinary least squares regression of the pari-mutuel odds on the starting prices produced a constant that was significantly different from zero and a slope coefficient that was significantly different from unity; again, suggesting that odds in these two parallel markets are different. Gabriel and Marsden (1990, 1991), therefore, argued that the UK betting market did not satisfy the conditions of semi-strong form efficiency. They suggested that the differences arose from: (i) different mechanisms of price formation in the two markets, (ii) reaction on the part of bookmakers to protect themselves from insiders, and (iii) barriers to arbitrage. Each of these explanations is now explored briefly. First, in the pari-mutuel market, odds are determined solely by the relative amount of money bet on each horse, and, hence, they are continuously priced. In bookmaker markets, on the other hand, the odds are offered in fixed discrete intervals. For example, pari-mutuel odds of 85/1 fall between two of the bookmakers discrete odds values (i.e., 66/1 and 100/1) and a profit maximizing bookmaker may well choose the lower of these. Differences in the price formation mechanism in the two markets may, thus, account for some of the discrepancy between odds in these parallel markets. Second, those who hold private (inside) information about a horse's enhanced prospects of success in a particular race are likely to bet with bookmakers rather than with the pari-mutuel operator, since their bet will be settled at the odds prevailing in the market at the

time the bet is struck. If they bet with the pari-mutuel operator, their returns could be eroded by other bettors observing a move in the market odds and betting on the horse themselves. Consequently, Gabriel and Marsden (1990, 1991) argued that bookmakers, in order to protect themselves from informed traders, may deliberately depress all their odds. Third, they argued that there are physical barriers to arbitrage. At the time of the study, Tote odds were not displayed electronically and pari-mutuel bettors had little information concerning trends in the pari-mutuel odds or what the final pari-mutuel odds would be. Bettors would, therefore, have little information on which to compare odds in the two markets and this would act as a significant barrier to arbitrage.

Cain et al. (2001) questioned the appropriateness of the regression methods employed by Gabriel and Marsden (1990, 1991). In particular, they argued that the results would be unduly influenced by large numbers and large outliers associated with longshots and that the error term would be heteroscedastic and non-normal. Their analysis of the same data employed by Gabriel and Marsden (1990, 1991) confirmed their concerns. Cain et al. (2001) corrected for these problems in three separate ways: using logarithms of tote and bookmaker odds rather than the odds themselves, using robust estimators, and trimming the datasets to omit outliers. Their refined results suggested that Tote odds did not consistently exceed bookmaker odds. However, they discovered that bookmakers' odds were more generous than those offered by the pari-mutuel operator for low odds winners, but the pari-mutuel odds were more generous on winners starting at higher odds. These results were confirmed by Blackburn and Peirson (1995), Bruce and Johnson (2000b), and Peirson and Blackburn (2003). Blackburn and Peirson (1995) compared the odds offered in the bookmaker and pari-mutuel market in the UK for the whole of the 1993 flat season and found that on average, bookmaker odds were higher at odds of less than 5/1 and the pari-mutuel odds were higher at odds of greater than 5/1. They suggested that these differences arose because demand elasticity for betting on longshots is relatively low. Consequently, profit maximizing bookmakers can offer reduced odds on these horses with little reduction in turnover. They argued that relatively high elasticity arises for short odds horses because past form is more likely to be available and heavily scrutinized for these horses than for longshots, leading to more consistent views about the short odds horses' probability of winning. On the other hand, they argued that less continuous and informative form records for longshots is likely to result in more diverse views about their chance of success, which results in bettors being less sensitive to the odds that bookmakers offer on longshots. Equally, they argued that bookmakers can attract bets on less favorably priced longshots because bettors with bookmakers (cf., the pari-mutuel operator) can secure their returns at the odds prevailing in the market at the time the bet is struck, rather than at the odds prevailing at the close of the market (which applies to pari-mutuel bets).

Peirson and Blackburn (2003) identified a very large difference between mean pari-mutuel (39.97/1) and bookmaker odds (26.27/1) for extreme longshots (bookmaker odds $\geq 20/1$). The relatively large standard deviation of the pari-mutuel odds compared with that for bookmaker odds (29.17 compared to 8.96) suggested that the large difference in mean odds arose from some very large pari-mutuel odds. In fact, Peirson and Blackburn (2003) argued that those horses that are unsupported drift out dramatically in the pari-mutuel market and to a far lesser extent in the bookmaker market.

This, they argued, arises partly because bookmakers are seeking to protect themselves from potentially large losses and partly because of bookmakers' incomplete menu of odds; bookmakers only updating odds to a new category when the exact odds they wish to offer exceeds an allowed odds category by a sufficient margin.

Bruce and Johnson (2000a, 2005) identified significant differences in the favorite-longshot bias in parallel bookmaker and pari-mutuel markets, and offered both demand and supply side explanations for why the odds may differ in these two markets.

Vaughan Williams and Paton (1997) confirmed Blackburn and Peirson's (1995) conclusion that bookmaker odds are significantly higher than pari-mutuel odds for shorter odds horses. However, Bruce and Johnson (2000b) argued that these results failed to capture data concerning non-winning horses that may have an important effect on the overall disparity in odds between the two markets. In addition, by failing to capture the idiosyncrasies of betting market structure and process, it was argued that the results may overestimate the opportunities for arbitrage between these markets.

Bruce and Johnson (2001) suggested that Gabriel and Marsden's (1990, 1991) approach of removing horses whose bookmaker odds exceed a certain level in order to reduce inflationary bias in the pari-mutuel mean odds fails to achieve its aim and that the results they obtained may, therefore, be suspect. In contrast, Bruce and Johnson (2001), using a sample of 2,109 races run in the UK in 1996, only removed horses whose *pari-mutuel* odds exceeded a certain level from the analysis. In addition, unlike all previous papers investigating differences in pari-mutuel and bookmaker odds, their dataset was not restricted to race winners, but included all runners in the race. Their subsequent distribution-based and distribution-free tests revealed that bookmaker odds exceeded the pari-mutuel odds where the pari-mutuel odds were less than 10/1 but that otherwise the pari-mutuel odds exceeded bookmaker odds. Bruce and Johnson (2001) argued that the market for shorter odds horses is particularly competitive, since most wagers would be placed on such horses. Consequently, since bookmakers are motivated by turnover as well as margin (Tuckwell, 1983) they are likely to offer relatively good value odds on short odds horses to attract customers. Equally, they offer relatively poor value odds on longshots for which there is less competition. Johnson and Bruce (2001) also pointed to a number of physical barriers to arbitrage which make the elimination of this inefficiency difficult in practice. One of the most important of these barriers resulted from the fact that neither the final pari-mutuel odds nor the final bookmaker odds can be observed until after betting activity has ceased and late bets into the pari-mutuel pool (both on and off-track) can significantly alter final odds (Bruce and Johnson, 2000a).

In Australian racetrack betting markets, where, once again, bookmaker and pari-mutuel markets operate alongside each other, it has been discovered that differences between odds in the two markets tend to reduce toward the close of the market (Bird and McCrae, 1994). However, disparities in odds remain, but here the disparities depend upon whether a horse's odds lengthen (bookmaker odds shorter) or shorten (pari-mutuel odds shorter). Despite these findings, Bird and McCrae (1994) could not construct a profitable trading rule that exploited the anomalies.

In summary, the studies exploring odds offered in parallel bookmaker and pari-mutuel markets suggests that persistent differences in the odds offered in these markets

remain. A range of explanations for the differences in odds observed in these two parallel markets have been provided, involving the mechanics of price formation, economic, behavioral, and arbitrage barrier aspects. The odds levels at which bookmaker odds exceed the pari-mutuel odds (and vice versa) differ between studies, and the precise nature of the bookmaker/pari-mutuel market anomaly appears sensitive to the methodology employed. However, the consistent disparity of odds observed between these parallel markets may be thought to suggest some level of semi-strong form inefficiency, although no study has explicitly tested to what extent these differences in odds can be used to earn abnormal returns. In addition, it has been shown that disparities between odds in these two markets are only a sign of inefficiency if the representative bettor is risk neutral (Cain et al., 2003). They demonstrated that a representative bettor exhibiting risk loving/averse preferences over favorites/longshots would lead to the observed anomalies between the pari-mutuel and bookmaker odds.

2.2. Professional Predictions

There is some debate concerning the status of professional predictions of race outcomes in the efficiency literature. There are those who consider predictions of experts, whether or not they are available to the general public, as the work of insiders. Consequently, they argue that tests of whether abnormal returns can be achieved using these predictions are strong-form tests of efficiency (Snyder, 1978a, 1978b). Equally, there are those who argue that predictions are generally based on publicly available information and that market efficiency tests based on predictions that are widely published (e.g., in daily newspapers) should be regarded as semi-strong efficiency tests (e.g., Losey and Talbott, 1980). It is suggested here that *any* professional prediction that is available to the public, whatever its cost (varying from the price of a daily newspaper, to a moderate price paid for a telephone prediction from a pundit, to perhaps a much higher price paid for a private prediction service) is, in effect, information available to the public. Clearly, this is a matter for debate, but in this section tests of the profitability of all professional predictions will be treated as tests of semi-strong form efficiency. The section is divided into those studies that focus on predictions provided in newspapers (Section 2.2.1), and those that are available on subscription (Section 2.2.2).

2.2.1. Newspaper Predictions

Snyder (1978a, 1978b) used a sample of 7,657 horses running in 846 races at Arlington Park, Chicago, USA to compare the rates of return obtainable from bets placed at the final pari-mutuel odds with bets placed at odds predicted by five representative experts: track handicappers (whose predicted odds are widely available at the racetrack), and four newspaper tipsters in the Daily Racing Form, Tribune, Daily News, and Sun Times. The main conclusion was that the experts' odds exhibited more bias than those of the general public; the former group underestimating favorites' odds and overestimating longshots' odds more than the latter group. Snyder (1978a) suggested a number of possible reasons for this. First, his interviews with the experts revealed that experts did

not try to predict the true winning probability of each horse; rather, they tried to predict what the odds would be in the pari-mutuel market. Second, there was an unwritten rule never to quote odds of greater than 30/1, which clearly depressed the odds on longshots. Third, their odds usually had to be forecast on the day before racing, and this exposed the experts to considerably more uncertainty than the general public, who were betting on the day of the race (e.g., changes in the weather or a withdrawal might significantly increase/decrease a horse's chance of success). The less information available, the more the odds on all runners are likely to be equal. Therefore, the greater exposure to uncertainty has the effect of underestimating (overestimating) the favorite's (longshot's) odds in the experts' predictions. In summary, Snyder (1978a, 1978b) concluded that positive profit could not be achieved using the experts' predictions. A similar conclusion was reached by Figlewski (1979) who analyzed the forecasts of 14 professional handicappers whose predictions appeared in the Daily Racing Form, the New York Post, and the New York Daily News. Figlewski (1979) developed multinomial logit models that incorporated (a) track odds, (b) handicapper odds, and (c) track odds and handicapper odds, respectively, using the outcome of 189 races. The results demonstrated that the handicappers' predictions contained valuable information but that this was to a large extent discounted in the final odds. Consequently, Figlewski (1979) concluded that these markets are semi-strong form efficient with respect to published predictions of professional handicappers.

Losey and Talbot (1980) used a similar dataset to that employed by Snyder (1978a, 1978b), based on races run at Arlington Park, Chicago in 1978, augmented with data from the following meetings in Hawthorne, Chicago. They investigated the profitability of a strategy of betting on horses predicted by the Racing Form handicapper to start at short odds ($\leq 3/1$) but actually go off at longer odds. The strategy yielded even larger losses (28.4%) than the track takeout rate (17%), suggesting that the experts were even worse at predicting horses' probability of winning than were the general public.

The studies discussed above were conducted in the pari-mutuel betting market, but studies by Bird and McCrae (1987), Tuckwell (1983), and Smith (2003) have extended these to bookmaker based markets. Bird and McCrae (1987) tested the extent to which the Australian bookmaker market was efficient with respect to 10 experts' predictions published in a Melbourne newspaper. A horse predicted to finish first, second, and third by each expert was awarded three, two, and one points, respectively. The horses in a given race were then ranked in order of favoritism on the basis of the aggregate opinion of the 10 experts, and on the basis of their degree of favoritism (based on odds) in the bookmaker market at the market close. The rate of return for horses in each category of favoritism across a sample of 1,026 races run at Melbourne racetracks during 1983–1984 was computed, assuming \$1 was bet on each horse in that category at bookmaker, market close odds. Positive returns were not made in any of the categories and there was no significant difference between the rates of return where horses were ranked on the basis of favoritism by bookmaker market odds, or experts' opinions. Bird and McCrae (1987) concluded that information provided by the newspaper handicappers had been fully incorporated into the bookmaker odds.

Tuckwell (1983) examined the extent to which winning probability predictions given by “Computercard,” which is published in the daily press for races run in Sydney and Melbourne, could be used to make abnormal returns. A strategy of betting one unit on any horse whose predicted winning probability was one of the three highest in the race and exceeded the winning probability implied by the starting price yielded a profit of 6.5% when applied to all races run in 1974. However, subsequent studies in Australia have failed to find profitable strategies based on forecast odds (e.g., Anderson et al., 1985).

Smith (2003) focused on the movements of bookmakers’ odds for horses selected by a leading newspaper tipster, “Winsome,” and other newspaper tipsters in the UK. He suggested that the tipsters’ selections may help to identify those horses whose odds are likely to contract during the betting period and that these might, therefore, offer an arbitrage opportunity. Based on a sample of 4,728 horses running in Saturday races over a three year period (1998–2000), he classified horses according to whether they had been selected by Winsome only, Winsome and other tipsters, other tipsters only, or no tipsters. The winning probability implied by bookmaker odds for those horses selected by Winsome increased on average by 1.89% whereas for those horses that were not selected by any tipster, their implied probability decreased by 0.26%. Smith (2003) calculated the rates of return for each category based on the bookmaker odds available in the morning before the race and at the bookmaker odds available at market close. For all those categories including Winsome selections a positive rate of return was identified (e.g., 37.03% and 7.53% for bets at bookmaker morning and starting prices, respectively, for the category “Winsome only”) whereas negative returns were identified for all other categories (“other tipsters only” and “no tipsters”). Statistical tests did not confirm positive returns for the Winsome selections but Smith (2003) argued that this may be due to the nonnormal leptokurtic nature of the returns’ distribution. Consequently, Smith (2003) concluded that the evidence pointed to bookmakers’ morning odds and even the closing market odds did not fully discount Winsome’s selections.

The papers exploring the extent to which selections by newspaper experts are discounted in horse race odds reveal a fairly consistent set of results across a variety of countries. The advice provided by newspaper experts appears to contain useful information but this information is, to a large extent, discounted in final odds in both pari-mutuel and bookmaker markets. Only two studies report that it *may* be possible to earn abnormal returns by trading on tipsters’ information. However, one of these (Smith, 2003) requires that bets are made in the early stages of a bookmaker betting market, and the second (Tuckwell, 1983) requires knowledge of the starting price, which is not available until after the race has started.

2.2.2. Predictions Available by Subscription

The few studies that have examined predictions available by subscription fall into two categories: those that examine the performance of selections indicated by betting systems, and those that examine the specific selections of professional tipping services. Betting systems do not strictly fall into the category of approaches employing only a

single variable (in addition to odds), since they often require several pieces of information concerning a horse's previous performances. However, these systems often employ simple elimination rules (eliminating runners from consideration in order to leave a single selection in a race), or selection rules (i.e., selecting the positive qualities of a single runner on which to bet); rather than simultaneously considering several variables in order to predict a horse's probability of success. These more sophisticated modeling approaches will be examined in Section 3.

Betting Systems Crafts (1994) examined the performance of three marketed betting systems that employed publicly available information to select a horse in a given race. All three systems produced negative post-tax rates of return for periods both before and after the systems were published. There was also a significant decline in the returns for the one system that appeared to yield pre-tax profits in the period before it was published. In addition, Crafts (1994) reported the results of a study conducted by Roberts and Newton (1987) that explored the profitability of nine racing systems. Eight of these produced negative returns and the one that produced positive returns was based on only 12 observations. Using a sample of 102 races run at Santa Anita Park, California in 1972, Vergin (1977) tested the profitability of six betting systems that required information concerning the horses' odds and/or previous performances. Only one of these produced profits to level stakes betting and the sample size was insufficient to produce a statistically significant result.

Professional Tipping Services Professional tipping services differ from betting systems, in that the latter provide a set of decision rules that the purchaser of the system is required to follow, whereas the former provide horse selections in a given race. Few studies have examined the degree to which abnormal returns can be earned using professional tipping services, and those that have been conducted failed to identify profits significantly different from zero. For example, Roberts and Newton (1987) examined five services offering advice on which specific horses to bet during the 1970s in the UK, but all of these produced negative returns. Vaughan Williams (2000) investigated the performance of five leading professional forecasting services in the UK in 1995 on the basis of bookmaker starting prices. A number of different trading strategies were explored and each of the services yielded a pre-tax profit at starting prices. It appeared that higher profits could be earned if the staking plan and minimum odds stipulations advised by some of the services were followed. However, only some strategies showed a post-tax profit. These profits were not statistically significant and the results did not account for the cost of the services, including subscriptions and/or the cost of premium-rate telephone calls. On the other hand, as Smith (2003) observed, relevant tipsters' information appeared to be incorporated into odds as the betting market develops. Consequently, it may be possible to earn positive returns in a bookmaker-based market by backing the professional forecasters' selections at the odds prevailing close to the start of the market (for further discussion of the role of professional tipsters see Deschamps and Gergaud, 2008).

2.3. Betting Volume

Trading volume is regarded as an important element of technical analysis in financial markets. However, until recently, few studies have explored the role of trading volume on horse race market efficiency. Notable exceptions include Busche and Walls (2000), Walls and Busche (2003), and Bruce et al. (2003). Busche and Walls (2000) were the first to observe the connection between betting volume and market efficiency (defined in terms of the degree to which returns were equalized across horses with different odds). In particular, they analyzed a sample of 10,000 races in pari-mutuel markets in North America, Japan, Hong Kong, and Macau and found that returns were equalized at all racetracks where the daily bet volumes exceeded \$1.8 million. In addition, they noted that previous papers, which had found discrepancies in pari-mutuel returns at different odds (the favorite-longshot bias), were all conducted at racetracks where the daily betting volume was between \$0.3 million and \$1.8 million. These conclusions were reinforced by their later study (Walls and Busche, 2003) which examined 13,000 races at 18 different racetracks across Japan, with pari-mutuel betting volumes varying from \$30,000 to \$3 million. Once again, returns were equalized at the high volume tracks. Busche and Walls (2000) argued that this results from the presence of professional bettors, who are attracted to the large volume markets where their large bets will not unduly reduce their own returns. However, at low volume tracks, they argued that casual bettors are likely to dominate. Since these bettors largely attended for leisure and social reasons, they tended to back horses on the basis of their own subjective preferences rather than on objective criteria. Consequently, for example, their preference for excitement may result in the over-betting of longshots.

Employing the ideas generated by Busche and Walls (2000), Bruce and Johnson (2005) sought to explore whether information contained in betting volume in the pari-mutuel market in the UK is fully discounted in the final pari-mutuel odds. Using a sample of 2,078 races from 49 racetracks across the UK, they developed conditional logit models incorporating (a) odds alone, and (b) both betting volume and odds. A likelihood ratio test revealed that information concerning betting volume was not fully incorporated into the final pari-mutuel odds and the market could, therefore, be regarded as semi-strong form inefficient. Bruce et al. (2003) demonstrated that the market is particularly inefficient in this regard where holders of private information are most likely to be present. In addition, using a model proposed by Camerer (1998), they suggested that this may result from the holders of private information manipulating the market in order to secure the best odds on their selection. These conclusions suggested that abnormal returns may be available by betting on horses that attract high betting volumes in races most likely to attract the holders of private information. Bruce and Johnson (2005) argued that in the UK these are the low class, non-handicap races where relatively little information is available about runners' previous history, where the prize money is insufficient as an incentive to owners (who, therefore, seek to maximize their gains from betting), where the race conditions (e.g., weight carried by horse) are to some extent under the control of the horse owner/trainer, and where media scrutiny is generally low.

In summary, the few studies that have explicitly explored the impact of betting volume in horse race markets suggest that its impact on market efficiency has been

underestimated. It has been demonstrated that betting volume impacts on the degree to which returns are equalized across odds and volume does not appear to be fully incorporated into odds in races that attract the holders of privileged information.

2.4. Post Position

Horses are generally required to begin races run on the flat from starting stalls. These devices ensure that all horses are released to start the race simultaneously. Each horse is randomly allocated a post-position within the starting stalls, and these are normally announced the day before the race. The post-position determines where in relation to the inside of the racetrack the horse's starting stall is located. Certain post-positions may be advantageous because of track configuration (e.g., short oval racetracks with sharp bends), or other racetrack topography (e.g., faster ground on the outside of the track). Semi-strong form efficiency in betting markets requires that it is not possible to construct profitable trading rules based on a horse's post-position.

Most racing publications agree that post-position needs to be considered carefully when assessing the chances of each horse. For example, Cotton (1990, p. 113), in discussing racing in the UK (where most races are run on turf), advised "make no mistake [post-position] can be the most important component in the outcome of many flat races." Similarly, Beyer (1983) observed that horses with inside post-positions in races run in North America (where most races are run on materials that incorporate sand) often have an advantage on tracks with sharp bends and short straights. However, the advantage gained from a shorter distance traveled by these inside runners is not guaranteed, since the banking of sand-based tracks can allow water to accumulate on the inside rail in wet conditions, and this can lead to softer, slower going. Despite this observation, Beyer's view was confirmed by Quirin (1979). He found that in 2,516, 6 furlong races run over tracks of 1 mi circumference in North America, post-position one was the most advantageous and positions 2–6 produced winners more often than could be attributed to chance. Similarly, Bolton and Chapman (1986) and Chapman (1994) both incorporated post-position in multi-variable conditional logit models to predict winning probabilities (see Section 3 for more details). The former developed and tested their model on 200 races run in North America and the latter on 2,000 races run in Hong Kong (with configuration and surfaces similar to tracks in North America). In both cases, the lower the post-position, the greater was the predicted probability of the horse winning. It is impossible to discern from Bolton and Chapman (1986) or Chapman (1994) whether a profitable trading rule based purely on post-position and odds could be constructed. However, in Quirin's (1979) study, strategies of betting a unit stake on most of the inner post-positions produced losses because the advantage was spotted and overbet by the public. However, horses drawn in post-position one produced a level stake profit of 8%, suggesting a degree of semi-strong form inefficiency.

Canfield et al. (1987) investigated the efficiency of betting markets with respect to post-position bias by examining the results of 3,345 races run at a very tight oval track in Vancouver (Exhibition Park, circumference of about two-thirds of a mile)

during 1982–1984. They ranked horses by post-position and observed the percentage of horses winning and the rates of return from each post-position. Inside post-positions produced a significantly greater percentage of winners than outside ones, and returns from wagers on post-positions one and two were higher than other post-positions. As might be expected, the bias was more pronounced on dry days (when the track was running fast, there was no disadvantage from slower ground conditions on the inside rail), and in longer races with more turns. Canfield et al. (1987) constructed some profitable win and exotic bet trading strategies for particular post-positions, under certain track conditions, for short periods of time (up to one season). However, these positive returns disappeared over time as bettors became aware of the advantage. In addition, taking transaction costs into account, the positive returns were not statistically significant.

Betton (1994) also examined races run at Exhibition Park (1,062 races during 1987). She compared the actual and expected finish position from each post-position and concluded that the inside post-position held a significant advantage. Betton (1994) also developed a probit model to explore the influence of post-position on winning probability. This analysis suggested that the accuracy of winning probabilities based on odds rankings or odds and the number of runners could be improved by information concerning post-position. However, in contrast to Canfield et al. (1987), she discovered that knowledge of race distance did not improve the probability estimates. Results given in her paper suggested that positive profits could be obtained by level stake win betting on post-positions one and two in *shorter* races (involving one turn). Despite the fact that these results related to only 40 bets in one season on each of the two inside positions, it is interesting to note that the greater advantage of inside post-positions in longer races at Exhibition park in 1982–1984 (noted by Canfield et al. 1987) appeared, by 1987, to have been fully discounted by the betting public.

Johnson et al. (2005) employed a conditional logit model, incorporating odds and post-position, to examine the degree to which post-position bias is accounted for in betting decisions at a tight (circumference of approximately 1 mi) oval racetrack in the UK (Wolverhampton) over the period 1995–2000. Their results showed that the bias actually changed each year as track management practices changed, but that bettors appeared to be able to adopt efficient heuristics to cope with the sporadically changing information and learn to adapt their betting strategies to capture the evolving post-position bias information.

Taken together, the post-position bias studies appear to support the notion that profits can be earned at some racetracks, under certain conditions, for short periods. However, in the longer term, bettors appear to be able to adapt their betting strategies to fully discount the bias.

2.5. Pedigree

A horse's breeding is believed by many racing enthusiasts to have a significant effect on a horse's speed and stamina. In an efficient market, relevant information concerning a horse's pedigree should be fully discounted in odds. It is surprising, given the coverage given to the impact of breeding in the racing press, that few academic papers have

explored the degree to which markets incorporate this information. The one exception is a study by Bain et al. (2006) that examined the degree to which odds for the Kentucky Derby over the period 1946–2000 accounted for the Dosage index, a measure derived from a horse's pedigree to predict a horse's stamina (over its speed). They argued that Derby winners are likely to have a Dosage index of less than 4.0 (indicating a good stamina/speed ratio). Consequently, they developed a model to predict winning probabilities based on whether the Dosage index is less than or equal to 4.0 and whether or not a horse meets certain quality tests based on its previous performances. Profitable betting strategies using the model were derived for horses that met the Dosage and quality tests, suggesting that the betting public do not fully incorporate this information into final odds. This is particularly surprising since no horses that run in the Kentucky Derby would have already run at the distance (hence, reliance on information from pedigree might be expected) and much publicity is given to the race, exposing previous form and pedigree to considerable pre-race analysis (for further discussion of the market efficiency implications of the Dosage Index, see Gramm and Ziemba, 2008a, 2008b), who show that the effect is strongest in the 1½-mile Belmont Stakes.

2.6. Distance Preference

A number of existing studies implicitly examine the impact of distance preferences of each horse on the likely outcome of a race; most of these investigate this factor together with other fundamental explanatory variables (e.g., Benter, 1994; Chapman, 1994). However, Edeleman (2005) focuses specifically on this variable. He develops a regression model to determine the average velocity of a particular horse running in a race (as a proxy, to measure a horse's distance preference), based on a number of input variables, such as weight carried, and variables that measure the "ability at 1,400 m," distance gradient, and distance convexity. However, Edelman (2005) did not develop a forecasting model to predict winning probability of horses based on the average velocity (computed by the regression). Consequently, the predictive ability of this information and the degree to which it is discounted in odds remains to be resolved.

2.7. Single Variable Models: Overview

The main conclusion to emerge from single variable semi-strong form efficiency studies in horse race betting markets is that most of the variables explored contain valuable information. However, in most cases, bettors react to this information and discount it fully in the short run (newspaper reporters', professional tipsters', and betting systems' predictions), or in the longer run (e.g., post-position), or partially (e.g., cross-track betting and parallel market discrepancies, betting volume, breeding) into final odds. Even where the information is not fully incorporated, such as in cross-track betting and parallel market discrepancies, physical and cost barriers to arbitrage make the construction of profitable arbitrage strategies problematic. In addition, for some of the variables that are not fully discounted in final odds (e.g., betting volume, parallel market discrepancies) there is no published evidence for profitable betting strategies that are able to exploit these anomalies. Furthermore, one of the studies that does demonstrate that abnormal

returns can be earned by exploiting the failure of bettors to fully incorporate information (i.e., concerning a horse's pedigree), was developed for a specific race (the Kentucky Derby). Further out of sample testing and exploration of the degree to which breeding is discounted in other races also needs to be conducted before it can be claimed that horse race markets are inefficient with respect to a horse's pedigree.

It appears particularly difficult to identify semi-strong form inefficiency in pari-mutuel betting markets, since bets are settled at the final odds prevailing in the market. Consequently, even if the market reacts slowly to a particular piece of information (e.g., a professional tipster's selections), provided the market eventually incorporates the information, abnormal returns cannot be earned. However, in bookmaker markets, it is possible to secure the odds at the time the bet is struck and there is some evidence that profitable trading on some information (e.g., certain newspaper and professional tipsters' selections) may be possible if bets are made early in the market before other bettors have reacted to the information.

The failure of many single variable studies to provide convincing evidence for semi-strong form inefficiency at odds prevailing at market close suggests that horse race betting markets are good at discounting relatively *simple*, single variable information. The pedigree study of Bain et al. (2006) supports this conclusion, since the betting strategy employed in that paper relies on the *combination* of two disparate pieces of information (i.e., pedigree and previous performances) and in this sense is not really a single variable study of market efficiency. The element of *complexity*, therefore, appears to be a feature that can lead to semi-strong form inefficiency. This is, again, demonstrated by studies that show that discrepancies between parallel betting pools can be effectively exploited (e.g., Hausch et al., 1981) using betting strategies that require complex calculations in a short time interval.

3. SEMI-STRONG FORM EFFICIENCY IN HORSE RACE BETTING MARKETS: MULTIPLE VARIABLE MODELS

A number of studies have sought to test more directly the ability of horse race betting markets to deal with complexity. More specifically, these papers have explored the extent to which bettors simultaneously impound a range of information concerning odds, previous performances of horses and jockeys, handicappers' assessments, and conditions at the racetracks, into the final odds. These studies may have been stimulated by the success of the more complex single variable papers discussed in Section 2 and by studies such as Tuckwell (1983). This latter study demonstrated that win probabilities generated from Computercard, a model incorporating publicly available information, past and present, concerning the horses competing in a race, could be used to construct a profitable betting strategy. The manner in which Computercard combined the various pieces of information was not made clear in Tuckwell (1983). However, subsequent studies using multiple variables have made explicit the means by which the information is used to make probability predictions. These studies are classified here under the two broad modeling approaches they adopt, namely, those that make assumptions about

the nature of the underlying statistical distributions (distribution-based methods, Section 3.1), and those that do not (distribution-free methods, Section 3.2). Broad conclusions that emerge from both of these modeling approaches are explored in Section 3.3.

3.1. Distribution-Based Methods

The majority of studies in this category employ multinomial logit or probit functions to develop a stochastic utility model that can be used to judge the “winningness” value of a horse. They differ in terms of the assumed distribution for the error term in the model. The relative “winningness” values of horses in a given race are then used to predict each horse’s probability of success. These models generally incorporate a range of publicly available information concerning the horse, its jockey, race characteristics, and odds. One advantage of these approaches is that they, to some extent, enable the manner and extent of a variable’s contribution (in terms of the sign, size, and significance of the coefficient of the variable in the model and the nature of any interaction between variables) to the winning probability to be discerned.

Bolton and Chapman (1986) were the first to employ a multinomial logit model to predict winning probabilities in a horse race. Their model incorporated 10 independent variables associated with previous performances of the horse and the jockey, together with information concerning the current race, including the weight the horse carried, the post-position, and whether the horse was running over a new distance. Bolton and Chapman (1986) developed their model using a sample of only 200 races drawn from five racetracks in North America. However, they employed an explosion process, whereby they used the information concerning not only which horse won the race but, having eliminated the winner, which horse “won” the race to finish second and, having eliminated the first two finishers, which horse “won” the race to finish third. Consequently, the model was developed from an exploded dataset of 600 races. The signs of all the variables in the model were consistent with the authors’ a priori expectation. Horse-related variables appeared more important than jockey-related variables in determining winning probability and, among the horse-related variables, the average speed rating achieved in a horse’s last four runs appeared to contribute most to winning probability. Bolton and Chapman (1986) estimated the model using 150 races from their sample and tested whether profitable wagering strategies could be constructed using the predicted probabilities on a holdout sample of 50 races. Various wagering strategies were employed, including those that did (Isaacs, 1953), and did not (Rosner, 1975), account for the effect of the model’s suggested bets on the final odds. None of these produced a profit. However, using a strategy that involved differential bet sizes based on the attractiveness of the wager, constraining bets to a level that would only marginally affect final odds (a constrained form of the the Rosner, 1975, strategy) and eliminating bets on horses where the model produced poor estimates (i.e., on horses with a predicted winning probability of less than 0.07) positive returns were obtained. The small dataset employed does not allow firm conclusions to be drawn from these results but Bolton and Chapman (1986) suggested that they did point to the fact that the horse race betting market may not be semi-strong form efficient.

Chapman (1994) employed a similar methodology to extend Bolton and Chapman's (1986) earlier study. He developed a 20-variable multinomial logit model; 13 of these independent variables relating exclusively to the horse's past performances (both historical and recent performance variables), three relating to the horse's previous performances relative to runners in the current race or in the conditions of the current race, two relating to the weight carried in the current race, and two to the jockey's previous performances. The model is estimated using 2,000 races run in Hong Kong between 1985 and 1991. Nineteen of the 20 variables were significant at the 5% level and, again, the signs of all variables were in line with a priori expectations. To test the degree to which abnormal returns could be made using the model, Chapman (1994) constructed five holdout samples of 400 races and developed the model using the remaining 1,600 races. A strategy of betting a unit stake on the horse with the highest expected return in a given race produced average returns of 17.2%. Bolton and Chapman (1986) demonstrated that multinomial logit models are unreliable in predicting winning probabilities of extreme longshots. Consequently, Chapman (1994) explored modified strategies by eliminating longshots from consideration. When bets were restricted to horses with a predicted winning probability greater than 0.05, mean returns increased to 29.3%.

Numerous studies across the world have demonstrated that the public's subjective probabilities correspond well with realized winning probabilities (e.g., U.S.: Ali, 1977; Asch et al., 1984, 1986; McGlothlin, 1956; Snyder, 1978a; Thaler and Ziemba, 1988; Ziemba and Hausch, 1986; Australia: Bird and McRae, 1987; Tuckwell, 1983; New Zealand: van Zijl, 1984; Gander et al., 2001; UK: Bruce and Johnson, 2001; Dowie, 1976; Vaughan Williams and Paton, 1997). Consequently, Chapman (1994) extended the fundamental variable model to incorporate the log of the public's subjective probability of winning, revealed in final odds. His results showed that this variable accounted for 21.4% of the total explanatory power of the model. No tests were conducted on the profitability of trading on the revised model since Chapman (1994) argued that this would be difficult to execute in practice, as final odds were not known until the race was underway.

Benter (1994), based on his own practical experience of model building and using it to bet in Hong Kong, outlined the conditions necessary for developing a multinomial logit model (involving independent variables based on publicly available information) which can be used to develop a profitable betting strategy. Benter (1994) did not provide precise details of the fundamental variables employed in his model. However, his experience shows that, unlike the models developed by Bolton and Chapman (1986) and Chapman (1994), a successful model often includes variables relating to horses' or jockeys' previous performances for which it is difficult to develop a clear a priori expectation. In fact, he indicated that several variables in his successful model were highly correlated and several had effects that were counterintuitive. Benter (1994) combined the log of the public's subjective probability judgements (revealed as final odds) with the log of the probability estimates derived from the first stage multinomial logit fundamental handicapping model into a second stage multinomial logit model. This was used to predict winning probabilities. The model was developed using the 2,000 races run in Hong Kong employed in the Chapman (1994) study. A modified Kelly betting strategy, designed to maximize the exponential growth rate of wealth, was employed to

test the profitability of the model's predictions on a further 2,500 races over a five year period. Bets with a positive expectation in both win and exotic bet pools were identified, and bets were restricted to some fraction of the Kelly bet (e.g., 1/3 or 1/4). For the five year holdout sample, four years produced net profits and over the whole period wealth increased by a factor of approximately 50. Benter (1994) concluded that it is possible to employ a model based on publicly available information to make abnormal returns.

Sung and Johnson (2005) developed a comprehensive 70 variable multinomial logit model for predicting the results of flat races in the UK. This model includes variables associated with the horses', jockeys', and trainers' previous performances, a variable to capture the degree of betting by the holders of privileged information, variables to capture current race conditions, and a number of interaction terms. The model was estimated using the results of 16,836 races run across the UK between 1996 and 1999 and the model's predictions are tested on a holdout sample of 1,929 races run in 2000. A Kelly betting strategy yielded returns of 16.38% and Sung and Johnson (2005) concluded that the betting public failed to account for a significant amount of publicly available information; particularly that associated with complicated transformations of variables associated with previous performances of horses, jockeys, and trainers.

Most of the studies discussed above, which have employed multinomial logit models to investigate semi-strong form efficiency, have now been in the public domain for several years. Consequently, it might be expected that profitable trading based on these models may no longer be possible. However, most of these studies were conducted in pari-mutuel betting markets in North America and Hong Kong and Sung et al. (2005) argued that conditions in UK horse race betting are so different that it might still be possible to construct profitable betting strategies based on a multinomial logit model incorporating a number of independent variables. In particular, they argued that in the UK (unlike the U.S. or Hong Kong) each racetrack is idiosyncratic in terms of market size, ground conditions, configuration, and knowledge of bettors. Consequently, different variables may exert different influences on a horse's winning chance at different racetracks. As a result, the task of predicting winning probabilities may be particularly difficult in the UK and this may offer scope for the construction of profitable betting strategies on a specific racetrack-based stochastic utility model, parameterized in the form of a multinomial logit model. Sung et al. (2005) tested their hypothesis by developing a multinomial logit model for Goodwood racetrack, incorporating six variables associated with the horses' previous performances, five variables to capture the conditions of the current race, and two variables relating to the jockeys' previous performances. In addition, they incorporated, unlike in previous studies, three variables to capture information concerning missing data (e.g., if the horse has not run at least four times before). The model is estimated using a sample of 200 races run at Goodwood racetrack between 1995 and 1997, but this dataset is exploded using the process described in Chapman and Staelin (1982) and used in Bolton and Chapman (1986). Predicted probabilities from this fundamental model are then combined with the log of the normalized winning probability implied by final bookmaker odds in a second stage multinomial logit model. This model was estimated using a second set of 200 races

run at Goodwood between 1997 and 1999. It was found that the fundamental model probabilities did add important information concerning winning probabilities over and above that contained in odds. The resulting model was then used to predict winning probabilities for a further set of 156 races run at the same racetrack between 1999 and 2000. A Kelly wagering strategy based on these predictions resulted in wealth increasing by a factor of 1.36, which compared favorably with a loss of 26.6% resulting from placing a unit bet on each runner. Sung et al. (2005), therefore, concluded that the study supported the view that the horse race betting market in the UK remained semi-strong form inefficient with respect to a multivariable model. They speculated that this may relate to the difficulties faced by bettors in understanding the different factors relevant to success at each racecourse in the UK and to the information overload experienced by bettors facing a choice between bookmaker and pari-mutuel odds.

One objection to the use of a multinomial logit model for predicting winning probability is that the assumed error term distribution (i.e., negative double exponential distribution) may not be the best for modeling horse race data. It has been proposed by Henery (1981) that the normal distribution may be a better approximation to the error term distribution. This conclusion has been confirmed in explorations by Lo et al. (1995) and Ali (1998). Gu et al. (2003) compared the performance of multinomial logit and probit models (which assume a normally distributed error term) for predicting the results of horse races in Hong Kong, using data from 1,000 races involving 12,850 horses. The models each consisted of 15 independent variables including the log of the probability implied by the final odds, four variables to capture information concerning the horse's previous performance history (e.g., the number of wins in the last two years), four to account for current race conditions (e.g., the horse's current age, the weight carried by the horse), two to capture the jockey's previous performance history, and four interaction terms (e.g., the horse's age times the distance of the race). The models were developed using 600 races. The predicted probabilities from the two models are then tested on a holdout sample of 400 races using two betting strategies. In the first, a unit bet is made on the horse in each race with the highest expected return. In the second, a modified Kelly betting strategy (following Rosner, 1975; and Bolton and Chapman, 1986) was employed, where the wealth level is fixed at unit value when calculating bet size. Betting strategies one and two, based on both the multinomial logit and probit models, were profitable. However, the multinomial probit model was more profitable in both cases (28.3% better under the unit bet strategy and 16.1% better under the modified Kelly strategy).

While the models examined in this section employed a variety of variables to capture information concerning a horse's previous performances, they all failed to adequately quantify changes in the class or competitiveness of previous races run by each horse. Changes in class determine how difficult a race is to win, and without such a measure, a comparison of previous performances is vague, at best. Edelman (2003) develops a model to address this weakness. He first developed a model to relate the beaten lengths for a given horse moving from race i to race j to the change in weight carried and the change in class of the two races. A ridge regression is performed over the complete history of races for all horses in the sample in order to estimate the model parameters

and, subsequently, to determine the change in class from race to race. Edelman (2003) was then able to assess the change in class from race to race and to use this to construct a competitive form variable for each horse. He then developed three multinomial logit models to include, respectively: (a) the competitive form variable only, (b) the logarithm of the bookmakers' final odds, and (c) the competitive form variable and the logarithm of the bookmakers' final odds. Edelman (2003) estimated the parameters for these models using a sample of higher grade sprint races (1000–1200 m) run in Australia between 1991 and 1998. The findings indicated that the winning probability increased as the competitive form variable increased and the competitive form variable added significantly to winning probability when included alongside the log of odds. Edelman (2003) tested the effectiveness of the competitive form variable by developing a strategy of betting enough on all runners with a favorable value of the competitive form variable to return one unit. This produced a profit of 27% on horses with odds of 2/1 or shorter (21% on horses longer than 2/1), compared to a loss of 18% from wagering on horses without reference to the form variable. In conclusion, Edelman (2003) suggested (p. 112) that the model gave rise to “apparently reliable profitable betting strategies” and that these profits could almost certainly be improved significantly with the addition of other class ratings variables.

3.2. Distribution-Free Methods

The distribution-based approaches discussed above suffer the potential disadvantage that they make assumptions concerning the distribution of the error terms in the models. If these assumptions are incorrect, then poor predictions may result. In addition, the direct modeling of win-loss outcomes using multinomial logit or probit does not use all the information concerning a horse's relative finish position (only whether it finishes first or is unplaced; see below for further discussion). This failure may also damage the accuracy of the resulting predictions. Their predictive ability can also be reduced by their reliance on linear approximations (to a greater degree than some distribution-free approaches); this can particularly reduce the value of their winning probability predictions for longshots (Bolton and Chapman, 1986). In addition, distribution-based methods, such as multinomial logit, often require considerable data for accurate estimation of estimates; this is less of a problem for some distribution-free methods. Equally, the distribution-free methods discussed below make no distribution assumptions and find alternative means of combining information from multiple variables in order to predict the outcomes of horse races.

One of these was developed by White et al. (1992) and involved combining forecasts from a number of diverse procedures to predict the outcome of a horse race. They developed a scaling procedure whereby each horse in a given race was ranked on the basis of a given criterion relative to the horse with the highest rating. For example, one of the criteria employed was earnings per start. The horse with the highest earnings per start would be assigned a score of 100 and the horse with lowest earnings per start would be assigned a score of 0. Using these two values as the base, all intervening earnings per start values were assigned an appropriate scaled score between 0 and 100.

In this manner, White et al. (1992) produced scaled scores for each horse in the race on the basis of 12 diverse approaches for distinguishing the horses' ability; five of these related to the horses' class, three to the horses' speed, two to the horses' pace, and two were based on subjective opinions (i.e., a consensus of handicappers' opinions and the final odds). The various scaled values for each horse are then combined into one measure for each horse using a simple arithmetic average. White et al. (1992) identify disadvantages associated with the scaling and averaging process, namely, that information is lost when scaling, finishing positions used by some of the measures to distinguish horses' ability may only accurately determine the relative performance of the first four or five finishers (those finishing down the field may cease to compete well before the finish), and, finally, the different approaches used are unlikely to be independent and may to some extent measure the same aspect of a horses' ability. However they also point to three important advantages of this approach. First, scaling results in equal ranges for all the methods of distinguishing the horses' ability; this is important since the combined forecast includes subjective methods which often show lower variance than actual results imply. Second, the scaled scores take account of the competitive element of a horse race. Third, the approach enables diverse types of methods for assessing horses' relative ability to be combined. The accuracy of the combined measure of horses' relative ability was tested on 25 sprint races run at Calder racetrack in Florida during 1990. The accuracy increased as the number of methods of assessing a horse's ability increased. The best individual forecast was shown, as expected, to be the final odds and the speed and pace variables showed the lowest mean error. Using a Kelly betting strategy, the average return from applying the 12 different methods for assessing horses' ability in isolation was a loss of 7%, but when all the methods were combined this produced a profit of 26%. Clearly, the holdout sample was very small but the results do suggest that a distribution-free approach which combines information concerning a horse's prospects from a variety of sources may enable profitable trading strategies to be developed.

The approach adopted by White et al. (1992) provides a simple distribution-free means of combining information from various sources in order to make horse race predictions, but it does not incorporate subtle interactions between the various data sources. Edelman (2007) sets out to develop a distribution-free approach that fills this gap. He develops a support vector machine regression (SVR) model to generate a continuous index value (a winningness index) for each horse based on a range of 12 input variables, eight of these being based on the performance of the horse in its previous race, and four on the conditions of the current race. This model is employed to predict the normalized finish position variable used by Benter (1994; i.e., horses' finishing position: first = 1, second = 2, etc., scaled to an interval $[-0.5, 0.5]$). This approach overcomes the drastic information loss suffered by approaches that directly model win-loss outcomes (as used, e.g., in Bolton and Chapman, 1986) since the relative performance of horses who do not win the race are ignored in these latter approaches. The winningness index developed from the regression of the fundamental values is then combined with information contained in final odds, using a standard multinomial logit model, to produce predicted probabilities of winning. Edelman (2007) developed his model using a sample of 200

races run in Australia in 1995 and tested the results on a holdout sample of 100 races. One of the practical drawbacks of this approach, compared to modeling probabilities directly using multinomial logit or probit, is that the contribution of the included independent variables cannot be measured. However, a Kelly betting strategy applied to the predicted probabilities resulted in an increase in wealth of 500%. However, it is argued that modeling the later finishing positions may introduce noise information (e.g., Bolton and Chapman, 1986; Sung and Johnson, 2007). This view challenges the reliability of SVR models. Consequently, Lessmann et al. (2007) adopt a similar methodology to Edelman (2007) but replace SVR with support vector machine classification (SVC). SVC in its simple form is a supervised learning method for distinguishing between winners and losers that maintains the merits of SVR; namely, those associated with overcoming most of the disadvantages of distribution-based approaches, discussed in the beginning of this section. Indeed, using a sample of 400 races run at Goodwood in the UK, Lessmann et al. (2007) develop SVR- and SVC-based models. They then use these two models to estimate winning probabilities for runners in a 156-race holdout sample. The returns produced by applying a Kelly wagering strategy using these estimated probabilities from the two models are compared. The results indicate that positive returns can be obtained using both SVR- and SVC-based models, but the profit derived using the SVC-based model estimates is 1.66 times higher than that using the SVR-based model estimates. This finding confirms the view that the noisy information contained in the minor placings can damage the predictive ability of a horse race model. The holdout samples in Edelman (2007) and in Lessmann et al. (2007) are small, but the profits achieved are strongly suggestive of semi-strong form inefficiency. The variables they employ either only account for information concerning a horse's *previous* run or they have been in the public domain for many years. Consequently, it is expected that other more sophisticated variables, that previous research has shown to be valuable for predicting winning probability, may well not be fully discounted in odds (e.g., a horse's lifetime earnings, a jockey's previous performances). The findings of Edelman (2007) and of Lessmann et al. (2007) also point to the existence of non-linear relationships among a range of explanatory variables, and this may help to explain the persistence of semi-strong form inefficiency in horse race betting markets.

3.3. Multiple Variable Models: Overview

The clear conclusion to emerge from the studies, which combine publicly available information concerning various aspects of a horse's and a jockey's previous performances with information associated with current race conditions, is that betting markets do not fully account for this information in final odds. In addition, it appears possible to develop profitable trading strategies based on these models' predictions. The results of the earliest multivariable models have been in the public domain for many years and most of the variables employed in these models are those that bettors would consider important in making their selections. Despite these observations, more recent attempts to model winning probabilities using multivariable approaches have also demonstrated profitable trading opportunities. It appears that the betting public is unable to handle

the complexity of weighting the various criteria appropriately (which the models are designed to do) when making their judgments concerning horses' chance of winning. More recent advances, such as approaches employing SVRs, capture "subtle non-linear and interactive effects of the input variables" (Edelman, 2007, p. 335) and these are the types of subtleties that are the least likely to be discovered and acted upon by the betting public.

4. SEMI-STRONG FORM EFFICIENCY IN HORSE RACE BETTING MARKETS: CONCLUSION

The broad conclusion to emerge from the studies examined in this chapter is that horse race betting markets incorporate a significant amount of publicly available information. Few studies examining the impact of a *single* type of publicly available information have found that abnormal returns can be earned from trading strategies based on information associated with these variables. While some studies have demonstrated that profitable trading is possible for short periods (even up to one year), evidence suggests that the information is discounted in odds in the longer run and the profits disappear. In this sense, the efficient market hypothesis cannot be rejected. In fact, the ability of bettors to utilize publicly available information has been well documented (e.g., Johnson and Bruce, 2001) and there are a number of reasons, based on previous research findings, to expect horse race bettors' decisions to produce odds that calibrate well with observed race outcomes. These include the long-term experience of bettors (see, e.g., Smith and Kida, 1991, on the role of experience), which means that many bettors are well-practiced in making probability judgments (e.g., see Ferrell, 1994, for the benefits of such practice), the regular and unequivocal feedback (i.e., their selection wins or loses) that they receive (e.g., McClelland and Bolger, 1994, discuss how this can improve judgments), they are well-motivated (monetary gains/losses) to make accurate judgment (e.g., Ashton, 1992, provides evidence of improved judgment where appropriate incentives are provided), and, finally, their judgments are made in a real world environment (cf. the laboratory, where calibration is often inferior; see, e.g., Gigerenzer et al., 1991). The absence of long-run abnormal returns from betting strategies based on a single variable derived from publicly available information suggests that the horse race betting market is efficient in its use of such information. However, efficiency can be defined in a more restrictive manner to require "that expected returns be equal across wagering prospects" on identical outcomes, where costs and risks are equal (Sauer, 1996, p. 2,024). If this definition is used, then a different picture emerges. The single variable studies suggest that there are a number of situations (e.g., cross-track betting, parallel bookmaker and pari-mutuel markets, parallel win and place/show markets) where expected returns are not equal. In part, this may result from the irrational behavior of horse race bettors, which has been observed in a number of studies (e.g., Johnson et al., 1999; Bruce and Johnson, 1995; Metzger, 1985). However, in some cases, cost and physical barriers exist to prevent arbitrage, and in others, it has been shown that the representative bettor must be risk neutral for the observed differences to represent a

sign of inefficiency. The latter assumption is clearly suspect, given bettors' motivation to continue to bet despite positive takeout rates. In most cases, therefore, the *simple*, single variable models fail to provide evidence for semi-strong form inefficiency.

The single variable studies that contradict this conclusion have either been conducted in the bookmaker market (where returns to information can be secured by accepting odds early in the market before other market participants react), or involve more *complex* strategies (e.g., differences between win and place/show pools exploited using complex calculations in a short time period). The results of the latter studies suggest the hypothesis that horse race betting markets are good at discounting simple, single variable information, but that they do not efficiently discount more complex information. This view is largely confirmed by the multivariable studies discussed above. In all cases, these studies show that profitable trading strategies can be developed using predictions from models that examine the simultaneous effect of several variables on race outcome. A striking feature is the robustness of these results through time. This may result from bettors' inability to perceive and react to the subtle relationships that exist between variables (e.g., Bruce and Johnson, 1996, 1997; Johnson and Bruce, 1997, 1998), and/or that these relationships change through time (Johnson et al., 2005).

Before dismissing the abilities of horse race bettors to discount publicly available information, it should be noted that odds remain the most successful single predictor of race outcome, even in comprehensive multivariable models. These results suggest that bettors do discount a significant amount of publicly available information. In addition, they appear to be particularly successful in discounting relatively simple, single variable information. In conclusion, the semi-strong form inefficiency that is observed in horse race betting markets appears to arise from bettors' failure to adequately account for the complex, subtle (or non-linear), and possibly changing relationships between the variables that affect race outcome.

References

- Ali, M. 1977. Probability and Utility Estimates for Racetrack Bettors, *Journal of Political Economy* 85, 803–815.
- Ali, M. 1979. Some Evidence of the Efficiency of a Speculative Market, *Econometrica* 47, 387–392.
- Ali, M. 1998. Probability Models on Horse Race Outcomes, *Journal of Applied Statistics* 25, 221–229.
- Anderson, D., R. Clarke, and P. Ziegler. 1985. Information, Equilibrium, Efficiency in Betting Markets. Working Paper, University of Queensland.
- Asch, P., B. G. Malkiel, and R. E. Quandt. 1984. Market Efficiency in Racetrack Betting, *The Journal of Business* 57, 165–175.
- Asch, P., B. G. Malkiel, and R. E. Quandt. 1986. Market Efficiency in Racetrack Betting: Further Evidence and a Correction, *The Journal of Business* 59, 157–160.
- Asch, P., and R. E. Quandt. 1987. Efficiency and Profitability in Exotic Bets, *Economica* 59, 278–298.
- Ashton, R. H. 1992. Effects of Justification and a Mechanical Aid on Judgment Performance, *Organizational Behavior and Human Decision Processes* 52, 292–306.
- Bain, R. S., D. B. Hausch, and W. T. Ziemba. 2006. An Application of Expert Information to Win Betting on the Kentucky Derby, *European Journal of Finance* 1, 283–301.
- Benter, W. 1994. Computer Based Horse Race Handicapping and Wagering Systems: A Report, in D. B. Hausch, V. S. Y. Lo, and W. T. Ziemba (eds.), *Efficiency of Racetrack Betting Markets*. Academic Press, San Diego, CA, pp. 183–198.

- Betton, S. 1994. Post-Position Bias: An Econometric Analysis of the 1987 Season at Exhibition Park, in D. B. Hausch, V. S. Y. Lo, and W. T. Ziemba (eds.), *Efficiency of Racetrack Betting Markets*. Academic Press, San Diego, CA, pp. 511–526.
- Beyer, A. 1983. *The Winning Horseplayer*. Houghton Mifflin, Boston.
- Bird, R., and M. McCrae. 1987. Tests of the Efficiency of Racetrack Betting Using Bookmaker Odds, *Management Science* 33, 1552–1562.
- Bird, R., and M. McCrae. 1994. Efficiency of Racetrack Betting Markets: Australian Evidence, in D. B. Hausch, V. S. Y. Lo, and W. T. Ziemba (eds.), *Efficiency of Racetrack Betting Markets*, Academic Press, San Diego, CA, pp. 575–582.
- Blackburn, P., and J. Peirson. 1995. Betting at British Racecourses: An Analysis of Semi-Strong Efficiency Between Bookmaker and Tote Odds. Discussion Paper, University of Kent at Canterbury, Department of Economics, 95/4.
- Bolton, R. N., and R. G. Chapman. 1986. Searching for Positive Returns at the Track: A Multinomial Logit Model for Handicapping Horse Races, *Management Science* 32, 1040–1060.
- Bruce, A. C., and J. E. V. Johnson. 1995. Costing Excitement in Leisure Betting, *Leisure Studies* 14, 48–63.
- Bruce, A. C., and J. E. V. Johnson. 1996. Decision-Making Under Risk: Effect of Complexity on Performance, *Psychological Reports* 79, 67–76.
- Bruce, A. C., and J. E. V. Johnson. 1997. The Complex Decision: Insights from Naturalistic Research, *The International Journal of Project and Business Risk Management* 1, 403–415.
- Bruce, A. C., and J. E. V. Johnson. 2000a. Investigating the Roots of the Favourite-Longshot Bias: An Analysis of Decision Making by Supply and Demand-Side Agents in Parallel Betting Markets, *The Journal of Behavioral Decision Making* 13, 413–430.
- Bruce, A. C., and J. E. V. Johnson. 2000b. Market Efficiency Analysis Requires a Sensitivity to Market Characteristics: Some Observations on a Recent Study of Betting Market Efficiency, *Applied Economics Letters* 7, 199–202.
- Bruce, A. C., and J. E. V. Johnson. 2001. Efficiency Characteristics of a Market for State Contingent Claims, *Applied Economics* 33, 1751–1754.
- Bruce, A. C., and J. E. V. Johnson. 2003. The Role of Trading Volume in Explaining Horse Race Outcomes: A Technical Analysis Perspective, in *The Twelfth International Conference on Gambling and Risk Taking*. Available at <http://eprints.soton.ac.uk/36439>.
- Bruce A. C., and J. E. V. Johnson. 2003. Investigating Betting Behaviour: A Critical Discussion of Alternative Methodological Approaches, in L. Vaughan Williams (ed.), *The Economics of Gambling*. Routledge, London, 224–246.
- Bruce A. C., and J. E. V. Johnson. 2005. Market Ecology and Decision Behaviour in State-Contingent Claims Markets, *Journal of Economic Behavior and Organization* 56, 199–217.
- Busche, K., and D. W. Walls. 2000. Decision Costs and Betting Market Efficiency, *Rationality and Society* 12, 477–492.
- Cain, M., D. Law, and D. A. Peel. 2001. The Incidence of Insider Trading in Betting Markets and the Gabriel and Marsden Anomaly, *The Manchester School* 69, 197–207.
- Cain, M., D. Law, and D. A. Peel. 2003. The Favourite-Longshot Bias and the Gabriel and Marsden Anomaly: An Explanation Based on Utility Theory, in L. Vaughan Williams (ed.), *The Economics of Gambling*. Routledge, London, pp. 2–13.
- Camerer, C. F. 1998. Can Asset Markets Be Manipulated? A Field Experiment with Racetrack Betting, *The Journal of Political Economy* 106, 457–482.
- Canfield, B. R., B. C. Fauman, and W. T. Ziemba. 1987. Efficient Market Adjustment of Odds Prices to Reflect Track Biases, *Management Science* 33, 1428–1439.
- Chapman, R. G. 1994. Still Searching for Positive Returns at the Track: Empirical Results From 2,000 Hong Kong Races, in D. B. Hausch, V. S. Y. Lo, and W. T. Ziemba (eds.), *Efficiency of Racetrack Betting Markets*. Academic Press, San Diego, CA, pp. 173–181.
- Chapman, R. G., and R. Staelin. 1982. Exploiting Rank Ordered Choice Set Data Within the Stochastic Utility Model, *Journal of Marketing Research* 19, 288–301.
- Cotton. M. 1990. *Value Betting*. Aesculus Press, Oswesry, UK.

- Crafts, N. F. R. 1994. Winning Systems? Some Further Evidence on Insiders and Outsiders in British Horse Race Betting, in D. B. Hausch, V. S. Y. Lo, and W. T. Ziemba (eds.), *Efficiency of Racetrack Betting Markets*. Academic Press, San Diego, CA, pp. 545–549.
- Deschamps, B., and O. Gergaud. 2008. Efficiency in Horse Race Betting Markets: The Role of the Professional Tipsters, in D. B. Hausch and W. T. Ziemba (eds.), *Handbook of Sports and Lottery Markets*. North-Holland, Amsterdam, 341–354.
- Dowie, J. 1976. On the Efficiency and Equity of Betting Markets. *Economica* 43, 139–150.
- Edelman, D. C. 2003. A Competitive Horse-Race Handicapping Algorithm Based on Analysis of Covariance, in L. Vaughan Williams (ed.), *The Economics of Gambling*, Routledge, London, pp. 106–113.
- Edelman, D. C. 2005. Modelling Distance Preference in Flat Racing via Average Velocity, in L. Vaughan Williams (ed.), *Information Efficiency in Financial and Betting Markets*. Cambridge University Press, Cambridge, UK, pp. 339–345.
- Edelman, D. C. 2007. Adapting Support Vector Machine Methods for Horse Race Odds Prediction. *Annals of Operations Research* 151(1), 325–336.
- Edelman, D. C., and N. R. O'Brian. 2004. Tote Arbitrage and Lock Opportunities in Racetrack Betting, *The European Journal of Finance* 10, 370–378.
- Ferrell, W. R. 1994. Discrete Subjective Probabilities and Decision Analysis: Elicitation, Calibration and Combination, in G. Wright and P. Ayton (eds.), *Subjective Probability*. Wiley, Chichester, UK, pp. 410–451.
- Figlewski, S. 1979. Subjective Information and Market Efficiency in a Betting Market, *Journal of Political Economy* 87, 75–89.
- Gabriel, P. E., and J. R. Marsden. 1990. An Examination of Market Efficiency in British Racetrack Betting, *Journal of Political Economy* 98, 874–885.
- Gabriel, P. E., and J. R. Marsden. 1991. An Examination of Efficiency in British Racetrack Betting: Errata and Corrections, *Journal of Political Economy* 99, 657–659.
- Gander, J. M., R. A. Zuber, and R. S. Johnson. 2001. Searching for the Favorite-Longshot Bias Down Under: An Examination of the New Zealand Pari-Mutuel Betting Market, *Applied Economics* 33, 1621–1629.
- Gramm, M., and W. T. Ziemba. 2008a. The Dosage Breeding Theory for Horserace Predictions, in D. B. Hausch and W. T. Ziemba (eds.), *Handbook of Sports and Lottery Markets*. North-Holland, Amsterdam, 307–340.
- Gramm, M., and W. T. Ziemba. 2008b. The Dosage Anomaly for the Triple Crown Races. *Mimeo*.
- Gigerenzer, G., U. Hoffrage, and H. Kleinbolting. 1991. Probabilistic Mental Models: A Brunswikian Theory of Confidence, *Psychological Review* 98, 506–528.
- Gu, M. G., C. Huang, and W. Benter. 2003. Multinomial Probit Models for Competitive Horse Racing, Working Paper, Chinese University of Hong Kong.
- Hausch, D. B., and W. T. Ziemba. 1990. Arbitrage Strategies for Cross-Track Betting on Major Horse Races, *The Journal of Business* 63, 61–78.
- Hausch, D. B., and W. T. Ziemba. 1991. Locks at the Racetrack, *Interfaces* 20, 41–48.
- Hausch, D. B., W. T. Ziemba, and M. Rubinstein. 1981. Efficiency of the Market for Racetrack Betting, *Management Science* 27, 1435–1452.
- Henry, R. J. 1981. Permutation Probabilities as Models for Horse Races, *Journal of the Royal Statistical Society* 43, 86–91.
- Hong, Y., and C. Chiu. 1988. Sex, Locus of Control and Illusion of Control in Hong Kong as Correlates of Gambling Involvement, *The Journal of Social Psychology* 128, 667–673.
- Isaacs, R. 1953. Optimal Horse Race Bets, *American Mathematical Monthly* 60, 310–315.
- Jackson, D., and P. Waldron. Pari-Mutuel Place Betting in Great Britain and Ireland: An Extraordinary Opportunity, in L. Vaughan Williams (ed.), *The Economics of Gambling*. Routledge, London, pp. 18–29.
- Johnson, J. E. V., and A. C. Bruce. 1997. A Probit Model for Estimating the Effect of Complexity on Risk Taking, *Psychological Reports* 80, 763–772.
- Johnson, J. E. V., and A. C. Bruce. 1998. Risk Strategy Under Task Complexity: A Multivariate Analysis of Behaviour in a Naturalistic Setting, *The Journal of Behavioral Decision Making* 11, 1–18.

- Johnson, J. E. V., R. O'Brien, and H. S. Shin. 1999. A Violation of Dominance and the Consumption Value of Gambling, *The Journal of Behavioral Decision-Making* 12, 19–36.
- Johnson, J. E. V., and A. C. Bruce. 2001. Calibration of Subjective Probability Judgements in a Naturalistic Setting, *Journal of Organizational Behavior and Human Decision Processes* 85, 265–290.
- Johnson, J. E. V., R. O'Brien, and J-F. Bozzetto. 2005. Toward an Understanding of Investors' Reactions to Changes in Information Affecting Market Outcomes: Impact of Post-Position Bias Changes on Betting Markets. In *The Sixth European Conference on Gambling Studies and Policy Issues*.
- Kelly, J. L. 1956. A New Interpretation of Information Rate, *The Bell System Technical Journal* 35, 917–926.
- Law D., and D. A. Peel. 2002. Insider Trading, Herding Behaviour and Market Plungers in the British Horse Race Betting Market, *Economica* 69, 327–338.
- Leong, S. M., and K. G. Lim. 1994. Cross-Track Betting: Is the Grass Greener on the Other Side? in D. B. Hausch, V. S. Y. Lo, and W. T. Ziemba (eds.), *Efficiency of Racetrack Betting Markets*. Academic Press, San Diego, CA, pp. 617–629.
- Lessmann, S., M. Sung, and J. E. V. Johnson. 2007. A New Method for Predicting the Outcome of Speculative Events. Discussion Papers in Centre for Risk Research, CRR-07-03. University of Southampton, Southampton, UK.
- Lo, V. S. Y., J. Bacon-Shone, and K. Busche. 1995. The Application of Ranking Probability Models to Racetrack Betting, *Management Science* 41, 1048–1059.
- Lo, V. S. Y., and K. Busche. 1994. How Accurately Do Bettors Bet in Doubles? in D. B. Hausch, V. S. Y. Lo, and W. T. Ziemba (eds.), *Efficiency of Racetrack Betting Markets*. Academic Press, San Diego, CA, pp. 465–468.
- Losey, R. L., and J. C. Talbott. 1980. Back on the Track with the Efficient Markets Hypothesis, *The Journal of Finance* 35, 1039–1043.
- McCulloch, B., and A. Van Zijl. 1986. Direct Test of Harville's Multi-Entry Competitions Model on Racetrack Betting Data, *Journal of Applied Statistics* May, 213–219.
- McClelland, A. G. R., and F. Bolger. 1994. The Calibration of Subjective Probabilities: Theories and Models 1980–1994, in G. Wright and P. Ayton (eds.), *Subjective Probability*. Wiley, Chichester, UK, pp. 453–482.
- McGlothlin, W. H. 1956. Stability of Choices Among Uncertain Alternatives, *American Journal of Psychology* 69, 604–619.
- Metzger, M. A. 1985. Biases in Betting: An Application to Laboratory Findings, *Psychological Reports* 56, 883–888.
- Peirson, J., and P. Blackburn. 2003. Betting at British Racecourses—A Comparison of the Efficiency of Betting with Bookmakers and at the Tote, in L. Vaughan Williams (eds.), *The Economics of Gambling*. Routledge, London, pp. 30–42.
- Quirin, W. L. 1979. *Winning at the Races: Computer Discoveries in Thoroughbred Handicapping*. William Morrow, New York.
- Roberts, P. M., and B. A. Newton. 1987. *The Intelligent Punter's Survey*. TIPS Publishing, Weymouth, UK.
- Rosner, B. 1975. Optimal Allocation of Resources in a Pari-Mutuel Setting, *Management Science* 21, 9, 997–1006.
- Sauer, R. D. 1998. The Economics of Wagering Markets, *Journal of Economic Literature* 36, 2021–2064.
- Shin, H. S. 1992. Prices of State Contingent Claims with Insider Traders, and the Favourite-Longshot Bias, *The Economic Journal* 102, 426–435.
- Shin, H. S. 1993. Measuring the Incidence of Insider Trading in a Market for State-Contingent, *The Economic Journal* 103, 1141–1153.
- Smith, M. A. 2003. The Impact of Tipster Information on Bookmakers' Prices in UK Horse-Race Markets, in L. Vaughan Williams (eds.), *The Economics of Gambling*, Routledge, London, pp. 67–79.
- Smith, J. F., and T. Kida. 1991. Heuristics and Biases: Expertise and Task Realism in Auditing, *Psychological Bulletin* 109, 472–485.
- Snyder, W. W. 1978a. Horse Racing: Testing the Efficient Markets Model, *The Journal of Finance* 33, 1109–1018.
- Snyder, W. W. 1978b. Decision-Making with Risk and Uncertainty: The Case of Horse Racing, *American Journal of Psychology* 91, 201–209.

- Sung, M., J. E. V. Johnson, and A. C. Bruce. 2005. Searching for Semi-Strong Form Efficiency in British Racetrack Betting Markets, in L. Vaughan Williams (eds.), *Information Efficiency in Financial and Betting Markets*. Cambridge University Press, Cambridge, UK, pp. 179–192.
- Sung, M., and J. E. V. Johnson. 2005. Searching for Information Inefficiency: Empirical Evidence from the UK Bookmaker Betting Market. In *The Fourth International Equine Industry Program Academic Conference*.
- Swidler, S., and R. Shaw. 1995. Racetrack Wagering and the “Uninformed” Bettor: A Study of Market Efficiency, *The Quarterly Review of Economics and Finance* 35, 305–314.
- Thaler, R., and W. T. Ziemba. 1988. Pari-Mutuel Betting Markets: Racetracks and Lotteries, *Journal of Economic Perspectives* 2, 161–174.
- Tuckwell, R. 1981. Anomalies in the Gambling Market, *Australian Journal of Statistics* 23, 287–295.
- Tuckwell, R. 1983. The Thoroughbred Gambling Market: Efficiency, Equity and Related Issues. Australian Economic Papers, June, pp. 106–118.
- Van Zijl, A. 1984. Returns and Weak Form Efficiency: Betting Markets. Working paper, Victoria University of Wellington.
- Vaughan Williams, L. 2000. Can Forecasters Forecast Successfully? Evidence from UK Betting Markets, *Journal of Forecasting* 19, 505–513.
- Vaughan Williams, L., and D. Paton. 1997. Does Information Efficiency Require a Perception of Information Inefficiency? *Applied Economics Letters* 4, 615–617.
- Vergin, R. C. 1977. An Investigation of Decision Rules for Thoroughbred Racehorse Wagering, *Interfaces* 8, 34–45.
- Walls, W. D., and K. Busche. 2003. Breakage, Turnover, and Betting Market Efficiency—New Evidence from Japanese Horse Tracks, in L. Vaughan Williams (eds.), *The Economics of Gambling*. Routledge, London, pp. 43–66.
- White, E. M., R. Dattero, and B. Flores. 1992. Combining Vector Forecasts to Predict Thoroughbred Horse Race Outcomes, *International Journal of Forecasting* 8, 595–611.
- Ziemba, W. T., and D. B. Hausch. 1984. *Beat the Racetrack*. Harcourt Bruce Jovanovich, San Diego, CA.
- Ziemba, W. T., and D. B. Hausch. 1986. *Betting at the Racetrack*, Dr Z Investments Ltd., San Louis Obispo, CA.
- Ziemba, W. T., and D. B. Hausch. 1987. *Dr Z's Beat the Racetrack*, rev. ed., William Morrow, New York.
- Ziemba, W. T., and D. B. Hausch. 1994. The Dr. Z Betting System in England, in D. B. Hausch, V. S. Y. Lo., and W. T. Ziemba (eds.), *Efficiency of Racetrack Betting Markets*, Academic Press, San Diego, CA, pp. 567–574.

CHAPTER 15

The Dosage Breeding Theory for Horse Racing Predictions

Marshall Gramm

Department of Economics and Business, Rhodes College, Memphis, TN, USA.

William T. Ziemba

Sauder School of Business, UBC, Vancouver, Canada, Mathematical Institute, Oxford University,
and ICMA Centre, University of Reading, UK.

1. Introduction	308
2. The Racetrack as a Sequence of Markets	310
3. The Dosage Index and Performance Measures	311
4. Data Acquisition	314
5. Application of Breeding Information and Performance Measures to Refine Estimated Win Probabilities for the Kentucky Derby	314
6. The Kelly Betting Model	320
7. The Kentucky Derby, 1981–2006	321
8. The Preakness Stakes, 1946–2006	327
9. The Belmont Stakes, 1946–2006	329
10. Conclusions	330
<i>References</i>	331
Appendix A: Data Sources	333
Appendix B: Kentucky Derby, Preakness, and Belmont Winners, 1946–2006	334

Acknowledgments

This chapter, which surveys part of the dosage literature, relies heavily on Bain et al. (2006) and Gramm and Ziemba (2008). Thanks go to Donald Hausch for useful comments on this chapter plus a longstanding relationship with the second author and to Rod Bain for his work with Hausch and Ziemba on their 2006 paper.

Abstract

This chapter surveys the dosage breeding theory pioneered by Vuilliers, Varola, and Roman with major emphasis on two top classic three-year-old thoroughbred races, namely, the Kentucky Derby and the Belmont Stakes. Run at $1\frac{1}{4}$ mi and $1\frac{1}{2}$ mi respectively, they typically are at least $\frac{1}{8}$ and $\frac{1}{4}$ mi longer than any of the horses has ever raced before. This extra distance, combined with the large fields (especially in the Derby), make these two races a difficult test of stamina for horses this young. Bettors are also challenged because there is no direct evidence of whether a horse has the stamina to compete effectively at these distances. The informational content of the publicly available, pedigree-based measure of stamina, the Dosage Index, is used with simple performance measures to identify a semi-strong-form inefficiency. Statistically significant profits, net of transaction costs, could have been achieved during 1946–2006. This can be compared to the middle leg of the Triple Crown, the Preakness, run at $\frac{13}{16}$, where the Dosage Index provided no advantage.

JEL Classifications: G10, G14

Keywords: semi-strong-form market efficiency, capital growth theory, speculative investments, sports betting

1. INTRODUCTION

The Kentucky Derby annually gathers many of the top three-year-old thoroughbred horses at Churchill Downs in Louisville, Kentucky on the first Saturday in May. For the horses entered, the race is a new challenge, since its distance of $1\frac{1}{4}$ mi is typically at least $\frac{1}{8}$ mi longer than any of them has ever raced. The extra distance of the Kentucky Derby, usually combined with a large field that includes many top-flight contenders, presents a significant test of stamina for these young horses. Two weeks after the Kentucky Derby, many of the same horses plus others compete in the Preakness Stakes run at the shorter distance of $\frac{13}{16}$ mi at Pimlico racetrack in Baltimore, Maryland. Then three weeks later, in early June, the $1\frac{1}{2}$ mi Belmont stakes is held at Belmont Park on Long Island, near New York City. And like the Kentucky Derby, typically none of the Belmont Stakes entrants has run this far this early in their careers.

Since the horses in the Kentucky Derby and the Belmont are running at longer distances than any earlier races, the public's assessment of their stamina cannot be easily based on their past performances. Without direct evidence of a horse's ability to run at the distances of these two races, bettors have looked to indirect evidence. One approach to assessing stamina that has received wide public attention looks at whether the sires in a horse's pedigree have demonstrated a pattern of progeny with stamina. This approach, called dosage theory and described in Section 3, is coupled with evidence of success in major races as a two-year-old horse to study semi-strong-form efficiency of the Kentucky Derby by Bain, Hausch, and Ziemba (Bain et al., 2006; hereafter BHZ), and of the Belmont Stakes by Gramm and Ziemba (2007; hereafter GZ).

We describe the empirical analyses in BHZ and GZ, and summarize their results that dosage theory and success as a two-year-old horse point to semi-strong-form inefficiency. Two categories of horses are studied: dual qualifiers and asterisk qualifiers.

Dual qualifiers are those with a speed/stamina balance consistent with the Derby and Belmont characteristics and who were strong enough as two-year-olds to be ranked within 10lb of the top two-year-old. Asterisk qualifiers pass the dosage test but not the 10lb test. However, they showed strength by winning a major race early in their three-year-old year.

BHZ and GZ show that both dual qualifiers and asterisk qualifiers did in fact provide positive wagering profits in the Derby and even more so in the Belmont. Since the dosage theory is silent on the Preakness, these strong positive results do not apply there. The Kentucky Derby results are strongest for the pre-1996 period while the Belmont results are strong through 2006. There were no dual or asterisk qualifiers in the 2007 Belmont.

Roberts (1967) defined a market as being *weak-form*, *semi-strong-form*, or *strong-form* efficient if it is not possible to devise a profitable investment scheme net of transactions costs based on prices (or, for the racetrack, publicly available odds), based on all publicly available information, or based on all information, respectively. For traditional financial markets, there is considerable evidence that points to weak-form and semi-strong-form efficiency, but little evidence for strong-form efficiency (see Fama, 1970, 1991; and Keim and Ziemba, 2000, for surveys).

Weak-form efficiency of the racetrack's win market means that betting systems based solely on the public's win odds, established through pari-mutuel betting, are not profitable. Evidence from many tracks over many years has pointed to weak-form efficiency, for example, Ali (1977) and Asch et al. (1982).¹ Weak-form efficiency of the win-betting market is a consequence of four of its features. First, transaction costs are high, about 13–20%, depending on track location, so a bettor needs to be considerably more successful than the average bettor just to break even.² Second, while the challenge is substantial, the concept of the win bet is relatively simple. Thus, bettors have no confusion about their task. Third, many racetrack bettors approach their wagering very seriously and some are very sophisticated. Fourth, for this serious audience, there is usually an abundance of relevant information, including records of past performances and workouts for all the horses, breeding, earnings, jockey records, and so on.

¹An exception may be extreme favorites at odds of 3/10 or less, which have been shown historically to produce a small average profit; see Ziemba and Hausch (1986). However, data such as that shown in Ziemba (2008) and in Snowberg and Wolfers (2008) indicates that such positive profits currently do not exist without rebates. Inefficiencies in other more complex markets are more common; see Hausch et al. (1994) and other chapters in this volume for such evidence. Weak and semi-strong market efficiencies are discussed in the chapters by Hausch and Ziemba (2008) and by Johnson and Sung (2008), respectively, in this volume.

²Large bettors can reduce this take by betting at rebate sites that return a portion of the bet to make the actual take about 10%. We do not deal with such bettors here nor with those outside the U.S. who wager on Betfair or other betting exchanges against other bettors directly rather than in a pari-mutuel pool as discussed here.

For the Kentucky Derby and the Belmont Stakes, the first two of these criteria are satisfied for the win market. While the third criterion is met, the Kentucky Derby and Belmont Stakes also receive much more interest in North America from casual fans than other races. Because, typically, none of the Derby or Belmont Stakes entrants has raced at $1\frac{1}{4}$ mi and $1\frac{1}{2}$ mi, respectively, it can be argued that the fourth criterion is not fully met.

The objective in BHZ and GZ is to determine whether the informational content of the Dosage Index, a pedigree-based measure of stamina that is publicly available, in conjunction with simple performance measures, is captured in the pari-mutuel win odds and, if not, whether it can be used to develop a profitable betting scheme.

The operation of the racetrack market is discussed in the first section. Section 3 describes the Dosage Index and performance measures, and their application to the Kentucky Derby and Belmont Stakes. The data used by BHZ and GZ in their analyses are discussed in Section 4. Section 5 describes BHZ's scheme for estimating each betting interest's win probability based on the public's odds, the Dosage Index, and the performance measures, with an application of this technique to the Kentucky Derby. The Kelly capital growth betting model is described in Section 6. The results from BHZ's joint application of their probability estimation scheme and Kelly wagering for the Kentucky Derby appear in Section 7. Sections 8 and 9 discuss GZ's results for the Preakness and Belmont Stakes for 1946–2007. Conclusions appear in Section 10.

2. THE RACETRACK AS A SEQUENCE OF MARKETS

Prior to a race, bettors engage in markets that establish prices for the various betting opportunities for that race. Betting closes immediately before the race begins, and payouts are calculated immediately following the race. For win betting, there are N betting interests in a race. Let W_i be the total amount bet to win on betting interest $i = 1, \dots, N$.³ The win pool is

$$W = \sum_{i=1}^N W_i. \quad (1)$$

The track payback, Q (generally 0.80 to 0.87 for win bets), is the fraction of each dollar bet that is returned to the bettors. The commission, or track take, is $1 - Q$. If betting interest k wins the race, then win bets on betting interests $i \neq k$ return zero, while each dollar bet on betting interest k returns approximately QW/W_k . The actual profit per dollar is rounded down to the nearest nickel or dime (this is called breakage). Together the track take and breakage constitute the transaction costs.⁴

³Our notation deals with one race only. In Section 5, to deal with several races simultaneously, we add a superscript to our notation to identify the race number.

⁴For each track there is a minimum payout, usually 5%, that the track must return even if there are insufficient funds available in QW . If there is a rebate, then the track take is effectively reduced to about 10% so Q is about 0.90.

Typically, each horse in a race runs as a separate betting interest. However, two or more horses in a race that have common ownership typically run as a single betting interest known as an entry. In addition, in a race where there would be more betting interests than a preset maximum, the horses with the least-impressive credentials are grouped as a single betting interest known as the field. A bet to win on an entry or the field pays off if any member of that betting interest wins the race. Both have been common in the Kentucky Derby but not in the Belmont Stakes. However, the long-time restrictions in Kentucky changed in 2001, so there were no entries and no field in the Kentucky Derby from 2001 on.

3. THE DOSAGE INDEX AND PERFORMANCE MEASURES

The fact that usually no Derby (or Belmont) entrant has raced at $1\frac{1}{4}$ ($1\frac{1}{2}$) mi prior to the race has led to the search for relevant information from alternative sources, including the horse's pedigree. One method of evaluating a thoroughbred's pedigree, commonly known as dosage theory, has its roots in the work of French cavalry officer Lt. Col. Jean-Joseph Vuillier, who studied the pedigrees of exceptional thoroughbreds of the late nineteenth and early twentieth centuries; see Vuillier (1902, 1906, 1928). The concept of thoroughbred dosage evolved through Varola, who developed a patented classification of prominent stallions according to the type of offspring that they produced in a series of articles in *The British Racehorse*; see also Varola (1974, 1980).

Roman's (1981) modifications of Varola's work are known as dosage theory. His work was outlined in Leon Rasmussen's Bloodlines column in the *Daily Racing Form* beginning before the 1981 Kentucky Derby. One product of Roman's pedigree analysis is the Dosage Index (*DI*), which is based on the categorization of prominent stallions in terms of whether they consistently sire offspring with distance proficiencies that are incongruous with the dosage profiles of those offspring when that stallion is excluded. Classified stallions are called *chefs-de-race* (or simply *chefs*);⁵ see Ziemba and Hausch (1987) and Roman's Website (<http://www.chef-de-race.com>) for the rationale behind the selection of recent chefs, and Roman (2002).

There are five categories for chef classification in Roman's system: Brilliant, Intermediate, Classic, Solid, and Professional. The categorization is based on "where they [sires] must lie on the speed-stamina spectrum to bring the figures of their descendants back in line with those of horses in the general population exhibiting similar performance traits" (Roman, 2001). A chef can be placed in one or two categories. Each time a chef appears in a four-generation pedigree, points are awarded in the appropriate category. Points are assigned on a scale of 16 for the first-generation sire, eight for each second-generation sire, four for each third-generation sire, and two for each fourth-generation sire. Sires that are classified in two categories have their points split. After

⁵Mares are not included because they are considered to have too few offspring to identify distance proficiencies, while it is not unusual for a stallion to sire 100–200 offspring in a year.

the 15 sires have been assigned points, the total for each category is entered into the Dosage Index formula

$$DI = \frac{\text{Brilliant} + \text{Intermediate} + 1/2 \text{ Classic}}{\text{Solid} + \text{Professional} + 1/2 \text{ Classic}} \quad (2)$$

Horses with a high *DI* have a pedigree that is weighted towards Brilliant and Intermediate chiefs, that is, sires who tend to produce offspring with greater sprinting ability than their pedigrees would suggest if that sire were eliminated from the pedigree. Horses with a low *DI* are predicted to have stamina. Very seldom will a stakes-quality horse have no dosage points, though some have so few that the *DI* is unreliable. The pedigree and the dosage profile for 2005 Belmont Stakes winner Afleet Alex are shown in Table 1. Each pedigree shows the sire and mare for each horse for four generations. For example, Afleet Alex's sire and mare were Afleet and Nurvette, with their respective sires and dams shown directly to their right in the pedigree.

After the initial classification of chiefs in 1981, Roman found that no Kentucky Derby winner from 1940 to 1980 had a *DI* exceeding 4.0, despite about one in seven entrants having a *DI* that high.

The Dosage Index is not a direct measure of the quality of a horse. One quality measure is the experimental free handicap (EFH), an annual ranking of two-year-old thoroughbreds that raced in select races in the United States. (see

TABLE 1 Pedigree and Dosage Index Calculation for 2005 Belmont Stakes Winner Afleet Alex

Northern Afleet	Afleet	Mr. Prospector (B/C)	Raise a Native (B)
			Gold Digger
	Polite Lady		Venetian Jester
			Friendly Ways
Nurvette	Nureyev (C)		Northern Dancer (B/C)
			Special
	Stellarette		Tentam
			Square Angel
Maggy Hawk	Hawkster	Silver Hawk	Roberto (C)
			Gris Vitesse
	Strait Lane		Chieftain
			Level Sands
	Hawaii		Utrillo II
			Ethane
Qualique	Dorothy Gaylord		Sensitivo
			Gaylord's Touch

Generation	Sire	Brilliant	Intermediate	Classic	Solid	Professional
1	Northern Affect					
2	Affect					
	Hawkster					
3	Mr. Prospector	2		2		
	Nureyev			4		
	Silver Hawk					
	Hawaii					
4	Raise a Native	2				
	Venetian Jester					
	Northern Dancer	1		1		
	Tentam					
	Roberto			2		
	Chieftain					
	Utrillo II					
	Sensitivo					
	Total	5	0	9	0	0

NOTE: Dosage Index = $(5 + 0 + 9/2)/(0 + 0 + 9/2) = 2.11$.

<http://www.jockeyclub.com/experimental.asp>). Conducted since 1933 by the Jockey Club, the EFH assigns the top runners a figurative weight on a scale that usually has the two-year-old champion weighted at 126 lb.⁶ Exceptional horses have been weighted up to 130 lb. Other top horses are assigned lower weights based on perceived ability until a cutoff is reached at about 100 lb beyond which no more horses are classified. Usually there are 15 to 30 horses classified within 10 lb of the top-weighted horse. Roman (1981) observed that starting in 1972, most Kentucky Derby winners were rated within 10 lb of the top-weighted horse. This observation led to the designation “dual qualifier” for any horse that was weighted within 10 lb of the top-weighted horse on the EFH (indicating the quality of the horse) and had a *DI* less than or equal to 4.0.⁷

Professional handicapper James Quinn offered a second measure of quality to add late-developers to the list. He defined what we call an “asterisk qualifier” to be any horse that: (1) won at least one of a selection of premier races prior to the Kentucky Derby or Belmont Stakes; (2) had a *DI* less than or equal to 4.0; and (3) was not rated within 10 lb on the EFH. A horse is a “dual-or-asterisk qualifier” if it qualifies for one of these two categories.

⁶Ranking horses by weight is a familiar concept at the racetrack. In handicap races, the top horses carry greater weight (jockey + saddle + additional weights if necessary) than the less-qualified horses. Handicapping of this sort occurs only in select races and is intended to make the race more competitive.

⁷Some people expand the dual qualifier category to include any horse that is declared a champion in a country other than the U.S. and has a *DI* less than or equal to 4.0. In this chapter, only the first definition was used.

The objective of BHZ and GZ was to study whether these widely publicized measures have any predictive power that is not incorporated into the public's pari-mutuel win odds. If not fully incorporated, then a further objective was to investigate whether these measures could be used to determine win probability estimates that are sufficiently superior to the public's so that a profitable wagering scheme based on win betting could be developed, despite the significant transaction costs.

4. DATA ACQUISITION

This section discusses the nature of the data in BHZ and GZ, while the sources of their data are described in the Appendix.

The public's win betting pool and results were collected for the period 1946 to 2007. For 54 of these years, dollar amounts that the public wagered were found which yielded precise values for q_i . For the other eight years, only the final win odds for each betting interest were available. In these cases it was possible to back out win probabilities that were consistent with these odds.

The EFH listing and pedigree information for each Derby participant were collected for each year from 1946 to 2007. The original list of chefs was published in 1981. For years prior to 1981, this list was used, which means that the classification of chefs for 1946 to 1980 is not completely out of sample. The Kentucky Derby hypothetical betting begins in 1981, so all betting is based on lists of chefs that were out of sample. For the period 1981–1986, the 1981 list was used (see Appendix for explanation). After 1986, an updated list of chefs was used each year. The Belmont calculations are for 1981–2006 and 1946–2006, with no dual or asterisk qualifiers in 2007.

The major races for asterisk-qualifier status, with their 2007 graded stakes classification and the years that they have been run over the interval 1946–2005, were the Blue Grass Stakes (G1; 1946–2005), the Flamingo Stakes (currently not run; 1946–1989, 1992–2001), the Florida Derby (G1; 1952–2005), the Santa Anita Derby (G1; 1946–2005), and the Wood Memorial Stakes (G1; 1946–2005). The Flamingo Stakes declined in importance before being cancelled, but was included because historically it was an important prep race.

5. APPLICATION OF BREEDING INFORMATION AND PERFORMANCE MEASURES TO REFINE ESTIMATED WIN PROBABILITIES FOR THE KENTUCKY DERBY

BHZ developed two models for estimating win probabilities that depended on whether a betting interest was a dual qualifier or a dual-or-asterisk qualifier.

The 1995 Kentucky Derby is used in Table 2 to illustrate the required information. Also evident is a complication with regard to accounting for pedigree with entries (and

TABLE 2 Sample Input Data: 1995 Kentucky Derby Field

Horse	W_i/W	Entry	EFH	Won specified	DI	Qualifier status	
						Dual	Asterisk
Jambalaya Jazz	0.044	1	115		1.15		
Pyramid Peak			—	Flamingo	3.00		•
Serena's Song	0.189	2	122		2.11	•	
Timber Country			126		3.29	•	
Mecke			107		4.50		
Knockadoon			—		3.57		
Citadeed	0.066	Field	—		1.60		
In Character			—		1.77		
Ski Captain			—		3.67		
Lake George			—		4.50		
Thunder Gulch	0.033		116	Florida Derby	4.00	•	
Tejano Run	0.087		121		2.38	•	
Jumron	0.126		115		3.80		
Eltish	0.070		123		3.00	•	
Afternoon Deelites	0.086		124		5.00		
Suave Prospect	0.059		113		4.60		
Talkin Man	0.167		114	Wood Memorial	3.00		•
Dazzling Falls	0.029		111		6.20		
Wild Syn	0.042		—	Blue Grass	4.33		

NOTE: W_i/W : Post-time fraction of win pool. Entry: Entry number or field. EFH: Experimental free handicap weight. Blank implies not weighted. (High weight for two-year-olds from 1994 was 126 lb.) Won specified race: • Winner of a major race prior to Kentucky Derby. DI : Dosage Index: see Equation (2). Qualifier status: • Implies meets qualifier requirements. Source: Bain et al., 2006.

the field): the horses in an entry may not have the same qualifier status. This difficulty was handled using the following scheme:

1. If all members of an entry had the same qualifier status, then the entry was considered as one horse with that qualification.
2. If one member of the entry was a dual qualifier plus had won any of the designated major races prior to the Kentucky Derby, the entry was considered to be a dual qualifier regardless of the qualifications of the other member(s) (based on the presumption that in most cases most of the public's attention on the entry was due to that horse).
3. If the members of an entry did not all have the same qualifier status, but each was either a dual qualifier or an asterisk qualifier, then the entry was viewed as a dual-or-asterisk qualifier.
4. In all other cases the entry was considered to be neutral, that is, neither an asterisk qualifier nor a dual-or-asterisk qualifier.

The qualifier status of the field was determined in the same manner. For the dual-qualifier model there are 67, 0, 10, and 22 betting interests in the respective categories, and for the dual-or-asterisk-qualifier model there are 57, 2, 10 and 30 betting interests in the respective categories.

With respect to the dual-qualifier model, of the winners there are 29 that are qualifiers, 26 that are not qualifiers, and 3 that are part of a neutral entry. With respect to the dual-or-asterisk-qualifier model, of the winners there are 41 qualifiers, 16 that are not qualifiers, and 3 that are part of a neutral entry. In 1998 and 2003 there were no dual qualifiers, so those years were ignored in the dual-qualifier modeling.

The base-case model relates a betting interest's win probability to the public's wagering to see if looking solely at the pools without the "expert information" could lead to a profitable betting scheme.

Let W_i^j be the public's win bet on betting interest i , W^j be the win pool in race j , and N^j be the number of betting interests in race j . For race j , define p_i^j to be the probability that betting interest i wins and define $q_i^j \equiv W_i^j / W^j$ to be the fraction of the win pool bet on betting interest i . For this base case, the following model was used for each race

$$p_i^j = \frac{(q_i^j)^\delta}{\sum_{m=1}^{N^j} (q_m^j)^\delta}. \quad (3)$$

If $\delta = 1$, then $p_i^j = q_i^j$.

BHZ used a standard maximum-likelihood approach to estimate optimal values of δ . Consider R independent races and define $K \equiv (k_1, \dots, k_R)$ to be an R -tuple representing the winners of the R races, that is, k_j is the number of the betting interest that won race j . Let $p_{k_j}^j$ represent the estimated probability based on Equation (3), evaluated before race j , that betting interest k_j wins race j . The probability that the vector K corresponds to the winners of the R races is

$$P(K|\delta) = \prod_{j=1}^R p_{k_j}^j. \quad (4)$$

Treating Equation (4) as a likelihood function that depends on δ gives

$$\varrho(\delta|K) \propto \prod_{j=1}^R p_{k_j}^j. \quad (5)$$

A maximum-likelihood point estimate for δ , namely δ_{ML} , can be found by maximizing the likelihood as a function of δ . The first value for δ_{ML} was calculated using the first 10 years of data, namely 1946–1955 inclusive. Thereafter, the value of δ_{ML} was updated for each year using data from 1946 to that year. The win pool fraction for the winner and values for δ_{ML} calculated after each year's race are shown in Figure 1.

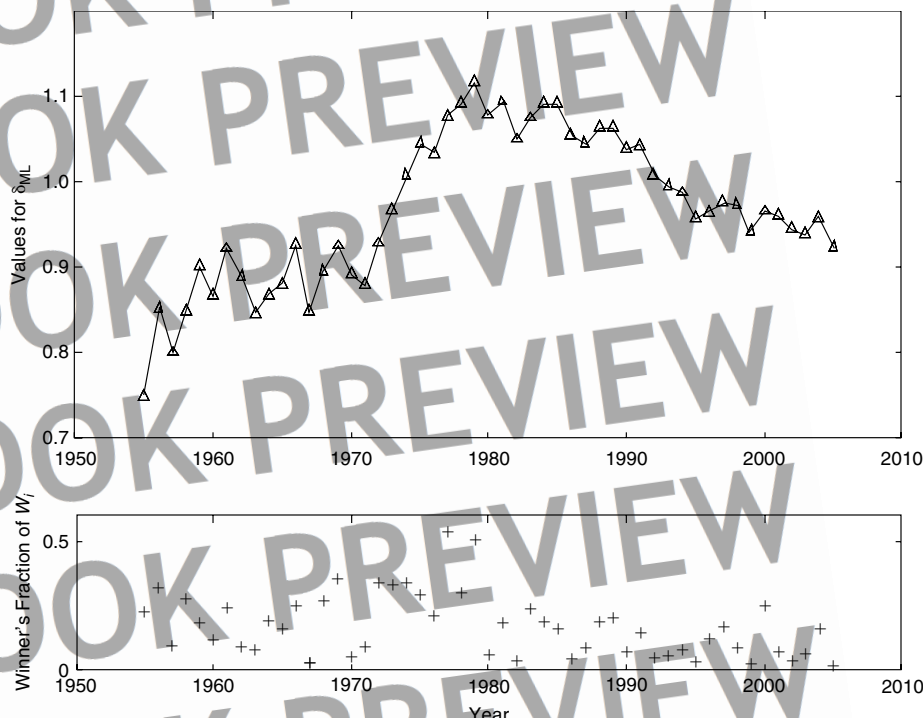


FIGURE 1 Value for δ_{ML} after each year's race and the fraction of the win pool bet on the winner, 1955–2005. *Source:* Bain et al., 2006.

The values for δ_{ML} are less than 1.0 for the years prior to 1974. This is a consequence of the public's more favored betting interests winning less often during this period than would have been expected based on the public's odds. The winners from 1972 to 1979 were dominated by favorites, culminating with Spectacular Bid in 1979, so δ_{ML} increases over this interval, reaching a maximum value of 1.12. During the period of 1980–2005, the public's favorite seldom won, so δ_{ML} tends to decrease to its final value 0.92.⁸

With values of δ_{ML} close to 1.0, Equation (3) generates revised win probabilities that differ only slightly from the fraction of the win pool. The greatest ratio p_i^j/q_i^j over the interval 1981–2005 is 1.12. This 12% edge is insufficient to offer a positive expected

⁸Griffith (1949), McGlothlin (1956), Ali (1977), Asch et al. (1982), and Ziemba and Hausch (1986), among others, have demonstrated that the public's wagering has a strong and stable bias of underbetting the favorites and overbetting the longshots. This results in $\delta > 1.0$. Ziemba and Hausch (1987) provided evidence that this "favorite-longshot bias" is exhibited at the Kentucky Derby but it is weaker, that is more flat, than in these earlier studies. The recent advent of rebate and betting exchange wagering has led to a flattening of the favorite-longshot bias in recent data since about 1998; see Ziemba (2004, 2008) and Snowberg and Wolfers (2008).

return on a win bet after accounting for the transaction costs; hence this simple model points to weak-form efficiency of the win-betting market over this period.

The main objective of BHZ was to use this procedure to create models that modified the win probability for each betting interest based on whether or not it was considered a dual qualifier or considered a dual-or-asterisk qualifier.

For this case, BHZ viewed the probability of betting interest i winning to be

$$p_i^j = \frac{(q_i^j)^{\gamma_i}}{\sum_{m=1}^{N_j} (q_m^j)^{\gamma_m}}, \quad \gamma_i, \gamma_m = \alpha, \beta, \text{ or } 1. \quad (6)$$

The variable $\gamma_i = \alpha$ if betting interest i was a dual qualifier (or dual-or-asterisk qualifier depending on the test), $\gamma_i = \beta$ if it was classified as not a dual qualifier (or not a dual-or-asterisk qualifier if applicable), and $\gamma_i = 1$ if the betting interest was an entry or field classified as being neutral.

Based on Equation (6), BHZ calculated annual maximum-likelihood values for α and β , denoted as α_{ML} and β_{ML} , were calculated each year. The initial estimate for 1956 used the first 10 years of data (1946–1955). Figure 2 illustrates the progression

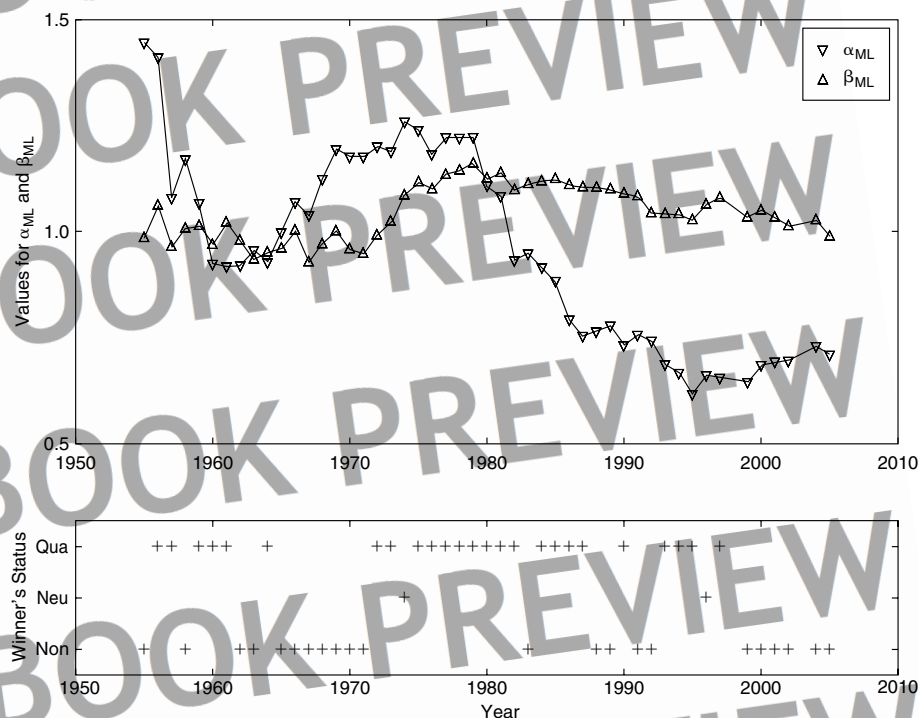


FIGURE 2 Values for α_{ML} and β_{ML} for dual-qualifier model after each year's race, 1955–2005.
NOTE: In the figure, Qua = dual qualifier; Neu = member of a neutral entry; and Non = non-qualifier.

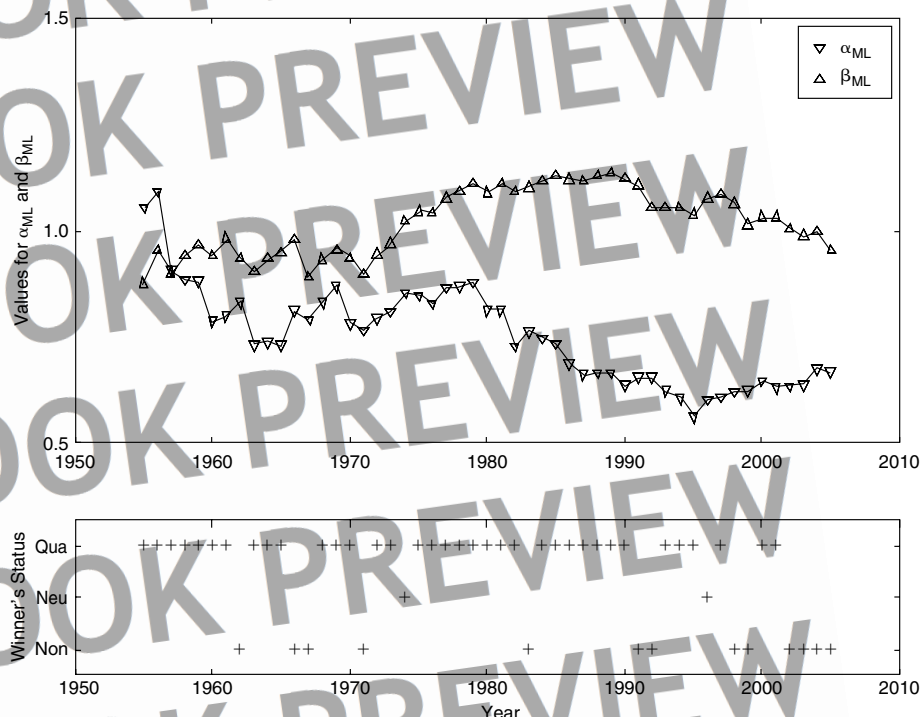


FIGURE 3 Value for α_{ML} and β_{ML} for dual-or-asterisk qualifier model after each year's race, 1955–2005. NOTE: In the figure, Qua = dual qualifier; Neu = member of a neutral entry; and Non = non-qualifier. Source: Bain et al., 2006.

of α_{ML} and β_{ML} values for the dual-qualifier model, and Figure 3 shows α_{ML} and β_{ML} values for the dual-or-asterisk-qualifier model.

In Figures 2 and 3, the critical pattern is the relative magnitude of α_{ML} and β_{ML} . In Figure 2, for nearly 20 years, α_{ML} exceeds β_{ML} . Consequently, for this period, the revised win probability for each dual qualifier is less than the fraction of the money bet on it in the win pool. This implies that betting on dual qualifiers, if they had been known, would not have been advantageous during that period. In the mid-1970s, dual qualifiers began to win consistently, eventually leading to β_{ML} exceeding α_{ML} for the remainder of the study period. For this later period, the revised win probabilities for dual qualifiers exceed their fraction of the win pool. Figure 3 for dual-or-asterisk qualifiers shows a similar pattern, although β_{ML} begins to exceed α_{ML} after only three years. Thus, after the third year, the model predicts win probabilities for dual-or-asterisk qualifiers that exceed their fraction of the win pool.

The original and revised estimates of win probabilities for 1995 are in Table 3, where a betting interest's estimated win probability rises if it meets the qualifier criterion. This is not necessarily the case if there are many qualifiers because the sum of the

TABLE 3 Original and Revised Estimated Win Probabilities for the 1995 Kentucky Derby

Betting interest	W_i/W	DQ	p_{DQ}	DAQ	p_{DAQ}
Entry 1	0.044	-1	0.029	0	0.027
Entry 2	0.189	1	0.242	1	0.220
Field	0.066	-1	0.044	-1	0.035
Thunder Gulch	0.033	1	0.075	1	0.076
Tejano Run	0.087	1	0.144	1	0.137
Jumron	0.126	-1	0.085	-1	0.068
Eltish	0.070	1	0.125	1	0.121
Afternoon Deelites	0.086	-1	0.057	-1	0.046
Suave Prospect	0.059	-1	0.039	-1	0.031
Talkin Man	0.167	-1	0.114	1	0.204
Dazzling Falls	0.029	-1	0.019	-1	0.015
Wild Syn	0.042	-1	0.027	-1	0.022

NOTE: W_i/W : fraction of win pool bet on the betting interest. DQ: Indicator is 1 for dual qualifiers, 0 for unclassified entries, and -1 for non-qualifiers. p_{DQ} : Revised estimated win probability based on dual-qualifier model. DAQ: Indicator is 1 for dual-or-asterisk qualifiers, 0 for unclassified entries, and -1 for non-qualifiers. p_{DAQ} : Revised estimated win probability based on dual-or-asterisk qualifier model. *Source*: Bain et al., 2006.

probabilities is unity. The effect of considering asterisk qualifiers is demonstrated by Talkin Man, an asterisk qualifier but not a dual qualifier. Talkin Man's estimated win probability varies from 0.114 to 0.204 depending on the criterion used.

The revised win probability estimates are occasionally sufficiently greater than the fraction of the win pool to allow a positive expected return even considering transaction costs.

The percent increase in the win probability over the fraction of the win pool for Thunder Gulch is much greater than for Entry 2. There is a general tendency for p_i/q_i to increase for qualifiers as q_i decreases, which is a consequence of the power function model in Equation (6) together with values of $\alpha < \beta$ and $\alpha < 1$.

6. THE KELLY BETTING MODEL

BHZ calculated betting amounts using the Kelly-optimal capital growth model which maximizes the expected logarithm of wealth on a race-by-race basis. This approach was proposed by Kelly (1956), and was extended and rigorously proved by Breiman (1961) and Algoet and Cover (1988). Among its properties are: (1) it maximizes the

asymptotic growth rate of wealth; (2) it asymptotically minimizes the expected time to reach any specific sufficiently large wealth level; and (3) in the long run, it outperforms any other essentially different betting strategy almost surely and asymptotically provides infinitely more final wealth than any other essentially different strategy. (See MacLean et al., 1992, 2006; and Thorp, 2006, for further properties, and Ziemba and Hausch, 1986, for simulation results for shorter time horizons.)

The revised probability of betting interest i winning based on dual-qualifier or dual-or-asterisk qualifier status is p_i . Let r_i be the gross return per dollar bet based on the win odds established by the public. (As in Section 1, the superscript indicating the race number is suppressed.) The model requires solving the following optimization problem for each race:

$$\begin{aligned} & \text{maximize}_{f \in \mathbb{R}^N} \sum_{i=1}^N p_i \log \left(1 - \sum_{m=1}^N f_m + f_i r_i \right) \text{ s.t. } f_i \geq 0 \forall i = 1, \dots, N \text{ and } \sum_{i=1}^N f_i \leq 1. \end{aligned} \quad (7)$$

The decision variable, f_i , is the fraction of the current wealth to bet on betting interest i . Suppose that the bettor's initial wealth is w and betting interest i wins. Then $w f_i r_i$ is returned to the bettor after having invested

$$w \sum_{m=1}^N f_m,$$

for a final wealth of

$$w \left(1 - \sum_{m=1}^N f_m + f_i r_i \right).$$

The objective function is the logarithm of final wealth for each betting interest winning, weighted by the probability of that betting interest winning. Initial wealth, w , can be disregarded in the formulation with the non-negative decision variables the fraction of wealth that is bet on each betting interest. The constraints comprise a budget constraint.

This formulation assumes that the bets are sufficiently small so that they do not influence the payout on any betting interest, that is, the bets on betting interest i do not reduce r_i . For the Kentucky Derby and the Belmont Stakes, the win betting pools are so large that a typical bet is unlikely to influence the payouts.⁹ The large pools also permit this assumption because the percent bet on each betting interest is assumed to vary little in the final few minutes.

⁹See Hausch et al. (1981) for a formulation that does account for the bettor's effect on payouts.

7. THE KENTUCKY DERBY, 1981–2006

In BHZ's study of semi-strong efficiency of the Kentucky Derby win market, they started by revising win probabilities using the base-case model [Equations (3–5)]. The revisions were sufficiently close to the public's win probabilities that expected returns were negative for all betting interests during the period 1981–2005. Thus, the optimization problem in Equation (7) with these revised probabilities led to no wagers.

For the models based on Equation (6) and on status as a dual or dual-or-asterisk qualifier, BHZ started with an initial wealth of \$2,500 in 1981. Wealth was updated each year based on the bets made and the actual outcome of the race. The wealth history for betting from 1981 to 2005 is shown in Figure 4 for dual qualifiers and for dual-or-asterisk qualifiers. The overall results are summarized in Table 4. For the years up to the mid-1980s, any advantage identified by the model is small and results in small bets. Comparing the values in Figure 4 to those in Figures 2 and 3 shows that α_{ML} and β_{ML} are relatively close over that interval. As the model predicts a greater advantage, the amount per bet, and consequently the volatility, grows.

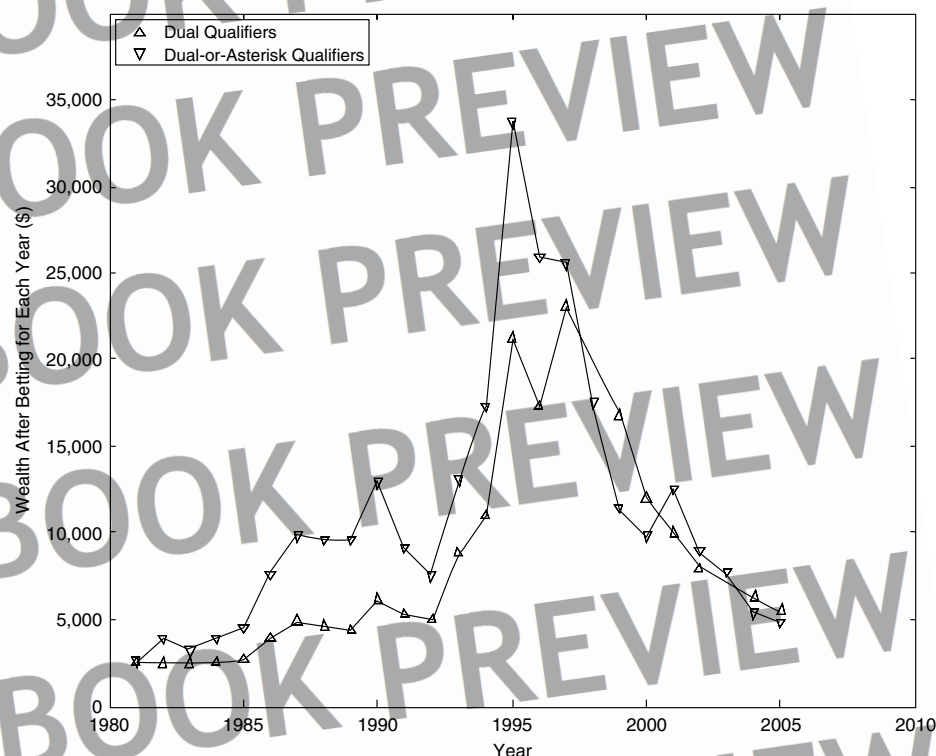


FIGURE 4 Wealth level history for Kelly win bets, 1981–2005. Source: Bain et al., 2006.

TABLE 4 Profits from Both Models Based on Kelly Bets and Revised Win Probabilities

	Model based on qualifier type	
	Dual	Dual-or-asterisk
Number of bets	61	107
Total amount bet	\$32,828	\$66,467
Number of bet that won	9	14
Initial wealth	\$2,500	\$2,500
Final wealth	\$5,514	\$4,889
Total profit	\$3,014	\$2,389
Percent return on investment	9.2	3.6

Source: Bain et al., 2006.

For comparison, betting \$2,500 on the favorite to win from 1981 to 2005 would yield a loss of \$41,500, betting \$200 to win on each dual qualifier would yield a profit of \$12,920 on \$13,200 bet, and a \$200 bet to win on each dual-or-asterisk qualifier would yield a profit of \$7,780 on \$23,000 bet (neutral entries excluded as qualifiers). The improved return on investment compared to the OCGM in the short run is mostly due to a few huge profits on qualifiers Gato del Sol in 1982, Ferdinand in 1986, and Thunder Gulch in 1995.

For both models the betting scheme produced profits during the 1980s and up to the mid-1990s, but the performance has been poor since. Several possibilities can be considered for this:

1. The sample size is small. This could mean that the sequence of successes for qualifiers for both models from 1972 to 1997 was a short-term run, so that in the long run there is nothing to be gained from using either model developed here. The limited sample size also implies that the final wealth is sensitive to individual race results. To give an idea of the scope, two extreme examples are (i) if a non-dual-qualifier had won in 1995, instead of Thunder Gulch, the final wealth for the dual-qualifier model would be \$2,221, that is, a slight loss overall, and (ii) if 2005 winner Giacomo had a favorable change of a single dosage point in any category, Giacomo would have been a dual qualifier and the final wealth would have been \$13,877.
2. It is extremely difficult to make a proper assessment of all of the two-year-olds, so the EFH can omit suitable horses. A pointed example occurred in 2003 where eventual 2004 Derby winner Smarty Jones was not rated on the EFH. Yet, he had overwhelmed a field of state-bred two-year-olds at Philadelphia Park, but that race is not counted when determining the EFH. Roman (2005a) lists other Derby winners such as Winning Colors and Sunday Silence who were superior two-year-olds but were not rated on the EFH.

3. Classification of chefs is an ongoing exercise. For example, Alydar was classified as a chef subsequent to Strike The Gold winning the Derby in 1991, so Strike The Gold, who won the Blue Grass Stakes, is not considered as an asterisk qualifier here ($DI = 9.00$), yet when Alydar was classified, Strike The Gold's DI was reduced to 2.60. Ziemba (1991) wrote a column about this on April 28, 1991, one week prior to the 1991 Derby arguing that Alydar should be a classic chef as Alydar had numerous classic distance winners. The reclassification of Alydar, and at what point, would possibly change other pedigrees. However, BHZ and GZ use Roman's classification, so Strike the Gold is neither a dual nor an asterisk qualifier.
4. Roman (2005b) has pointed out the gradual rise in the DI of Derby winners over time, so the failure of the system in the last few years of the study may reflect a shift of the overall breed in North America toward speed at shorter distances. Real Quiet in 1998, Charismatic in 1999, and Giacomo in 2005 all had DI values greater than 4.00.
5. The Flamingo Stakes decreased in significance in the final years that it was run. Including the Flamingo winner as an asterisk qualifier in recent years was unwarranted in retrospect. Two possible solutions are to drop the Flamingo at some point in the analysis, or switch to the Arkansas Derby as the fifth significant prep race.

Random betting generates expected losses in excess of 16%, due to the 16% track take plus breakage. BHZ showed both qualifier designations approximately doubling wealth over the betting period. They also used two approaches to address the statistical significance of these profits. The first approach treats a betting interest's win or loss as a binomial random variable and then uses a normal approximation. The second simulates the set of races assuming random wagering.

Before considering their first approach, observe that the data in Figure 4 are not ideal for addressing statistical significance. Wealth generally grows until the mid-to-late 1990s, and then dramatically falls. This pattern of wins and losses leads to wealth that is highly variable. Focusing on just the dual-qualifier case, Figure 5 superimposes on Figure 4 the wealth level history assuming that the races were run in reverse order, that is, we started in 2005 with \$2,500, then updated wealth based on our results in 2005 and went to the 2004 race, and so on. Thus, the string of large losses occurs early with lower wealth, after which wins are more common. The final wealth is identical, since the optimal capital growth system simply determines the optimal fraction of wealth to bet each year. (For example, losing 10% one year and gaining 20% the other year leads to an overall return of 8%, whatever the order of the win and the loss.) Despite the final wealth being the same, the wealth histories are very different, as is the appearance of any statistical significance to the profits.

To eliminate the effect of varying wealth, which can dampen or intensify the variance in profits, BHZ's test of statistical significance uses bets and returns in each race

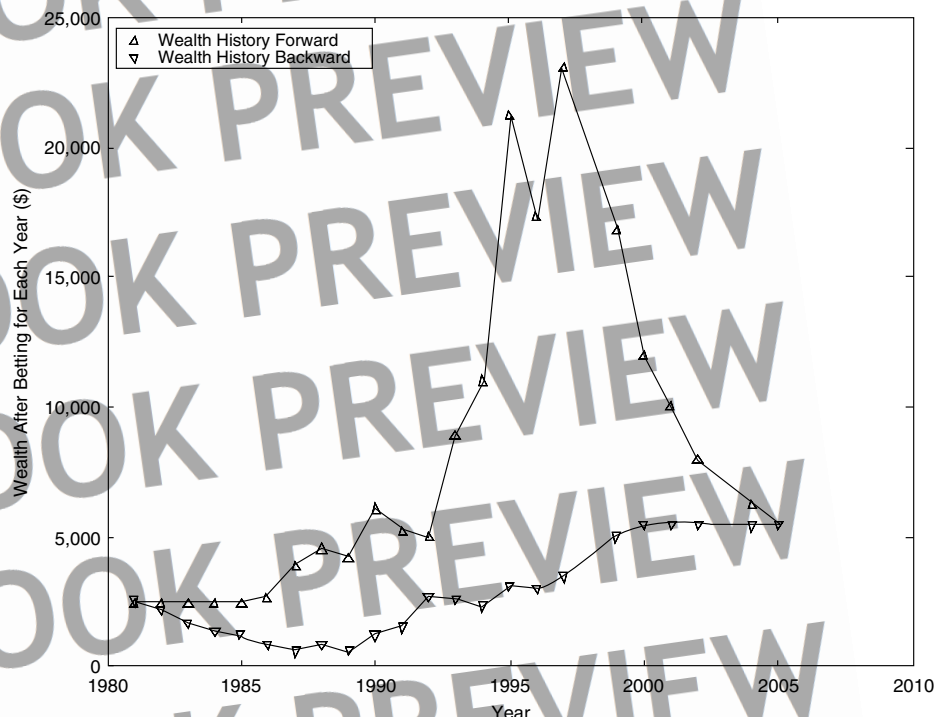


FIGURE 5 Betting wealth for Kelly win bets on dual qualifiers with races run forward (1981–2005) and run backward (2005–1981). *Source:* Bain et al., 2006.

assuming an identical initial wealth each race. They do not update wealth year by year (as in Figure 4); instead the initial wealth each year is assumed to be \$2,500.

Let q be the probability of winning a bet in each trial, n be the number of trials, c be the amount wagered each trial, and r be the gross return upon winning,¹⁰ and let X be the random variable representing the number of wins. The probability of profits exceeding a constant π is

$$P[rX - nc > \pi]. \quad (8)$$

Assume that the trials are independent. For bets in different races, this assumption is reasonable. For multiple bets on the same race—which are common—wins are negatively correlated, since if one betting interest wins then the others must lose. Negative correlation leads to a tighter distribution of wins, so this analysis based on independent

¹⁰In practice, q , c , and r vary across races and even within races if there are multiple wagers. We approximate the sequence of wagers by using the average values of these parameters.

trials underestimates the statistical significance of the results. Since X is binomially distributed, the normal distribution approximates Equation (8) as

$$1 - \Phi \left[\frac{\frac{\pi + nc}{r} - nq}{\sqrt{nq(1-q)}} \right], \quad (9)$$

where Φ is the cumulative distribution function of a standard $N(0, 1)$ variable.

For dual qualifiers and assuming an initial wealth of \$2,500 each year, there were 61 bets totaling \$7,079, and of these bets, nine won for a gross return of \$11,357 and a profit of \$4,278. Thus, $c = 7,079/61 = 116.0$ and $r = 11,357/9 = 1,262$. If the system were no better than random betting, then q satisfies $rq - c = -0.16c$, recognizing the 16% track take, giving $q = 0.07725$. Then, by Equation (9), the probability of profits of at least the observed level of \$4,278 given random betting is 2.0%. Suppose instead that the system is better than random betting but only good enough to offer zero expected profits. Then q solves $rq - c = 0$, or $q = 0.09192$, and, by Equation (9), the probability that such a system would produce at least the observed profits is 6.7%.

For dual-and-asterisk qualifiers, with initial wealth of \$2,500 each year, there were 107 bets totaling \$13,268; 14 of these bets won for a gross return of \$18,253 and a profit of \$4,985. Thus, $c = 124$ and $r = 1,304$. If the system were no better than random betting, then $q = 0.07988$ and the probability of profits of at least the observed level is only 2.6%. Assuming instead that the system is better than random betting but only good enough to offer zero expected profits, then $q = 0.09509$ and the probability that such a system would produce at least the observed profits is 10.4%.

The second approach used by BHZ to address the statistical significance of the results involved two simulations for each qualifier designation. The first simulation dealt with the question of how likely it would be that profits at the observed level would have been generated if their approach was vacuous and, therefore, was essentially random wagering. The second simulation asked how likely it would be that the observed profits would have been generated if the system was able to improve upon random wagering, but only enough to achieve zero expected return on each wager (excluding breakage). The algorithm for the first simulation was

1. Start with a betting wealth of \$2,500 in 1981.
2. Determine the fraction of wealth to wager on each betting interest i for the current year based on the Kelly criterion and the (wrong) assumption that our probability estimate, p_i , is correct.
3. Randomly select the winner, with the probability of winning for betting interest i being q_i .
4. Based on the simulated winner, its payout and our wagers, update wealth.
5. Repeat steps (2) to (4) for each year in order up to 2005.

The second simulation differed only in step 3, where the simulation used q_i/Q as the correct win probability for any betting interest i that received a wager in step 2. The expected return on every wager was zero, before accounting for breakage. The collective

TABLE 5 Results from 10,000 Betting Simulations with \$2,500 Initial Wealth

	Dual qualifier		Dual-or-asterisk qualifier	
	Simulation 1	Simulation 2	Simulation 1	Simulation 2
Final wealth \leq \$1,000 (%)	54.3	37.0	73.8	49.3
Final wealth \leq \$2,500 (%)	84.2	72.6	91.5	76.1
Final wealth \geq system's final wealth (Table 4) (%)	3.9	9.2	3.1	11.4
Mean final wealth	\$1,555	\$2,449	\$1,030	\$2,473
Median final wealth	\$892	\$1,371	\$450	\$1,022
Maximum final wealth	\$67,601	\$161,777	\$97,336	\$409,294

Source: Bain et al., 2006.

win probability of the other betting interests was such that probabilities summed to one. For example, in a three-horse race with $[q_1, q_2, q_3] = [0.42, 0.21, 0.37]$ and Kelly bets having been placed on the first two horses based on $[p_1, p_2, p_3]$, the probability that the simulation would select each betting interest as the winner was $[0.42, 0.21, 0.37]$ for Simulation 1, and was $[0.5, 0.25, 0.25]$ for Simulation 2 after dividing the first two fractions by 0.84. The simulations were run 10,000 times each. The results are in Table 5.

For all simulations, losses occurred more than 70% of the time. For Simulation 2 and for both types of qualifiers, the mean final wealth was close to \$2,500, which is expected given the modification in step 3 of this simulation. For dual qualifiers, only 3.9% of the time did Simulation 1 realize profits as high as our observed profit. For Simulation 2, the corresponding value is 9.2%. For dual-or-asterisk qualifiers, these values for Simulations 1 and 2 are, respectively, 3.1% and 11.4%.

As a final test, the analyses were conducted using limited data. For example, if the year being considered was 1997 and the interval was 25 years, only information from 1972 to 1996 would have been applied. For the dual-qualifier model, using the entire dataset produced the greatest final wealth; while for the dual-or-asterisk model, using an interval of 56 years produced a final wealth of \$5,305.

8. THE PREAKNESS STAKES, 1946–2006

The Preakness Stakes, unlike the Kentucky Derby and Belmont Stakes, does not provide a new test of stamina for its competitors. The race is run two weeks after the Kentucky Derby and is 1/16 mi shorter in distance. The application of the dosage system would likely be negated by the fact that entrants in the Preakness exiting the Kentucky Derby have a quantifiable result of their ability to run longer distances. GZ find that dual qualifiers and asterisk qualifiers outperform non-dual qualifiers and non-asterisk qualifiers

but have not historically earned a positive wagering return. Using a test for differences between two population proportions given in Equation (10), we are able to show that dual qualifiers win a statistically greater proportion of races than non-dual qualifiers and asterisk qualifiers win a statistically greater proportion of races than non-asterisk qualifiers. The test statistic is

$$z = \frac{(\bar{p}_{dq} - \bar{p}_{ndq})}{\sqrt{\bar{p}(1 - \bar{p}) \left(\frac{1}{n_{dq}} + \frac{1}{n_{ndq}} \right)}} \quad (10)$$

where \bar{p} is a proportion of winners from dual qualifier (dq), non-dual qualifiers (ndq), and the pooled sample and n_{dq} is the number of dual qualifiers and n_{ndq} non-dual qualifiers. The 116 dual qualifiers won 28 of 61 Preaknesses between 1946 and 2006. This translates into 24.1% (28/116) of dual qualifiers were winners versus only 7.5% (33/440) of non-dual qualifiers which is significantly different at a 1% level ($z = 5.10$). The results were similar for asterisk qualifiers ($z = 5.36$) with 22.0% winners (36/164) and non-asterisk qualifiers with 6.4% winners (25/392).

While dual and asterisk qualifiers won the Preakness more than non-qualifiers, the wagering returns were negative for flat win bets on dual and asterisk qualifiers from 1946 to 2006 (see Table B2 in the Appendix and Figure 6). Wagers on each of the 116 dual qualifiers would yield a loss of \$14.90 (−12.8%). The interval from 1981 to 2006 which corresponds to the first Roman/Rasmussen publication on dosage in the *Daily Racing Form*, however, does show a net positive return of \$2.80 (7.6%) for dual qualifiers (see Table B2 in the Appendix and Figure 7). Win wagers on asterisk qualifiers lose \$39.80 (24.3%) from 1946 to 2006 and \$17.10 (−10.4%) from 1981 to 2006.

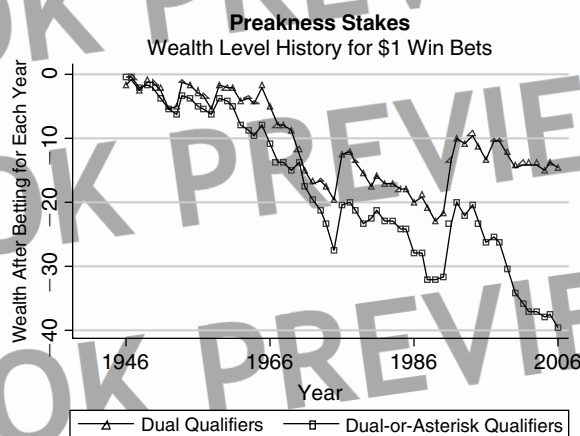


FIGURE 6 Preakness wealth level history for \$1 win bets, 1946–2006.

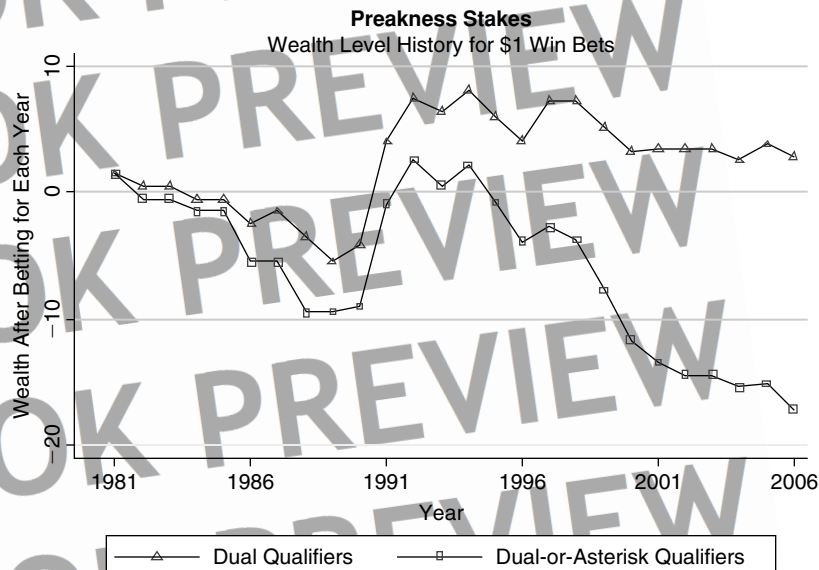


FIGURE 7 Preakness wealth level history for \$1 win bets, 1981–2006.

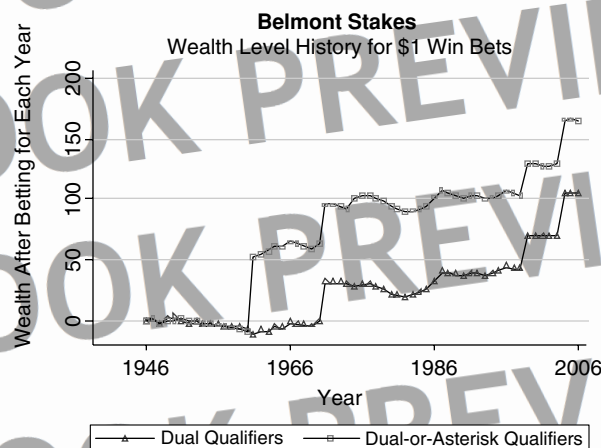


FIGURE 8 Belmont Stakes wealth level history for \$1 win bets, 1946–2006.

9. THE BELMONT STAKES, 1946–2006

The Belmont Stakes provides a unique test of stamina where competitors run a full 1/4 mi farther than they have before. GZ find that the application of expert information from the dosage theory does result in a significant advantage. Table B3 and Figures 8 and 9 give the wealth level histories for 1946–2006 and 1981–2006, respectively. Both

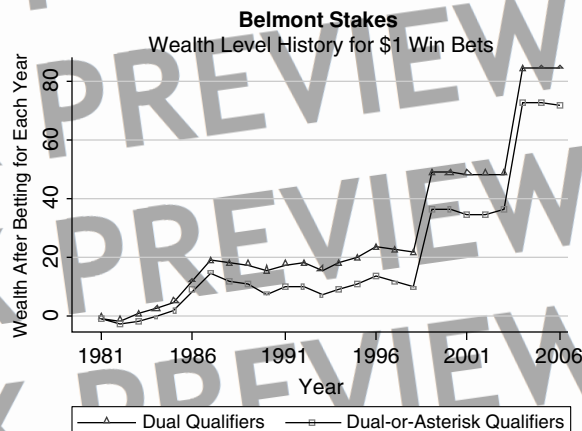


FIGURE 9 Belmont Stakes wealth level history for \$1 win bets, 1981–2006.

demonstrate considerable statistically significant positive profits. Year by year dual and asterisk results for 1946–2006 have \$1 flat win bets beginning in 1946 growing to resulting in \$105.55 (98.6%) and \$164.85 (113.7%) in profit for dual and asterisk qualifiers, respectively. The results for 1981–2006 yield \$80.85 (216.3%) and \$67.95 (129.7%) in profit for dual and asterisk qualifiers, respectively.

Using differences between two population proportions given earlier in Equation (10), gives that the proportion of dual qualifier winners is statistically greater than the proportion of non-dual qualifying winners at less than the 1% significance level ($z = 6.73$). Indeed, the dual qualifiers won 29.9% of the time vs. 6.8% for the non-dual qualifiers. We reject the hypothesis at $P_{dq} = P_{ndq}$ at a level well below 1% significance. The proportion of asterisk qualifier winners is also statistically greater than the proportion of non-asterisk qualifier winners at less than the 1% significance level ($z = 6.26$). The asterisk qualifiers won 25.5% of the time versus 6.2% for the non-asterisk qualifiers.

10. CONCLUSIONS

The racetrack is a useful financial market for testing market efficiency and considerable evidence exists in support of the track's win market being weak-form efficient. This chapter, however, summarizes the work of BHZ and GZ that show the win market is not semi-strong efficient.

BHZ and GZ focused on a particular aspect of the Kentucky Derby and Belmont Stakes, whose distances of $1\frac{1}{4}$ and $1\frac{1}{2}$ miles are typically farther than any entrant has ever raced, and the Preakness Stakes, which is run between these two races at $1\frac{3}{16}$ mi. This lack of direct evidence of an entrant's stamina for this race has motivated the search for indirect evidence. Dosage theory, which analyzes a horse's pedigree, has

been offered as such evidence but it has also been controversial, both in general and in its relation to the Kentucky Derby. Other evidence that has been offered includes well-publicized rankings of horses and results from recent high-caliber races.

BHZ did not evaluate the criticisms or the justifications offered for the dosage concept and for the ranking of two-year-olds, nor did they attempt to refine their application to the Kentucky Derby. Instead, they simply merged this publicly available information with the public's win odds to establish "adjusted" win probabilities. They then tested these win probabilities within a betting system based on the optimal capital growth model and showed statistically significant profits.

GZ applied this procedure to the 1½ mi Belmont Stakes, which is run five weeks after the Kentucky Derby. From the 1980s to the mid-1990s when the dual qualifiers were having very good success in the Derby, GZ's results in the Belmont were not as good. However, in recent years the situation has reversed with superior results in the Belmont than the Derby. The betting systems discussed here are two of many strategies used by bettors. In the ¹³/₁₆ Preakness, the dosage breeding theory is less of a factor, which is understandable given that the race is shorter than the Kentucky Derby, and thus evidence of a horse's stamina exists. Even so, a positive return on dual qualifiers exists from 1981–2006. The procedure outlined shows that given the pools from a set of races for which the strategy is applicable, the simple model given in Equation (6) can be used to test the validity of the strategy.

References

- Ali, M. M. 1977. Probability and Utility Estimates for Racetrack Bettors, *Journal of Political Economy* 85, 803–815.
- Algoet, P., and T. Cover. 1988. Asymptotic Optimality and Asymptotic Equipartition Properties of Log-Optimum Investments, *Annals of Probability* 16, 876–898.
- Asch, P., B. G. Malkiel, and R. E. Quandt. 1982. Racetrack Betting and Informed Behavior, *Journal of Financial Economics* 10, 187–194.
- Bain, R., D. B. Hausch, and W. T. Ziemba. 2006. An Application of Expert Information to Win Betting on the Kentucky Derby, 1981–2005, *European Journal of Finance* 12, 283–301.
- Breiman, L. 1961. Optimal Gambling System for Favorable Games. *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability* 1, 63–68.
- Camerer, C. F. 1998. Can Asset Markets be Manipulated? A Field Experiment with Racetrack Betting, *Journal of Political Economy* 106, 457–482.
- Chew, P. 1974. *The Kentucky Derby: The First 100 Years*. Houghton Mifflin, Boston.
- Fama, E. 1970. Efficient Capital Markets: A Review of Theory and Empirical Work, *Journal of Finance* 25, 383–417.
- Fama, E. 1991. Efficient Capital Markets: II, *Journal of Finance* 46, 1575–1617.
- Gramm, M., and W. T. Ziemba. 2008. A Test of Semi-Strong Form Market Efficiency: The Triple Crown Races Betting Markets. Working Paper. University of British Columbia, Vancouver, British Columbia, Canada.
- Griffith, R. M. 1949. Odds Adjustments by American Horse racing Bettors, *American Journal of Psychology* 62, 290–294.
- Hausch, D. B., W. T. Ziemba, and M. Rubinstein. 1981. Efficiency of the Market for Racetrack Betting, *Management Science* 27, 1435–1452.
- Hausch, D. B., V. S. Lo, and W. T. Ziemba (ed). 1994. *Efficiency of Racetrack Betting Markets*, Academic Press, San Diego, CA.

- Keim, D. B., and W. T. Ziemba (eds.). 2000. *Security Market Imperfections in Worldwide Equity Markets*, Cambridge University Press, Cambridge UK.
- Kelly, J. L. 1956. A New Interpretation of Information Rate, *Bell System Technical Journal* 35, 917–926.
- MacLean, L. C., W. T. Ziemba, and G. Blazenko. 1992. Growth Versus Security in Dynamic Investment Analysis, *Management Science* 38, 1562–1585.
- MacLean, L. C., and W. T. Ziemba. 2006. Capital Growth Theory and Practice, in S. A. Zenios and W. T. Ziemba (eds.), *Handbook of Asset and Liability Management, Vol. 1: Theory and Methodology*. North-Holland, Amsterdam, pp. 429–473.
- McGlothlin, W. H. 1956. Stability of Choices Among Uncertain Alternative, *American Journal of Psychology* 69, 604–619.
- Roberts, H. 1967. Statistical Versus Clinical Prediction of the Stock Market. Mimeo, University of Chicago.
- Roman, S. A. 1981. Dosage: A Practical Approach, *Daily Racing Form*, May.
- Roman, S. A. 2001. Personal Communication to R. S. Bain. May 17, 2001.
- Roman, S. A. 2002. Dosage: Pedigree and Performance, The Russell Meerdink Company, Ltd., Neenah, Wisconsin.
- Roman, S. A. 2005a. Dosage and Two-Year-Old Form: A Predictor of Classic Performance. Available at http://www.chef-de-race.com/dosage/classics/di_and_2yo_form.htm.
- Roman, S. A. 2005b. Dosage Figures for Kentucky Derby Winners Since 1940. Available at http://www.chef-de-race.com/dosage/classics/derby_dosage.htm.
- Snowberg, E., and J. Wolfers. 2008. Examining Explanations of a Market Anomaly: Preferences or Perceptions? in D. B. Hausch and W. T. Ziemba, *Handbook of Sports and Lottery Markets*. North-Holland, Amsterdam, pp. 103–133.
- Sung, M., and J. E. V. Johnson. 2008. Semi-Strong Form Information Efficiency in the Horse Race Betting Markets, in D. B. Hausch and W. T. Ziemba, *Handbook of Sports and Lottery Markets*. North-Holland, Amsterdam, pp. 275–306.
- Thorp, E. O. 2006. The Kelly Criterion in Blackjack, Sports Betting and the Stock Market, in S. A. Zenios and W. T. Ziemba (eds.), *Handbook of Asset and Liability Management, Vol. 1: Theory and Methodology*. North-Holland, Amsterdam, pp. 385–428.
- Varola, F. 1974. *Typology of the RaceHorse*. J. A. Allen, London.
- Varola, F. 1980. *Functional Development of the Thoroughbred*. J. A. Allen, London.
- Vuillier, J.-J. 1902. *Les Croisements Rationnels dans la Race Pur*. Legoupy, Paris.
- Vuillier, J.-J. 1906. *Les Croisements Rationnels dans la Race Pur*. Legoupy, Paris.
- Vuillier, J.-J. 1928. *Les Croisements Rationnels dans la Race Pur*. Maulde, Paris.
- Ziemba, W. T. 2004. Behavioral finance, betting racetrack, and options and futures trading. Presentation to the Mathematical Finance Seminar, Stanford University, January.
- Ziemba, W. T., and D. B. Hausch. 1986. *Betting at the Racetrack*, Dr. Z Investments Inc, San Luis Obispo, CA.
- Ziemba, W. T., and D. B. Hausch. 1987. *Dr. Z's Beat the Racetrack*. William Morrow, New York.
- Ziemba, W. T. 1991. Triple Crown Dosage Analysis Review, *Win Magazine*, August.
- Ziemba, W. T. 2008. Efficiency of Racing, Sports, and Lottery Betting Markets, in D. B. Hausch and W. T. Ziemba, *Handbook of Sports and Lottery Markets*. North-Holland, Amsterdam, pp. 183–222.

APPENDIX A: Data Sources

A.1. Public's Wagering

For races from 1946 to 1991, betting data were taken from tables published in *The Courier-Journal*, a Louisville, Kentucky, newspaper, usually on the Sunday after the Kentucky Derby. The pools for 1970 could not be found. There were several discrepancies in the data for which the published pools did not sum to the totals, or did not correspond with published win odds. Adjustments were made for errors for which an apparent revision could be made. For 1999–2002, complete pools were obtained from the Bloodstock Research Information Services Website (<http://www.bris.com>). In 2001, the pools also appeared on the Website for the home of the Kentucky Derby, Churchill Downs (<http://www.churchilldowns.com>). The 2003 pools were sent directly by Churchill Downs, and the 2004 and 2005 pools were obtained courtesy of John Swetye who obtained them from Philadelphia Park's Phonebet service. From 1992 to 1998, the pools recorded in *The Courier-Journal* did not have all of the bets included. While these totals are not available, win odds based on total wagering are available. Therefore, for 1970 and 1992–1998 we estimated the total win pool and backed out a set of win pool fractions that are consistent with the published win odds. There were no dual qualifiers in 1998 and 2003 so those years were excluded from the dual-qualifier modeling.

A.2. Pedigrees

Pedigree information was taken from *The Blood-Horse* magazine, the *American Produce Records*, a software database called "The Pedigree Program," the pedigree query Website, <http://owl.net scout.com/pedigree> (no longer active), the Del Mar Turf Club Website, <http://www.dmtc.com/dmtc98/Pedigree/>, thoroughbred registries, and Roman's Website, <http://www.chef-de-race.com>. The 2004 data were sent by personal communication from Roman to John Swetye who forwarded them to us.

A.3. Chef-de-Race Listings

Classifications of chefs were taken from the original 1981 list (Roman, 2000), the *American Racing Manual* for each year from 1986 (the first year that the list was included) to 1993, and from Roman's Website. For the period 1981–1986, the 1981 list was used. For years prior to 1981, the original list was used. For 2001 to 2003, and 2005, Dosage Indices and EFH rankings tabulated by Roman were taken from his Website, <http://www.chef-de-race.com>, and for 2004 they were sent via email from Roman (see Section A.2, above).

A.4. Experimental Free Handicap Listings

The EFH listings were taken from the *American Racing Manual*, the *Blood-Horse* magazine (print and online at <http://www.bloodhorse.com>), the Thoroughbred Times Website, <http://www.thoroughbredtimes.com>, and Roman's Website, <http://www.chef-de-race.com>.

A.5. Results of the Kentucky Derby and Major Races Prior to the Kentucky Derby

The results of the Kentucky Derby were taken from the *Daily Racing Form*, both print and online (<http://www.drf.com>), press materials from Churchill Downs, and from Chew (1974). Recent results charts were obtained from the following Websites:

About.com Inc.	http://horseracing.about.com
Sportsline.com Inc.	http://www.sportsline.com
CNN/Sports Illustrated	http://www.sportsillustrated.cnn.com
Equibase Company, LLC	http://www.equibase.com
Daily Racing Form, LLC	http://www.drf.com

The results for the major races prior to the Derby were taken from the *American Racing Manual* and lists from the following Websites:

Blue Grass Stakes:	http://www.keeneland.com/liveracing/history.asp
Flamingo Stakes:	http://hialeahpark.com/99/HallofFame/flamingo.htm
Florida Derby:	http://www.thoroughbredchampions.com/library/fladerby.htm
Santa Anita Derby	http://www.revistahipodromo.com/santaanita.html
Wood Memorial Stakes	http://www.nyra.com/aqueduct/index2.html

and the out-of-date site, <http://www.iglou.com/tbred/tc97/preps>, which was run by the *Thoroughbred Times*.

APPENDIX B: Kentucky Derby, Preakness, and Belmont Winners, 1946–2006

TABLE B1 Kentucky Derby, 1946–2006

Year	Winner	Odds	Dual qualifiers
1946	Assault*	8.2	5
1947	Jet Pilot	5.4	4
1948	Citation	0.4	4
1949	Ponder	16	5

TABLE B1 (continued)

Year	Winner	Odds	Dual qualifiers
1951	Count Turf	14.6	2
1952	Hill Gail	1.1	2
1953	Dark Star	24.9	3
1954	Determine	4.3	5
1955	Swaps*	2.8	2
1956	Needles	1.6	4
1957	Iron Leige	8.4	5
1958	Tim Tam*	2.1	1
1959	Tomy Lee	3.7	8
1960	Venetian Way	6.3	5
1961	Carry Back	2.5	3
1962	Decidedly	8.7	2
1963	Chateaugay*	9.4	3
1964	Northern Dancer	3.4	6
1965	Lucky Debonair*	4.3	5
1966	Kauai King	2.4	2
1967	Proud Clarion	30.1	4
1968	Forward Pass*	2.2	4
1969	Majestic Prince*	1.4	3
1970	Dust Commander*	15.3	4
1971	Canonero II	8.7	1
1972	Riva Ridge	1.5	3
1973	Secretariat	1.5	2
1974	Cannonade	1.5	1
1975	Foolish Pleasure	1.9	1
1976	Bold Forbes	3	4
1977	Seattle Slew	0.5	4
1978	Affirmed	1.8	3
1979	Spectacular Bid	0.6	3
1980	Genuine Risk	13.3	5
1981	Pleasant Colony	3.5	4
1982	Gato del Sol	21.2	3
1983	Sunny's Halo	2.5	3
1984	Swale	3.4	3
1985	Spend A Buck	4.1	2
1986	Ferdinand	17.7	2
1987	Alysheba	8.4	5

(continued)

TABLE B1 (continued)

Year	Winner	Odds	Dual qualifiers
1988	Winning Colors *	3.4	3
1989	Sunday Silence*	3.1	3
1990	Unbridled	10.8	4
1991	Strike The Gold*	4.8	4
1992	Lil E. Tee	16.8	3
1993	Sea Hero	12.9	3
1994	Go For Gin	9.1	5
1995	Thunder Gulch	24.5	6
1996	Grindstone	5.9	6
1997	Silver Charm	4	2
1998	Real Quiet	8.4	0
1999	Charismatic	31.3	6
2000	Fusaichi Pegasus*	2.3	5
2001	Monarchos*	10.5	3
2002	War Emblem	20.5	3
2003	Funny Cide	12.8	0
2004	Smarty Jones	4.1	5
2005	Giacomo	50.3	2
2006	Barbaro*	6.1	2

NOTE: Bold indicates dual qualifier, * indicates asterisk qualifier.

TABLE B2 Preakness Stakes, 1946–2006

Year	Winner	Odds	Number	Dual qualifiers		Number	Asterisk qualifiers	
				Return 46–06	Return 81–06		Return 46–06	Return 81–06
1946	Assault*	1.4	2	–\$2.00		3	–\$0.60	
1947	Faultless	4.2	4	–\$0.80		5	–\$0.40	
1948	Citation	0.1	3	–\$2.70		3	–\$2.30	
1949	Capot	2.5	2	–\$1.20		3	–\$1.80	
1950	Hill Prince	0.7	2	–\$1.50		2	–\$2.10	
1951	Bold	4.1	1	–\$2.50		2	–\$4.10	
1952	Blue Man*	1.6	3	–\$5.50		4	–\$5.50	
1953	Native Dancer	0.2	1	–\$5.30		2	–\$6.30	
1954	Hasty Road	5	2	–\$1.30		3	–\$3.30	

TABLE B2 (continued)

Year	Winner	Odds	Dual qualifiers			Asterisk qualifiers		
			Number	Return 46–06	Return 81–06	Number	Return 46–06	Return 81–06
1955	Nashua	0.3	2	−\$2.00		2	−\$4.00	
1956	Fabius	2.5	1	−\$3.00		1	−\$5.00	
1957	Bold Ruler	1.4	3	−\$3.60		3	−\$5.60	
1958	Tim Tam*	1.1	2	−\$5.60		3	−\$6.50	
1959	Royal Orbit	6.6	4	−\$2.00		5	−\$3.90	
1960	Bally Ache	1.7	3	−\$2.30		3	−\$4.20	
1961	Carry Back	1	2	−\$2.30		3	−\$5.20	
1962	Greek Money	10.9	2	−\$4.30		3	−\$8.20	
1963	Candy Spots	1.5	2	−\$3.80		3	−\$8.70	
1964	Northern Dancer	2.1	4	−\$4.70		4	−\$9.60	
1965	Tom Rolfe	3.6	2	−\$2.10		3	−\$8.00	
1966	Kauai King	1	3	−\$5.10		3	−\$11.00	
1967	Damascus	1.8	3	−\$8.10		3	−\$14.00	
1968	Forward Pass*	1.1	0	−\$8.10		2	−\$13.90	
1969	Majestic Prince*	0.6	1	−\$9.10		3	−\$15.30	
1970	Personality*	4.5	3	−\$12.10		4	−\$13.80	
1971	Canonero II	3.4	3	−\$15.10		4	−\$17.80	
1972	Bee Bee Bee	18.7	2	−\$17.10		2	−\$19.80	
1973	Secretariat	0.3	1	−\$16.80		3	−\$21.50	
1974	Little Current	13.1	1	−\$17.80		2	−\$23.50	
1975	Master Derby	23.4	2	−\$19.80		4	−\$27.50	
1976	Elocutionist	10.1	4	−\$12.70		4	−\$20.40	
1977	Seattle Slew	0.4	1	−\$12.30		1	−\$20.00	
1978	Affirmed	0.5	3	−\$13.80		3	−\$21.50	
1979	Spectacular Bid	0.1	3	−\$15.70		3	−\$23.40	
1980	Codex*	2.7	2	−\$17.70		3	−\$22.70	
1981	Pleasant Colony	1.5	1	−\$16.20	\$1.50	1	−\$21.20	\$1.50
1982	Aloma's Ruler	6.9	1	−\$17.20	\$0.50	2	−\$23.20	−\$0.50
1983	Deputed Testamony	14.5	0	−\$17.20	\$0.50	0	−\$23.20	−\$0.50
1984	Gate Dancer	4.8	1	−\$18.20	−\$0.50	1	−\$24.20	−\$1.50
1985	Tank's Prospect	4.7	0	−\$18.20	−\$0.50	0	−\$24.20	−\$1.50
1986	Snow Chief	2.6	2	−\$20.20	−\$2.50	4	−\$28.20	−\$5.50
1987	Alysheba	2	2	−\$19.20	−\$1.50	3	−\$28.20	−\$5.50
1988	Risen Star	6.8	2	−\$21.20	−\$3.50	4	−\$32.20	−\$9.50
1989	Sunday Silence*	2.1	2	−\$23.20	−\$5.50	3	−\$32.10	−\$9.40
1990	Summer Squall	2.4	2	−\$21.80	−\$4.10	3	−\$31.70	−\$9.00

(continued)

TABLE B2 (continued)

Year	Winner	Odds	Number	Dual qualifiers		Number	Asterisk qualifiers	
				Return 46–06	Return 81–06		Return 46–06	Return 81–06
1991	Hansel	9.1	2	–\$13.70	\$4.00	2	–\$23.60	–\$0.90
1992	Pine Bluff	3.5	1	–\$10.20	\$7.50	1	–\$20.10	\$2.60
1993	Prairie Bayou	2.2	1	–\$11.20	\$6.50	2	–\$22.10	\$0.60
1994	Tabasco Cat	3.6	3	–\$9.60	\$8.10	3	–\$20.50	\$2.20
1995	Timber Country	1.9	2	–\$11.60	\$6.10	3	–\$23.50	–\$0.80
1996	Louis Quatorze	8.5	2	–\$13.60	\$4.10	3	–\$26.50	–\$3.80
1997	Silver Charm	3.1	1	–\$10.50	\$7.20	3	–\$25.40	–\$2.70
1998	Real Quiet	2.5	0	–\$10.50	\$7.20	1	–\$26.40	–\$3.70
1999	Charismatic	8.4	2	–\$12.50	\$5.20	4	–\$30.40	–\$7.70
2000	Red Bullet	6.2	2	–\$14.50	\$3.20	4	–\$34.40	–\$11.70
2001	Point Given	2.3	3	–\$14.20	\$3.50	5	–\$36.10	–\$13.40
2002	War Emblem	2.8	0	–\$14.20	\$3.50	1	–\$37.10	–\$14.40
2003	Funny Cide	1.9	0	–\$14.20	\$3.50	0	–\$37.10	–\$14.40
2004	Smarty Jones	0.7	1	–\$15.20	\$2.50	1	–\$38.10	–\$15.40
2005	Afleet Alex	3.3	3	–\$13.90	\$3.80	4	–\$37.80	–\$15.10
2006	Bernardini	12.9	1	–\$14.90	\$2.80	2	–\$39.80	–\$17.10

NOTE: Bold indicates dual qualifier, * indicates asterisk qualifier.

TABLE B3 Belmont Stakes, 1946–2006

Year	Winner	Odds	Number	Dual qualifiers		Number	Asterisk qualifiers	
				Return 46–06	Return 81–06		Return 46–06	Return 81–06
1946	Assault*	1.4	1	–\$1.00		2	\$0.40	
1947	Phalanx	2.3	2	\$0.30		2	\$1.70	
1948	Citation	0.2	4	–\$2.50		5	–\$2.10	
1949	Capot	5.6	2	\$2.10		4	\$0.50	
1950	Middleground	2.7	3	\$2.80		4	\$0.20	
1951	Counterpoint	5.15	3	–\$0.20		4	\$2.35	
1952	One Count	12.8	2	–\$2.20		3	–\$0.65	
1953	Native Dancer	0.45	1	–\$1.75		1	–\$0.20	
1954	High Gun	3.45	2	–\$3.75		3	–\$3.20	
1955	Nashua	0.15	1	–\$3.60		1	–\$3.05	
1956	Needles	0.65	2	–\$3.95		2	–\$3.40	
1957	Gallant Man	0.95	1	–\$4.95		1	–\$4.40	
1958	Cavan	4.5	0	–\$4.95		1	–\$5.40	

TABLE B3 (continued)

Year	Winner	Odds	Dual qualifiers			Asterisk qualifiers		
			Number	Return 46-06	Return 81-06	Number	Return 46-06	Return 81-06
1959	Sword Dancer	1.65	3	-\$5.30		4	-\$6.75	
1960	Celtic Ash	8.4	2	-\$7.30		2	-\$8.75	
1961	Sherluck*	65.05	4	-\$11.30		5	\$52.30	
1962	Jaipur	2.85	1	-\$8.45		2	\$54.15	
1963	Chateaugay*	4.5	1	-\$9.45		2	\$57.65	
1964	Quadrangle	6.55	3	-\$4.90		4	\$61.20	
1965	Hail to All	2.65	1	-\$5.90		1	\$60.20	
1966	Amberoid	5.5	2	-\$1.40		2	\$64.70	
1967	Damascus	0.8	1	-\$2.40		1	\$63.70	
1968	Stage Door Johnny	4.4	1	-\$3.40		2	\$61.70	
1969	Arts and Letters	1.7	1	-\$4.40		2	\$59.70	
1970	High Echelon	4.5	2	-\$0.90		2	\$63.20	
1971	Pass Catcher	34.5	3	\$31.60		3	\$95.70	
1972	Riva Ridge	1.6	3	\$31.20		3	\$95.30	
1973	Secretariat	0.1	1	\$31.30		3	\$93.40	
1974	Little Current	1.5	1	\$30.30		2	\$91.40	
1975	Avatar*	13.2	2	\$28.30		4	\$101.60	
1976	Bold Forbes	0.9	1	\$29.20		1	\$102.50	
1977	Seattle Slew	0.4	1	\$29.60		1	\$102.90	
1978	Affirmed	0.6	3	\$28.20		3	\$101.50	
1979	Coastal	4.4	3	\$25.20		3	\$98.50	
1980	Temperence Hill	53.4	4	\$21.20		5	\$93.50	
1981	Summing	7.9	1	\$20.20	-\$1.00	1	\$92.50	-\$1.00
1982	Conquistador Cielo	4.1	1	\$19.20	-\$2.00	2	\$90.50	-\$3.00
1983	Caveat	2.6	1	\$21.80	\$0.60	3	\$91.10	-\$2.40
1984	Swale	1.5	1	\$23.30	\$2.10	1	\$92.60	-\$0.90
1985	Crème Fraiche	2.5	1	\$25.80	\$4.60	1	\$95.10	\$1.60
1986	Danzig Connection	8	2	\$32.80	\$11.60	2	\$102.10	\$8.60
1987	Bet Twice	8	2	\$39.80	\$18.60	3	\$108.10	\$14.60
1988	Risen Star	2.1	1	\$38.80	\$17.60	3	\$105.10	\$11.60
1989	Easy Goer	1.6	3	\$38.40	\$17.20	4	\$103.70	\$10.20
1990	Go And Go	7.5	2	\$36.40	\$15.20	3	\$100.70	\$7.20
1991	Hansel	4.1	3	\$38.50	\$17.30	3	\$102.80	\$9.30
1992	A.P. Indy	1.1	2	\$38.60	\$17.40	2	\$102.90	\$9.40
1993	Colonial Affair	13.9	2	\$36.60	\$15.40	3	\$99.90	\$6.40

(continued)

TABLE B3 (continued)

Year	Winner	Odds	Number	Dual qualifiers		Number	Asterisk qualifiers	
				Return 46–06	Return 81–06		Return 46–06	Return 81–06
1994	Tabasco Cat	3.4	2	\$39.00	\$17.80	2	\$102.30	\$8.80
1995	Thunder Gulch	1.5	1	\$40.50	\$19.30	1	\$103.80	\$10.30
1996	Editor's Note	5.8	3	\$44.30	\$23.10	4	\$106.60	\$13.10
1997	Touch Gold	2.65	1	\$43.30	\$22.10	2	\$104.60	\$11.10
1998	Victory Gallop	4.5	1	\$42.30	\$21.10	2	\$102.60	\$9.10
1999	Lemon Drop Kid	29.75	3	\$70.05	\$48.85	4	\$129.35	\$35.85
2000	Commendable	18.8	0	\$70.05	\$48.85	0	\$129.35	\$35.85
2001	Point Given	1.35	3	\$69.40	\$48.20	4	\$127.70	\$34.20
2002	Sarava	70.25	0	\$69.40	\$48.20	0	\$127.70	\$34.20
2003	Empire Maker*	2	0	\$69.40	\$48.20	1	\$129.70	\$36.20
2004	Birdstone	36	1	\$105.40	\$84.20	1	\$165.70	\$72.20
2005	Afleet Alex	1.15	2	\$105.55	\$84.35	2	\$165.85	\$72.35
2006	Jazil	6.2	0	\$105.55	\$84.35	1	\$164.85	\$71.35

NOTE: **Bold** indicates dual qualifier, * indicates asterisk qualifier. Crème Fraiche was coupled with dual qualifier Stephan's Odyssey.

CHAPTER 16

Efficiency in Horse Race Betting Markets: The Role of Professional Tipsters

Bruno Deschamps

University of Bath, Claverton Down, Bath, UK.

Olivier Gergaud

Université de Reims, Champagne-Ardenne, France.

1. Introduction	342
2. The Model	343
3. Data	345
3.1. <i>Tips and Rewards Rules</i>	345
3.2. <i>Public Information</i>	346
3.3. <i>Measuring Forecast Originality</i>	347
4. Results	348
4.1. <i>Results of Frequency Tests</i>	348
4.2. <i>Originality and Accuracy</i>	348
4.3. <i>Anti-Herding and Excess Originality</i>	350
5. Conclusion	351
<i>References</i>	351
Appendix: Proof of Equation (5)	353

Abstract

We use a dataset of more than 11,000 horse racing forecasts from 35 professional tipsters and investigate whether they make excessively original forecasts. We find that tipsters do exaggerate and make forecasts that are excessively distant from the public information given their private information. This result has implications for the efficiency of betting markets.

1. INTRODUCTION

It is well established that horse race betting markets are inefficient. A large empirical literature has documented odds inefficiencies, including the well documented favorite-longshot bias. According to the favorite-longshot bias, betting on favorites yields a higher return than betting on longshots. Several theoretical explanations have been advanced, including the existence of risk-loving bettors, or the overestimation of the longshots' probability of winning.¹

In this chapter, we investigate the role played by professional horse racing tipsters and their influence on the odds. Our hypothesis is that career concerns lead tipsters to make biased forecasts, which in turn induces bettors to overbet on the longshots, and this generates odds inefficiencies. There are indeed several theoretical reasons to believe that tipsters are induced to make biased forecasts. Since Scharfstein and Stein (1990), numerous articles have shown that forecasters are induced either to herd (conservatism) or anti-herd (exaggeration) the public information in order to maximize their reputation.² Evidence of anti-herding behavior has been found among financial analysts by Bernhardt et al. (2006) and Chen and Jiang (2006). In this chapter, we find evidence of exaggeration among professional horse racing tipsters.

Our analysis is based on a dataset of more than 11,000 horse racing forecasts made by 35 French professional tipsters. These tipsters participate in a famous yearly tournament organized by Paris-Turf, the most influential French betting newspaper. After each race, each tipster receives a number of points depending on the accuracy of his or her tip and the difficulty of forecasting the race outcome. Performing well in this tournament is good for a tipster's career. Assessing the effects of this renowned tournament on behavior is beyond the scope of this chapter, and is analyzed in Deschamps and Gergaud (2007). The novelty of this chapter is to specifically investigate whether tipsters are excessively original, in the sense that they deviate excessively from the public information given their private information.

We provide two types of excessive originality evidence. First, we analyze the relationship between forecast precision and originality. We first develop a simple model in which we compare the distance between forecasts and the public information to the distance between the final odds and the public information. We find that the former

¹See Ottaviani and Sørensen (2008) and Vaughan Williams (2005) for a survey of the literature.

²See, among others, Ottaviani and Sørensen (2006), Prendergast and Stole (1996), and Ehrbeck and Waldmann (2001).

distance is larger than the latter, meaning that tipsters make forecasts that are excessively distant from the public information. Second, we find that tipsters include in their forecasts horses that do not perform well compared to horses that are not included and that are ex ante likely to perform well. Hence, forecasts would be more accurate if tipsters decided to deviate less from the public information.

Arguably, professional tipsters have a significant impact on the final odds, given that bettors often base their betting strategy on professional tips. If, for instance, tipsters report that a particular horse is the favorite, we might expect bettors to bet large amounts on that horse. Given that the odds are determined by the betting volumes, the odds will consequently be particularly short for that horse. Hence, we believe that excessive originality among professional tipsters partially explains the well documented favorite-longshot bias.

The chapter is organized as follows. Section 2 presents the model. In Section 3 we describe the data. Section 4 presents the results, and Section 5 concludes.

2. THE MODEL

We develop a model of forecasting with a testable prediction to detect exaggeration. Consider a tipster having to predict the race order of arrival, so that his or her forecast consists of an ordered list of horses, ranked according to their perceived quality. We call q_i the objective quality of horse i , while Q is as an ordered list of horses, from the highest q_i to the lowest. The Q vector is also the race order of arrival, so that the highest quality horse finishes first and so on. Tipsters do not directly observe horses' quality, so their task is to estimate Q in order to predict the race result. We call c_i the public information (or consensus forecast) on the quality of horse i . It represents the information that is common knowledge to all tipsters regarding that horse's quality. This information is available before tipsters make their forecast and is therefore the prior on q_i . Specifically, we assume that q_i is normally distributed around the public information c_i :

$$q_i \sim N(c_i, \sigma_c^2) \quad (1)$$

where σ_c^2 is the imprecision of the public information. We assume for simplicity that σ_c^2 is independent of i . Hence, c_i can be written as $q_i = c_i + \varepsilon_{ci}$, and $E(c_i - q_i)^2 = \sigma_c^2$. Note that c_i is not the odds. Indeed, the odds are published only after tipsters make their forecasts, while the public information is available before forecasts are made. In addition to the consensus forecast, each tipster receives a private signal s_i on each horse. Private signals provide additional information on horses' quality, and its precision is what differentiates good tipsters from bad ones. Private signals are also assumed to be normally distributed:³

$$s_i \sim N(q_i, \sigma_s^2) \quad (2)$$

³We assume here that $s_i - q_i$ and $c_i - q_i$ are not correlated, but allowing a positive correlation would actually make the result stronger.

Given that tipsters observe both s_i and c_i before making a forecast, the standard normal learning model implies that the belief on q_i is a weighted average of the public and the private information:

$$E(q_i | s_i, c_i) = \frac{\sigma_c^2}{\sigma_c^2 + \sigma_s^2} s_i + \frac{\sigma_s^2}{\sigma_c^2 + \sigma_s^2} c_i \quad (3)$$

This shows that the more precise the public information is (σ_c^2 low), the lower is the weight put on the private signal. On the other hand, tipsters whose private information is very reliable (σ_s^2 low), tend to attach less importance to public information. Based on their information, tipsters have to tell which horses are the best and which are the worst. We call f_i the forecast that a tipster makes on q_i , and F an ordered list of horses from the highest forecast quality (high f_i) to the lowest. If forecasts were truthful, we would by definition have f_i equal to the ex post belief, that is, $f_i = E(q_i | s_i, c_i)$. Note that, under truthtelling, the distance between the forecast and the public information would be $E(f_i - c_i)^2 = \frac{\sigma_c^2}{\sigma_c^2 + \sigma_s^2} \sigma_c^2$.⁴ Indeed, the more precise the private signals (σ_s^2 low), the more tipsters distance themselves from the public information.

We define k_i as the quality of horse i that can be inferred from the final odds. Given that many bettors base their betting strategy on tipsters' recommendation, k_i is partially determined by f_i . However, further information can be released between the moment F is made public and the moment the odds are computed, so k_i is also affected by this new information. Formally, we call this new information ϕ_i , and we consider that it is unbiased: $\phi_i \sim N(q_i, \sigma_\phi^2)$. Then, depending on the relative precision of f_i and ϕ_i , we have that $k_i = \beta \phi_i + (1 - \beta) f_i$, where $1 - \beta$ represents the influence of f_i on k_i and depends on σ_ϕ^2 , σ_c^2 , σ_s^2 .⁵ In that case, after some algebra, we get that $k_i - c_i = -\varepsilon_{ci} + (1 - \beta) \left[\frac{\sigma_s^2}{\sigma_c^2 + \sigma_s^2} \varepsilon_{ci} + \frac{\sigma_c^2}{\sigma_c^2 + \sigma_s^2} \varepsilon_{si} \right] + \beta \varepsilon_{\phi_i} = (1 - \beta) \frac{\sigma_c^2}{\sigma_c^2 + \sigma_s^2} \varepsilon_{si} + \left[(1 - \beta) \frac{\sigma_c^2}{\sigma_c^2 + \sigma_s^2} - 1 \right] \varepsilon_{ci} + \beta \varepsilon_{\phi_i}$. Hence, $E(k_i - c_i)^2 = (1 - \beta)^2 \left(\frac{\sigma_c^2}{\sigma_c^2 + \sigma_s^2} \right)^2 \sigma_s^2 + \left[(1 - \beta) \frac{\sigma_s^2}{\sigma_c^2 + \sigma_s^2} - 1 \right]^2 \sigma_c^2 + \beta^2 \sigma_\phi^2$. The result of this model is thus that, as long as the forecasts are truthful:

$$E(f_i - c_i)^2 < E(k_i - c_i)^2 \quad (4)$$

⁴Proof: $f_i - c_i = \frac{\sigma_c^2}{\sigma_c^2 + \sigma_s^2} c_i + \frac{\sigma_s^2}{\sigma_c^2 + \sigma_s^2} s_i - c_i = -\frac{\sigma_c^2}{\sigma_c^2 + \sigma_s^2} c_i + \frac{\sigma_s^2}{\sigma_c^2 + \sigma_s^2} s_i = -\frac{\sigma_c^2}{\sigma_c^2 + \sigma_s^2} \varepsilon_{ci} + \frac{\sigma_s^2}{\sigma_c^2 + \sigma_s^2} \varepsilon_{si} = \frac{\sigma_c^2}{\sigma_c^2 + \sigma_s^2} (\varepsilon_{ci} - \varepsilon_{si})$. Hence, $E(f_i - c_i)^2 = \left(\frac{\sigma_c^2}{\sigma_c^2 + \sigma_s^2} \right)^2 (\sigma_c^2 + \sigma_s^2) = \frac{(\sigma_c^2)^2}{\sigma_c^2 + \sigma_s^2}$. This assumes that ε_{si} and ε_{ci} are not correlated. Our result would actually be stronger if they were positively correlated.

⁵Formally, the standard normal learning model implies that $1 - \beta = \frac{\sigma_\phi^2}{\sigma_\phi^2 + \sigma_s^2 + \sigma_c^2} = \frac{\sigma_\phi^2 \sigma_s^2 + \sigma_\phi^2 \sigma_c^2}{\sigma_\phi^2 \sigma_s^2 + \sigma_\phi^2 \sigma_c^2 + \sigma_s^2 \sigma_c^2}$, and $\beta = \frac{\sigma_s^2 \sigma_c^2}{\sigma_\phi^2 \sigma_s^2 + \sigma_\phi^2 \sigma_c^2 + \sigma_s^2 \sigma_c^2}$. So the less precise is ϕ_i , the more k_i depends on f_i .

Indeed, after simple algebra, $\frac{\partial E(k_i - c_i)^2}{\partial \beta} = 2 \left(\frac{\sigma_s^2 \sigma_c^2}{\sigma_s^2 + \sigma_c^2} + \sigma_\phi^2 \right) \beta > 0$. And if β goes to 0, Equation (4) tends to an equality. Thus Equation (4) holds for any $\beta < 1$. Said differently, under truthtelling, the public information is on average closer to f_i than to k_i . If, for instance, private signals are infinitely precise, $\sigma_s^2 = 0$, and $f_i = q_i = k_i$. Then, $|f_i - c_i| = |k_i - c_i|$. If private signals become less precise, $|f_i - c_i|$ decreases and Equation (4) holds. The general intuition is that f_i is determined uniquely by s_i and c_i , while k_i also relies on an extra signal ϕ_i . Hence, k_i puts less weight on c_i than f_i does, which explains Equation (4).

We call C the vector that ranks the horses according to their c_i , and K is the vector that ranks horses according to the odds. Given that Equation (4) holds for each horse, we show in the appendix that, on average, the footrule distance (i.e., the absolute distance between two rank vectors) between F and C is smaller than the distance between K and C :

$$E\|F - C\| < E\|K - C\| \quad (5)$$

Indeed, given that the f_i are relatively close to c_i , F will not be very distant from C . Instead, ranking horses according to k_i , given that k_i are weakly related to c_i , will lead to K being more distant from C . This is the prediction in Equation (5) as long as forecasts are truthful.

If tipsters exaggerate and put excessive weight on their private signal and deviate excessively from the consensus forecast, we could eventually observe that the consensus is closer to the odds than to the forecasts. More precisely, if a tipster tries to distance himself or herself from the herd, he or she will bias the forecast away from the consensus. In that case, the distance between F_i and C will increase. If the anti-herding behavior is sufficiently strong, at some point the difference between the forecasts and the consensus will be larger than the distance between the odds and the consensus, and we will observe $\|F - C\| > \|K - C\|$. This could occur if tipsters have strong incentives to outperform their peers, and strategically forecast far from the public information. Section 4 tests the prediction in Equation (5) using our professional tips data. If we find that $\|F - C\| > \|K - C\|$, we could argue that tipsters anti-herd the public information and release untruthful forecasts.

3. DATA

3.1. Tips and Rewards Rules

The data come from the leading French horse racing daily newspaper, Paris-Turf. This newspaper publishes the tips by 35 professional tipsters the day before each race. The dataset that is used in this chapter covers all daily tips made during the 2004 tournament. This represents a total of 318 races and more than 11,100 tips. A tip (or forecast F) is an ordered list of eight horses that are expected to be the most competitive during the race. Paris-Turf keeps track of tipsters' successive performances. The number of points scored by a tipster is (1) positive when they win the tiercé (triple forecast), or the quarté

TABLE 1 Paris-Turf Rewards Rules

Aggregate forecast	Race outcome	Tipster 1	Tipster 2	Tipster 3
1	8	2	5	8
2	3	1	3	3
3	5	3	6	5
4	1	6	8	6
5	4	7	7	7
6				
7				
8				
Points		0	16	32

(quadruple forecast) or the quinté (quintuple forecast), (2) doubled if the forecast is in the exact order, and (3) higher when they outperform their peers.

The following basic example illustrates the rules. Consider a race involving eight horses only and a set of tipsters giving rise to an aggregate forecast as in Table 1 (column 1). We consider for simplicity tips made of five horses only instead of eight. In such a ranking, horse #1 is called the first favorite, as it was the most tipped, horse #2 the second favorite, and so on. With a race outcome as reproduced in column 2, tipster #1 does not get any points as his or her forecast is unsuccessful. On the contrary, tipster #2 (column 4) and tipster #3 (column 5) would get 16 and 32 points for a triple forecast inexact order and exact order, respectively. These results are computed as simply as adding each horse rank found in the aggregate forecast, that is to say, $5 + 3 + 8$ and $(5 + 3 + 8) \times 2$.

In this chapter, we analyze whether tipsters exaggerate the strength of their private information when making forecasts. In order to analyze the tipsters' forecasting behavior, we need to observe the public information.

3.2. Public Information

To proxy this public information (C), we rank—per race—each registered horse (between 15 and 20 horses, depending on the race) on the basis of their likelihood of winning the race from a set of 12 dummy variables: whether or not the horse is suited to the track, whether or not the horse is on form, whether the jockey/driver performed well in the recent past, and so on. We compute the sum of these 12 dummies and rank horses according to these statistics. This ordered list constitutes what we call consensus forecast, or public information. The source of information for these criteria is Paris-Turf as well.

We assume that the public information is known by all tipsters, which implies that all of these 12 variables are common knowledge. We strongly believe that this is the

case. First, this information is public and most of these indicators are published in monthly magazines⁶ well before tipsters make their forecasts. Second, these variables concern the most fundamental characteristics of the horse and their dummy nature means that they are easy to measure. Third, tipsters' forecasts are on average much more accurate than the consensus forecast, suggesting that they are well informed. Such a degree of accuracy makes it unlikely that they are not informed of these fundamental variables.

It is important to stress that C does not capture the entirety of the public information, since it is likely that some information is common knowledge but is not captured by our 12 dummies. This does not affect the result, since, for any tipster, the way we split the information between s_i and c_i does not affect f_i . Hence, f_i is not sensitive to what we call public or private information. The same holds for k_i , and we have shown that Equation (4) holds for any σ_c^2 and σ_s^2 .

3.3. Measuring Forecast Originality

Table 2 shows an example of the way a forecast's originality is calculated. We compute the Spearman footrule distance between the forecast vector and the public information vector. Imagine that the forecast is the ordered list of horses numbered 5, 6, 1, 13, 14, 12, 9, 2. The public information column shows how these eight horses are ranked in the consensus forecast. The last column is the absolute difference between the rank of the horse in the forecast and his or her rank in the consensus forecast. The sum of these differences (14 in this example) measures the distance between the forecast and the consensus forecast. We call this distance forecast originality.

TABLE 2 Measuring Forecast Originality

Horse	Rank in F	Rank in C	Absolute difference
5	1	3	2
6	2	4	2
1	3	5	2
13	4	1	3
14	5	2	3
12	6	6	0
9	7	8	1
2	8	7	1
Total			14

⁶Such as Stato Tierce.

4. RESULTS

4.1. Results of Frequency Tests

This section tests the prediction of our forecasting model. If tipsters make truthful forecasts, the consensus forecast should on average be closer to the forecasts than to the odds. We find that this is not the case, meaning that the forecasts are excessively far from the consensus and that tipsters anti-herd the public information.

Let us denote forecast originality FC_{ij} , that is the distance between the forecast F_i of expert i for race j and the consensus forecast C . The average distance between the forecast and the consensus for race j is $FC_j = E_i(FC_{ij})$. Let us define OC_j as the distance between the odds and the consensus forecast for race j . We compute OC_j the same way as for FC_{ij} , that is, by summing the rank differences between the odds and the public information as shown in Table 2. For the entire sample, we find that OC_j is on average 27.99. Crucially, we find that the forecast originality FC_j is on average 30.35. The difference is significant at 5% and this is inconsistent with truthful forecasting since $\|F_i - C\| > \|K - C\|$. Hence, this shows that the forecasts are excessively far from the consensus given tipsters' private information. We also look at the number of races for which $OC_j > FC_j$ and the number of races for which $FC_j > OC_j$. We find that, out of 318 races, FC_j is higher than OC_j 213 times, while OC_j is higher than FC_j only 105 times. Said differently, the consensus is closer to the odds than to the most of the forecasts for most races: $Pr(OC_j > FC_j) = 66.98\%$. This is empirical evidence of exaggeration.

We then analyze whether this behavior is widespread among all tipsters. To do so, we compute how often F "overshoots" K for each individual tipster. The results appear in Table 3. It shows, for each of the 35 tipsters, how often $FC_{ij} > OC_j$ and how often $OC_j > FC_{ij}$.⁷ All tipsters but three produce on average forecasts that are more distant from the consensus than the consensus is distant from the odds. This suggests that most tipsters deviate excessively from the public information.

The theoretical literature on strategic forecasting provides some insights on why tipsters may decide to exaggerate. Ottaviani and Sørensen (2006) show that forecasters participating in a forecasting contest will be induced to take high risk, in order to differentiate themselves from the other forecasters and increase their likelihood of winning. The literature on asymmetric rank-order tournaments also makes predictions that are consistent with our results. For instance, Gilpatric (2005) shows that, in asymmetric tournaments (which is the case for Paris-Turf), contestants will pursue high risk strategies as long as the prize from finishing first is large enough compared with the penalty of finishing last.

4.2. Originality and Accuracy

This section provides additional evidence of excessive originality. Theoretically, tipsters are expected to tip the eight horses they believe to have the best chance of winning the

⁷The reason why the total for each tipster does not add up to 318 is that for some races, $OC = FC$.

TABLE 3 Exaggeration Among Tipsters

Tipster	$OC > FC$	$FC > OC$	Tipster	$OC > FC$	$FC > OC$
1	79	224	19	126	178
2	114	187	20	130	164
3	117	186	21	103	192
4	154	145	22	127	173
5	139	164	23	125	171
6	126	171	24	123	190
7	114	195	25	111	202
8	128	175	26	77	238
9	106	204	27	89	219
10	128	179	28	148	156
11	119	189	29	129	178
12	99	202	30	100	211
13	147	155	31	145	152
14	89	214	32	105	189
15	151	145	33	117	196
16	120	186	34	108	198
17	154	146	35	103	203
18	128	173			

NOTE: Columns 1 and 4 include the tipster identifier. Columns 2 and 5 measure the number of races where $OC_i > FC_i$. Columns 3 and 6 measure the number of races where $OC_i < FC_i$.

race. If tipsters minimize forecast errors, their forecasts will finely weight public and private information. In that case, tipsters will achieve the highest possible frequency of successful tips. If, instead, tipsters overweight private information, we would expect the forecast success frequency to decrease. This section tests directly whether excess originality leads to less accuracy.

If forecasts were efficient, we would expect every single difference between F and C to be based on private information. Therefore every decision to drop a horse from C and replace it by another one in F should on average improve accuracy. Imagine, for instance, that the top eight horses according to C are $\{1,2,3,4,5,6,7,8\}$, and F is $\{2,3,8,5,1,7,10,9\}$. The difference between F and C is that horses #4 and #6 have been replaced by horses #9 and #10. Consequently, both horse #9 and #10 should be more likely to be top five finishers than horses #4 and #6. If, instead, tipsters have a taste for originality, horses #9 and #10 would not necessarily be more likely to be top five finishers.

In order to establish whether tipsters distance themselves excessively from the public information, we investigate whether the frequency of success of the horses tipped would rise if tipsters decided to deviate less from C . For every tipster, we compare the top five finish frequency between the lowest placed horse in F that is not included in C (#9 in

TABLE 4 Success Frequency and Originality

Tipster	#Top five for horse in <i>F</i> not in <i>C</i>	#Top five for horse in <i>C</i> not in <i>F</i>	Tipster	#Top five for horse in <i>F</i> not in <i>C</i>	#Top five for horse in <i>C</i> not in <i>F</i>
1	79	97	19	93	112
2	95	106	20	103	106
3	93	111	21	91	110
4	99	110	22	105	105
5	104	113	23	98	112
6	101	108	24	96	109
7	91	108	25	90	102
8	107	100	26	94	110
9	97	107	27	102	107
10	85	101	28	105	106
11	101	116	29	88	116
12	97	100	30	92	111
13	89	108	31	97	110
14	89	109	32	89	114
15	106	96	33	94	112
16	94	103	34	105	94
17	101	112	35	98	103
18	90	102	Average	95.9	107

NOTE: Columns 1 and 4 include the tipster identifier. Columns 2 and 5 measure the number of times the lowest horse included in *F* but not included in *C* finishes in the top five. Columns 3 and 6 measure the number of times the best placed horse included in *C* but not included in *F* finishes in the top five.

our example) and the best placed horse according to *C* that is not included in *F* (#4 in our example). Table 4 shows the result for each individual tipster. Columns 2 and 5 show the number of times that the lowest placed horse in *F* that is not included in *C* finishes in the top five. Columns 3 and 6 show the number of times that best placed horse in *C* that is not included in *F* finishes in the top five. It is striking that for 32 of the 35 tipsters, the frequency of tipped horses finishing in the top five would rise if they decided to stay closer to *C*. Importantly, the difference is significant at 5% for just a few individual tipsters. However, the average of columns 2 and 4 (95.9 top five finishes) is different at the 1% level from the average of columns 3 and 6 (107 top five finishes).

4.3. Anti-Herding and Excess Originality

We have shown that tipsters exaggerate. The most natural explanation is that, in order to win the tournament, tipsters need to tip longshots in order to differentiate themselves from the other tipsters. However, if a tipster expects the other tipsters to massively forecast the longshots, his or her best anti-herding strategy could be to make a

TABLE 5 Correlation Between Originality and Distance to the Other Tipsters

	$\text{Corr}\left(\ F_i - C\ , \frac{\sum \ F_i - F_j\ }{34}\right)$
Across races for each tipster	0.44
Across tipsters for each race	0.37

conservative forecast. Therefore, excess originality is not necessarily equivalent to anti-herding. In order to clarify the issue, we compute the correlation between originality (as measured in Table 2) and the average distance from the other tipsters. For tipster i , the latter distance is measured by the average footrule distance $\frac{\sum_j \|F_i - F_j\|}{34}$, where $j = [1, \dots, i - 1, i + 1, \dots, 35]$ are the tipsters other than tipster i . Table 5 shows the correlation between $\|F_i - C\|$ and $\frac{\sum \|F_i - F_j\|}{34}$. Across races the correlation is on average 0.44 for individual tipsters, meaning that tipsters are more likely to anti-herd when they are original. Note that the correlation is positive for every individual tipster. We also compute the correlation between originality and $\frac{\sum_j \|F_i - F_j\|}{34}$ across the tipsters for each of the 318 races. This correlation is on average 0.37, meaning that for the typical race, the most original tipsters are indeed the most likely to anti-herd. This suggests that most often, the best way to anti-herd the other tipsters is to make an original forecast.

5. CONCLUSION

This chapter analyzes the role played by professional tipsters in horse racing betting markets. We find two types of direct evidence of anti-herding behavior in forecasting. First, the distance between their forecasts and the public information is inconsistent with truthful forecasting. Second, the negative relationship between forecast originality and forecast precision is also a clear indication of biased forecasts. Overall, our findings are that tipsters do tip low ability horses, that is, horses that are unlikely to win the race. A possible explanation for such a behavior is that tipsters are careerist and want to outperform their peers by making risky forecasts. This has potentially large implications for the efficiency of betting markets. Indeed, the odds are determined by the betting volumes, and punters do rely on professional tipsters to determine their bets. Hence, the fact that tipsters favor low ability horses should also somewhat bias the odds in favor of the longshots.

References

Bernhardt, D., M. Campello, and E. Kutsoati. 2006. Who Herds? *Journal of Financial Economics* 80, 657–675.

Chen, Q., and W. Jiang. 2006. Analysts’ Weighting of Private and Public Information, *Review of Financial Studies* 19(1), 319–355.

- Deschamps, B., and O. Gergaud. 2007. Risk-Taking in Rank-Order Tournaments: Evidence from Horse Racing Tipsters. Working Paper, Université de Reims.
- Effinger, M., and M. Polborn. 2001. Herding and Anti-Herding: A Model of Reputational Differentiation, *European Economic Review* 45, 385–403.
- Gilpatric, S. 2005. Tournaments, Risk Taking, and the Role of Carrots and Sticks, Working Paper, University of Tennessee.
- Ottaviani, M., and P. Sørensen. 2006. The Strategy of Professional Forecasting, *Journal of Financial Economics* 81(2), 441–466.
- Ottaviani, M., and P. Sørensen. 2008. The Favorite-Longshot Bias: An Overview of the Main Explanations, in D. B. Hausch and W. T. Ziemba (eds.), *Efficiency of Sports and Betting Markets*, Elsevier: B.V. Amsterdam, pp. 83–101.
- Prendergast, C., and L. Stole. 1996. Impetuous Youngsters and Jaded Old-Timers: Acquiring a Reputation for Learning, *Journal of Political Economy* 104, 1105–1134.
- Scharfstein, D., and J. Stein. 1990. Herd Behavior and Investment, *American Economic Review* 80(3), 465–479.
- Vaughan Williams, L. 2005. Weak Form Information Efficiency in Betting Markets, in L. Vaughan Williams (ed.), *Information Efficiency in Financial and Betting Markets*. Cambridge University Press, Cambridge, MA, pp. 84–122.

APPENDIX: Proof of Equation (5)

To show the proof of Equation (5), let us first consider a two-horse race. Imagine that $c_2 > c_1$, so that $C = \{2, 1\}$. Hence, as in Table 2, the footrule distance is $\|F - C\| = 0$ if $f_2 > f_1$, and $\|F - C\| = 2$ if $f_1 > f_2$. Similarly, $\|K - C\| = 0$ if $k_2 > k_1$, and $\|K - C\| = 2$ if $k_1 > k_2$. We want to show that $E\|F - C\| < E\|K - C\|$, which is equivalent to $\Pr(f_1 > f_2 | c_2 > c_1) < \Pr(k_1 > k_2 | c_2 > c_1)$.

We can develop: $\Pr(f_1 > f_2) = \Pr\left(\frac{\sigma_c^2}{\sigma_c^2 + \sigma_s^2} s_1 + \frac{\sigma_s^2}{\sigma_c^2 + \sigma_s^2} c_1 > \frac{\sigma_c^2}{\sigma_c^2 + \sigma_s^2} s_2 + \frac{\sigma_s^2}{\sigma_c^2 + \sigma_s^2} c_2\right) = \Pr\left(\frac{\sigma_c^2}{\sigma_s^2} (s_1 - s_2) > (c_2 - c_1)\right)$.

Similarly,

$$\begin{aligned} \Pr(k_1 > k_2) &= \Pr\left(\beta \phi_1 + (1 - \beta) \frac{\sigma_c^2}{\sigma_c^2 + \sigma_s^2} s_1 + (1 - \beta) \frac{\sigma_s^2}{\sigma_c^2 + \sigma_s^2} c_1 > \beta \phi_2 \right. \\ &\quad \left. + (1 - \beta) \frac{\sigma_c^2}{\sigma_c^2 + \sigma_s^2} s_2 + (1 - \beta) \frac{\sigma_s^2}{\sigma_c^2 + \sigma_s^2} c_2\right) \\ &= \Pr\left(\frac{\beta}{(1 - \beta) \frac{\sigma_s^2}{\sigma_c^2 + \sigma_s^2}} (\phi_1 - \phi_2) + \frac{\sigma_c^2}{\sigma_s^2} (s_1 - s_2) > (c_2 - c_1)\right) \\ &= \Pr\left(\frac{\sigma_c^2}{\sigma_\phi^2} (\phi_1 - \phi_2) + \frac{\sigma_c^2}{\sigma_s^2} (s_1 - s_2) > (c_2 - c_1)\right). \end{aligned}$$

By the properties of the sum of two normal distributions,

$$\frac{\sigma_c^2}{\sigma_s^2} (s_1 - s_2) \sim N\left(0, 2\sigma_s^2 \left(\frac{\sigma_c^2}{\sigma_s^2}\right)^2\right) = N\left(0, 2\frac{\sigma_c^4}{\sigma_s^2}\right) \quad (6)$$

and

$$\begin{aligned} \frac{\sigma_c^2}{\sigma_s^2} (s_1 - s_2) + \frac{\sigma_c^2}{\sigma_\phi^2} (\phi_1 - \phi_2) &\sim N\left(0, 2\sigma_s^2 \left(\frac{\sigma_c^2}{\sigma_s^2}\right)^2 + 2\sigma_\phi^2 \left(\frac{\sigma_c^2}{\sigma_\phi^2}\right)^2\right) \\ &= N\left(0, 2\frac{\sigma_c^4}{\sigma_s^2} + 2\frac{\sigma_c^4}{\sigma_\phi^2}\right) \quad (7) \end{aligned}$$

Since Equation (7) has a larger variance than Equation (6) and $c_2 - c_1 > 0$, it is immediately clear that $\Pr\left(\frac{\sigma_c^2}{\sigma_s^2} (s_1 - s_2) > (c_2 - c_1)\right) < \Pr\left(\frac{\sigma_c^2}{\sigma_\phi^2} (\phi_1 - \phi_2) + \frac{\sigma_c^2}{\sigma_s^2} (s_1 - s_2) > (c_2 - c_1)\right)$. Therefore $E\|F - C\| < E\|K - C\|$.

The result can be generalized for more than two horses. Indeed, Equation (5) holds for any possible pair of horses. With eight horses, $C = \{1, 2, 3, 4, 5, 6, 7, 8\}$, the order of the pair $\{3, 7\}$ or any other pair is more likely to be reversed in K than in F . It is then immediately clear that more positions will change between C and K than between C and F .

DOWNLOAD FULL VERSION HERE: <https://bettingbooks.org/12>

PART VI: Prediction Markets

Download More Sports Betting Books at: www.bettingbooks.org

DOWNLOAD FULL VERSION HERE: <https://bettingbooks.org/12>

This page intentionally left blank

Download More Sports Betting Books at: www.bettingbooks.org

CHAPTER 17

Index Betting for Sports and Stock Indices

John Haigh

University of Sussex, Falmer, Brighton, UK.

Leighton Vaughan Williams

Nottingham Business School, Nottingham Trent University, Nottingham, UK.

1. Background	358
2. How Index Betting Operates	360
3. Setting Spreads	362
4. Spreads in Performance Indices	365
5. Advantageous Bets	368
6. Regulation, Taxation, and Biases in Spread Betting Markets	372
<i>References</i>	374
Appendix A	377
Appendix B	378
Appendix C	380

Abstract

We investigate a relatively recent format developed in the UK for wagering on sports, that is, *index* or *spread* betting. It has parallels with commodity trading, where the difference between the buy and sell prices of a market-maker indicates the margin. In sports betting, the “commodity” may be the distance by which a racehorse is beaten, the number of goals in a soccer match—almost anything. Returns or losses to the bettor are defined by the difference between the traded figure and the actual outcome.

To illustrate, suppose the offered spread for the number of points scored by a basketball team is (89, 93). A bettor who expects a high score might *buy* (at 93) for, say, \$10 per point. If 100 points are scored, seven higher than the buy price, his profit is $10 \times 7 = \$70$; if 80 points are scored, his loss is $10 \times 13 = \$130$. A bettor expecting a low total might *sell* (at 89). At, say, \$20 per point, his profit is $20 \times 9 = \$180$ from 80 points, or his loss is $20 \times 11 = \$220$ with 100 points. The sell bettor’s loss is apparently unlimited, but there are ways to cap it.

We examine spread betting on sports and on financial variables. Topics include information efficiency, decision-making under uncertainty and opportunities for insider trading. We explore similarities and contrasts between different forms of betting, from both the market-makers’ and the bettors’ perspectives.

For the empirical researcher, a sports betting market gives a rapid accumulation of data through a sequence of short-lived markets with well-defined endpoints. Spread betting markets may well be populated by bettors whose motivations are similar to financial traders, as the stake at risk is normally much higher than with conventional bookmakers. Any evidence of inefficiency in betting markets may offer insights into financial markets.

1. BACKGROUND

In addition to the well-established pari-mutuel (or Tote) system of betting, and definite odds offered by bookmakers and on betting exchanges, another method of betting on horse racing and other sports events is UK *index* or *spread* betting. Despite the name similarity, spread (index) betting is far removed from the notion of “betting on the spread” for matches between two teams, common in the U.S.

In spread (index) betting in a sports context, the “commodity” that is the subject of the “trade” may be almost anything. Obvious examples include the winning margin in a soccer match, the number of victories in a season by a given team, the winning distance in a horse race, or the total number of points scored in a rugby match. Less obviously, it might be the sum of the times, in minutes, that goals are scored in a given soccer match, the number of races in a meeting before the favorite wins for the first time, an artificial Performance Index, with different rewards for different finishing positions in a horse race, even wildly speculative quantities, such as the number of seconds before the first throw-in in a soccer match.

For stock prices or stock indices, the commodity is usually the price of a stock, or the level of an index, at some specific time in the future, say on the third Friday of every third month. Spread betting enables a trader to take a position on the price of a stock without any need ever to trade the stock itself, or even to trade on an option on the stock.

Whatever the context, the market-makers set a spread interval for a specified quantity, and their clients are invited to *buy* at the top end of the spread, or *sell* at the bottom end. The outcome of the trade is the product of the size of the unit stake chosen by the client, and the difference between the actual outcome of the event in question and the level at which the trade is enacted. The width of the spread interval is set by the market-maker. Even a narrow spread may enable insider traders, who possess superior information to the market-maker, to make easy profits, but a wide spread will reduce trading volume, by discouraging noise or liquidity traders.

Jackson (1994) pointed out an important contrast between sports betting and commodity trading. In the latter, the day to day fluctuations in prices are generally small, and the actual commodity price, at the time of delivery, will be close to this range, catastrophes aside. In sports betting, outcomes far removed from the buy-sell trading range are frequent—for example, in betting on the total goals scored in a soccer match, the spread intervals offered are almost invariably close to (2.5, 2.8), but about 8% of matches have no goals, and totals of seven or more happen frequently. This volatility has implications for how firms and gamblers control their risks.

Whether in sport or in the stock market, spread betting is regulated by the Financial Services Authority. Profits are free from income or capital gains taxes, there are no dealing or commission charges, stamp duty, or other deductions. One consequence of the regulatory regime is that a spread betting transaction is a legally enforceable contract; bettors may incur heavy losses if they make a large error.

A gambler wishing to participate in spread betting must normally arrange a suitable amount of credit with a spread betting firm. Specific transactions are often conducted by telephone, with the calls recorded to settle any dispute. The firm may decline any bet a gambler proposes, or accept it only at a lower unit stake, or may ask the gambler to increase his line of credit before the bet can be accepted. Clients may open or close trades at any time in dealing hours, with instant execution.

A common use of this flexibility when trading in single shares is to make a bet that a price will fall, or rise, at the immediate cost, as a margin call, of say, 7.5% of the contract value. For example, if the buy price of a share is 500p, a gambler who expected the price to rise could decide to buy at the rate of £10 for each 1p movement in the price. His maximum loss, if the share became worthless, would be $500 \times £10$, but the margin call would be only 7.5% of this, or £375. Risk-averse traders have several alternative actions: they may consent to the quoted spread being widened, with the proviso that the trade be closed automatically if their losses reach some predetermined level; or they may simultaneously make a pair of trades, one which is tied to the particular share of interest, and the other offsets this by being tied to a different future index. (This “index combination bet” has the advantage that it has a lower margin requirement than a standard trade on a single share and lower dealing costs. Moreover, non-standard quantities can be traded, hence futures bets on single stocks can accurately be offset.)

The mirror image is when a trader places a *limit order*, either when making the bet or at any subsequent time, asking for the bet to be closed if profits ever reach a preordained level. In this way, the trader guards against subsequent market falls, but at the expense of foregoing further profits.

Index trading is also a means of hedging against a possible short-term fall in the market, which would reduce the value of a trader's portfolio. The trader can sell the futures index that is linked to his portfolio as a means of avoiding brokerage charges.

2. HOW INDEX BETTING OPERATES

Denote by X the quantity (goals, time, price) that is the subject of the transaction. The market-maker offers the *spread interval* (a, b) on the outcome of X , and the bettor has the option either to *buy* X at a chosen unit amount β ($\beta > 0$), or to *sell* X , again at a chosen unit amount β ($\beta > 0$). When the value of X is known, bets can be settled: the *buy* trader gains $\beta(X - b)$, while the *sell* trader gains $\beta(a - X)$; negative values correspond to losses.

Example 1. X is the number of minutes to the first goal in a soccer match, and the spread is $(35, 38)$. Barry buys X at £10/min, Sophie sells X at £5/min. If the first goal is scored after 48 min, Barry *wins* $\pounds(48 - 38) \times 10 = \pounds100$, while Sophie *loses* $\pounds(48 - 35) \times 5 = \pounds65$. Had the match ended goalless after 90 min, the value of X would be taken as 90, so Barry would then have won $\pounds(90 - 38) \times 10 = \pounds520$, while Sophie's loss would be $\pounds(90 - 35) \times 5 = \pounds275$.

If the commodity of interest is the value at a fixed date in the future of a continuously changing quantity, such as the level of a stock market index, or the price of an individual stock, the spreads offered by the firms will naturally be updated in line with these movements. Similarly, in many sporting events, the spreads offered will change as the event unfolds, to reflect both the current state of the game, and the reduced period of time left to affect the outcome. Making bets at such times is termed *betting in running*.

Example 2. Consider the situation from Example 1. Suppose that, after 30 min, no goals have been scored, and the spread now on offer for X is $(58, 61)$.

Barry could decide to *close his bet* by now selling X , at the same unit stake of £10/min, at the new sell price of 58. Whatever the eventual value of X , he would then make a profit of $\pounds(58 - 38) \times 10 = \pounds200$. Or, he might hedge his bets by selling X at £6/min; the net effect here is a guaranteed profit of $\pounds(58 - 38) \times 6 = \pounds120$, along with whatever is the outcome of the residual £4/min at the original buy price of 38, with the knowledge that X cannot be less than 30. He might even choose to increase his exposure and buy further units of X at the new price of 61, at some suitable unit stake.

Unless a goal is scored soon, Sophie is in danger of suffering a large loss. She could also close her bet by buying X at £5/min at the new buy price of 61, leading to a certain

loss of $\pounds(61 - 35) \times 5 = \pounds130$. She might prefer this to the possible loss of $\pounds(90 - 35) \times 5 = \pounds275$ if the score remained 0–0 for the rest of the game.

Some commodities, such as that in Examples 1 and 2, have natural lower and upper bounds, so the maximum possible gains and losses are both finite. At other times, one or other of these natural bounds is not present: for example, the total number of goals scored in a soccer game cannot be less than zero, but has no upper limit; and the *difference* in the scores of two teams is unbounded in both directions. The market-makers will wish to ensure that each bettor has the funds to meet any liability, so may introduce a *stop loss* for these unbounded quantities, by treating any value of X that exceeds D (say) as equal to D , and any value of X less than some C as equal to C . Especially in spread betting on financial markets, the firm may offer several different stop loss bounds, possibly with different spread intervals, to suit the tastes of different clients.

Example 3. To illustrate how the language of the stock market and the sports arena has become integrated, a bettor may be interested in the total number of goals scored in the soccer World Cup Finals. Perhaps the spread quoted is (80, 82), but an “80 call option” may also be offered at (3.5, 5), and a “79 put option” at (1.5, 3).

Hugh, who thinks the number of goals will be high, has the choice of the spread bet at $\pounds10$ per point at a price of 82, or the 80 call option at $\pounds10$ per point at a price of 5. The latter would cost $\pounds10 \times 5 = \pounds50$. If only 60 goals are scored, he would lose $\pounds10 \times (82 - 60) = \pounds220$ on the spread bet; but if he had bought the option, his loss is only the $\pounds50$ cost of the option—which he would not exercise. If 95 goals are scored, he would win $\pounds130$ on the spread bet; his net profit from the call option would be $\pounds100$, arising from the $\pounds150$ he gains by exercising his option to buy at 80, less the $\pounds50$ cost of the option. Effectively, buying the option gives insurance against a heavy loss, at the cost of reducing his potential profit.

Larry, who thinks the total goals will be low, has the option to sell at $\pounds10$ per point at a price of 80, or buy the put option at $\pounds10$ per point at a price of 3. If 60 goals are scored, he would make a 20 point, or $\pounds200$, profit from the spread bet; if he had bought the option, he would exercise it and make a $\pounds160$ net profit, arising as $19 \times \pounds10 = \pounds190$ from the spread bet, offset by the $\pounds30$ cost of the option. With 95 goals, he would lose just the $\pounds30$ cost of the option (and not exercise it), but his loss on the spread bet would be $\pounds10 \times (95 - 80) = \pounds150$.

What about the arithmetic from selling call or put options? If, in Example 3, Larry sold an 80 call option at $\pounds10$ per point at 3.5, the maximum profit he can make is the $\pounds35$ cost of the option, which he hopes not to be exercised. However, his loss is unlimited, as if more than 80 goals are scored, those who bought the call option from him will exercise it, and get $\pounds10$ for every goal in excess of 80. Similarly, if Hugh sold a 79 put option at $\pounds10$ per point at 1.5, he receives the $\pounds15$ cost, and has no liabilities so long as at least 79 goals are scored; but every goal less than 79 costs him $\pounds10$. From the spread betting firm’s perspective, matching buyers and sellers of the call and put options simply guarantees a profit proportional to the width of the spread on these options, just as it does on trades in the original spread of (80, 82)—see below.

3. SETTING SPREADS

A market-maker's ideal is a large profit, with zero or minimal risk. He selects the commodities on which to trade, and their spreads. We follow Haigh (1999). At the simplest level, with the notation of the previous section, suppose the total of the unit buy stakes is B , and the total of the unit sell stakes is S . The market-maker's profit is then

$$B(b - X) + S(X - a) \quad (1)$$

Hence, if buy and sell contracts match exactly, so that $B = S$, this profit is

$$B(b - a) = \frac{(B + S)(b - a)}{2}$$

whatever the outcome—the risk is zero. Market-makers have a powerful incentive to set spreads so that buy and sell propositions are equally attractive to bettors. And if the market-maker successfully judges the preferences of bettors as a whole in this manner, the width of the spread, $b - a$, plays a central role in maximizing profit. A narrow spread can be expected to attract higher total stakes, but the specific expression $B(b - a)$ shows that halving the spread width will not increase profits unless total unit stakes are more than doubled.

Even if no information of significance to the sporting event in question arises after spreads are first offered but before the event begins, firms will adjust their buy and sell prices in response to bets placed. The initial choice of spread will be informed by an assessment of the likely actions and expectations of the betting public, and also by the firm's objective assessment of the chances of the different outcomes. We assume that both spread setters and punters have equal access to relevant information.

Let the quantity X in Expression (1) above have mean μ and variance σ^2 . Then, if B and S are the total unit stakes for buy and sell at the spread (a, b) , Expression (1) shows that the mean profit to the firm is

$$B(b - \mu) + S(\mu - a) \quad (2)$$

and hence, so long as $a < \mu < b$, the mean profit is positive. Individual punters can be at an advantage if $\mu > b$ or $\mu < a$, but so long as B and S are such as to leave expression (2) positive, the overall advantage is with the firm. But if μ is outside the spread interval, this mean outcome could be negative. A firm may thus be expected to set the spread interval so as to contain μ . In particular, if $\mu = (a + b)/2$ so that the mean outcome is at the center of the spread interval, the mean profit is

$$\frac{(B + S)(b - a)}{2} \quad (3)$$

Expression (3) is also the actual profit when buy and sell bets attract equal support.

The variance of the profit, from (1), is $(B - S)^2 \sigma^2$, and so when $\mu = (a + b)/2$, the coefficient of variation of the profit can be written as

$$\frac{\sigma}{b - a} \times \frac{2|B - S|}{B + S} \quad (4)$$

which suggests that $(b - a)$, the *width* of the spread, might be expected to be proportional to σ . It is often the case that a firm is prepared to offer spread bets on some quantity both for individual sports games, and also on the total of this quantity over a collection of n games. Then, so long as the games are independent, and the variances in all games can reasonably be taken as equal, we might expect the spread width for bets on the total over n games to be about \sqrt{n} times as large as the spread width for an individual game.

Example 4. In American football, typical spreads for the *total points* scored in a single game are in the range (37, 40) to (42, 45), depending on the reputations of the teams for offense and defense capabilities. Almost invariably, spreads for a single game are 3 points wide. When a spread is offered for the total points scored over a weekend's fixtures of $n = 12$ games, it is commonly 10 points wide; these figures fit with the observation above, as 10 is quite close to $\sqrt{12} \times 3$.

Example 5. Between well-matched soccer teams at a professional level, little variation is seen in spreads offered for the time T of the first goal, typical spreads being in the range (34, 37) to (37, 40), suggesting an expected value of T of about 37 min. Associated with these bets on individual matches, we find spreads offered on the time until the first goal in any one of n specified matches (perhaps the $n = 10$ Premier League games to be played over one weekend, or the $n = 31$ games to be played in a particular tournament).

Empirical evidence (e.g., Dixon and Robinson, 1998) is that T can be well modeled as having an exponential distribution $E(\lambda)$, whose mean and standard deviation are both equal to $1/\lambda$ (ignoring the finite length of any game). A well-known property of the exponential distribution is that the *minimum* of n independent exponential variables with parameters $\{\lambda_1, \dots, \lambda_n\}$ is also exponential, with parameter $\lambda_1 + \dots + \lambda_n$. The similarity of the spreads typically offered suggests that taking all values of λ as equal is a reasonable step, in which case the mean and standard deviation of the time to the first goal across n matches are both $1/(n\lambda)$. (When n is large, the finite length of the game is no longer an issue.)

This indicates that we might expect the spread width for the time of the first goal in n matches to be about $1/n$ of the width of the spread for one match, typically 3 min. Data for the European Championship (1996), when 31 matches were played, show that the index firms were much more cautious than this; rather than a spread of about $180/31 \sim 6$ sec, the spreads were typically 15 sec wide.

As a separate point, we would also expect the mean time until the first goal in 31 matches to be about $1/31$ of the mean time to the first goal in a single match,

in this case 37/31 min, or about 72 sec. Because teams often begin matches rather cautiously, we might expect the center of the spreads to be a little later than this. But the actual initial spreads on offer for that tournament were centered around 240 sec, which is so much larger than 75 sec as to suggest that the index firms were not using this approach. (The fastest goal in that 1996 tournament actually occurred after only 115 sec.)

In betting on a pari-mutuel system, a fixed proportion of stakes is returned to the punters, so the market-makers carry no risk, and the punters know precisely the mean amount they would lose from random bets. In a race with n horses, in which bookmakers offer horse i at odds of $\alpha(i)$ to 1, the *overround* is defined at that quantity m such that

$$\sum_i 1/(1 + \alpha(i)) = 1 + m. \quad (5)$$

Bookmakers ensure that $m > 0$. One interpretation is that $1 + m$ is the smallest quantity that a punter can stake to guarantee a return of one unit, whatever the race outcome; another is that if these odds have been so carefully judged that the total stakes on horse i are proportional to $1/(1 + \alpha(i))$, then $m/(1 + m)$ is the fraction of all stakes that are retained by the bookmakers, whatever the race outcome. Bookmakers have no expectation that punters will behave as conveniently as in this ideal scenario, but the smaller the value of m , the larger the degree of risk the collection of odds represents. In general, there is no equivalent quantity within spread betting.

For example, if stop losses are not in operation, and a spread bet is made on a score *difference*, potentially unbounded in both directions, there is no maximum loss, so no definite sum can be said to be at stake.

However, in a two-horse race, in which X takes the values $R(R > 0)$ or 0, according to which horse wins, suppose the spread for one horse is (a, b) . Inevitably, the spread for the other is $(R - b, R - a)$, as selling one horse is here exactly the same as buying the other. Buying at price b is equivalent to betting at odds of $R - b$ to b , and buying at $R - a$ is equivalent to odds of a to $R - a$. This spread bet is identical to a set of odds with overround $(b - a)/R$.

Haigh (2000) has explored how this may extend to two-person contests in which X may take more than two different values. Spreads (a, b) and (c, d) are offered for the two contestants, and suppose the rewards are $R = R_1 > R_2 \geq \dots \geq R_{n-1} > R_n = 0$, paired so that if one contestant scores R then the other scores 0, and, if one scores R_i , the other scores some definite value, S_i , for each $i = 2, \dots, n - 1$.

In Appendix A, we show that, under very mild conditions, there is a set of odds corresponding to the n distinct outcomes, so that for each of the four possible spread bets, a combination of odds bets can be constructed that has exactly the same result as the spread bet; and these odds bets have overround $(b - a)/R$.

This extends the correspondence between spread bets and odds, but this second correspondence is one way only. Given a set of odds bets, it is not always possible

to find a collection of rewards for the different outcomes, along with spreads for the two contestants, such that a gambler could make one or more spread bets that give results identical to those for every odds bet.

But the correspondence does suggest how to make a comparison of the overall risks being taken when a variety of firms offer sets of odds or collections of spread bets in this sort of two-person contest. For any set of odds, the overround can be calculated as in Equation (5), and for any reward structure from a spread betting firm, calculating $(b - a)/R$. The smaller these quantities evaluate to, the larger the risk the firm is taking, and the more likely it is that a gambler can find a favorable bet.

Example 6. (Cricket, South Africa vs. England, Fifth Test Match, January 2000.) The reward structure from IG Index allocated 10 points to each team in a drawn match, otherwise the winner scored 25 points and the loser scored zero. Initial spreads of (16, 17.5) for South Africa and (6, 7.5) for England were offered.

Consider the set of odds 21:29 against a win for South Africa, 41:9 against a win for England, and 7:3 against a draw. The overround for these odds is $6\% = (17.5 - 16)/25$, as noted above. The four possible spread bets at unit stake could be replaced by the combinations of odds bets listed below. In each case, the spread bets and the odds bets give *identical* outcomes to the gambler, whatever the match result.

Buy South Africa, OR bet 14.5 on South Africa to win, bet 3 on a draw.

Sell South Africa, OR bet 4.5 on each of England to win, or a draw.

Buy England, OR bet 4.5 on England to win, 3 on a draw.

Sell England, OR bet 14.5 on South Africa to win, 4.5 on the draw.

Suppose a different firm offered rewards of 50 for a win, 30 for a draw, and 0 for a loss, with spreads (35, 37) and (16, 18). The calculation $(b - a)/R = 2/50 = 4\%$ indicates that this second firm has a slimmer margin than the 6% for the first—and hence may be more likely to yield a favorable bet to a punter.

If a firm offers odds against the three results of a win for South Africa, a win for England, or a draw, it would not be possible to construct a spread bet reward structure that would always give the same outcome as an odds bet.

4. SPREADS IN PERFORMANCE INDICES

Spread betting firms often approach multi-team contests via some Performance Index. In a horse race, or a team competition, rewards $R_1 \geq R_2 \geq \dots \geq R_n$ will be scored by the winner, second, and so on, respectively. For example, in horse racing, rewards of 50, 30, 20, and 10 points may be associated with the first four horses, the rest scoring zero. In a soccer league of 20 teams, rewards may range from 60 points to the champions down to 5 points for the eighth placed team, zero to the rest. In knockout competitions, decreasing rewards may be offered for the winner, losing finalist, both

losing semi-finalists, the four losing quarter-finalists, and so on. Bets based on the i th entrant are defined by the spread (a_i, b_i) .

The firms must ensure that certain necessary conditions are met, to avoid the possibility that a bettor can guarantee a profit, either by selling all the most expensive horses, or by buying the cheapest ones. In symbols, this translates to the firms ensuring that, for all $i = 1, 2, \dots, n$,

$$a_{(1)} + a_{(2)} + \dots + a_{(i)} \leq R_1 + R_2 + \dots + R_i \quad (6a)$$

$$b_{(n)} + b_{(n-1)} + \dots + b_{(i)} \geq R_n + R_{n-1} + \dots + R_i \quad (6b)$$

where the subscript (i) denotes the order statistic. There are further restrictions in knockout tournaments where the complete draw is known from the outset: for example, there are blocks of eight contestants, exactly one of whom will get through the first three stages; the sum of the rewards for the other seven is fixed, while the one contestant gets at least some minimum reward. Thus similar sets of conditions on these blocks of two, four, eight, and so on, contestants are required. We assume the spread betting firms have ensured that all these consistency conditions are met.

Using (6a) when $i = n$, and (6b) when $i = 1$, in particular we must have

$$\sum_{i=1}^n a_i \leq \sum_{i=1}^n R_i \leq \sum_{i=1}^n b_i \quad (6c)$$

One way to ensure this is to have $\sum (a_i + b_i)/2 = \sum R_i$, a condition automatically satisfied if the spread for each contestant is centered at the contestant's mean reward. We have already noted the advantages to firms of taking this precaution, but they will also have regard to the expected weight of bets: this may incline them to offer a spread that does not include this mean value if, by doing so, buy and sell liabilities on that contestant become more closely matched. Whatever the reason, if one contestant attracts so many buy bets that the firm moves its spread interval upwards, it is reasonable to expect compensating downwards movements in other spread intervals, resulting in the sum of the mid-points of all the spreads remaining fairly close to $\sum R_i$.

Suppose that this condition $\sum (a_i + b_i)/2 = \sum R_i$ is satisfied, and that the maximum reward R_1 is given to one contestant only. In Appendix B, we show that, provided some (usually undemanding) technical conditions are satisfied, there is a set of odds bets against each runner winning, with an overround of $m = \sum (b_i - a_i)/[2(R_1 - R_n)]$, which correspond (in the sense described in Appendix B) to the spread bets actually on offer. There may be many collections of odds that have this correspondence, but the overround for each of them is the same.

Example 7. In the 1998 Super Bowl, the winner would be awarded 100 points, the runner-up 75 points, and the losing semi-finalists 50 points each. Green Bay played San Francisco, Denver played Pittsburgh, and the spreads were (73, 76), (70, 73), (64, 67) and (62, 65), respectively. The mid-points of these spreads do sum to 275, the

sum of the rewards. The technical condition here is that each team's winning chance should exceed 3%, as should each team's chance of elimination at the semifinal stage—surely met! Our result is that these spreads correspond to an overround of $(4 \times 3)/(2 \times 50) = 12\%$.

For the final, Green Bay was initially quoted at (94.5, 96.5), with Denver at (78.5, 80.5), which, in this two-horse race, exactly corresponds to odds of 7:43, and 39:11. These odds have an overround of 8%, that is, $(2 \times 2)/(2 \times 25)$.

Even when some of the conditions of the argument in Appendix B are not satisfied, we can still calculate the quantity $m = \sum (b_i - a_i)/[2(R_1 - R_n)]$, and use it to compare sets of spread bets from different Performance Indices on the same event. We outline the circumstances where some of these conditions might not be met.

1. For obvious reasons, the firms use rounded values for the spreads $[(a, b)]$, so they can only be at best an approximation of more exactly calculated numbers. Even if the firms have completely followed the steps described in Appendix B, the quoted spreads may thus not exactly correspond to these calculations.
2. The psychology of betting means that many gamblers will wish to support one particular runner, rather than specifically oppose that runner, and, indeed, we frequently find that $\sum b_i - \sum R_i > \sum R_i - \sum a_i$. In such cases, the sell bets, as a whole, offer better value to the bettor than the buy bets, and the condition $\sum (a_i + b_i)/2 = \sum R_i$ fails. However, by using an appropriate scaling, we can justify parallel steps in the argument of Appendix B, leading to the same value for m .
3. Appendix B requires that the winning chance for a contestant with spread (a, b) be at least $(b - a)/[2(R_1 - R_n)]$. This is not a stringent requirement: for example, in a common Performance Index for a 10-horse race, $R_1 - R_n = 50$, and the spread for the weakest runner is often at most 1.5 units wide: this condition asks only that its winning chance be at least 1.5%. But plainly, this condition will sometimes fail. Note there is also the parallel requirement that the chance that even the best horse scores minimal points shall be at least

$$(b - a)/[2(R_1 - R_n)].$$

4. Suppose a contestant is so strongly fancied that the offered spread takes the form (a, R) . A firm would be in great difficulty in constructing an equivalent spread (needed in the argument in Appendix B) for the artificial bet that this contestant will *not* finish last, but would readily be able to give a corresponding spread for a bet on whether the contestant would win, or not. Since this last consideration is the only construction that is needed in the calculation of the overround m , that calculation can still proceed. However, if some contestant is so weak that its spread has the form $(0, b)$, it will not be possible to suggest a meaningful equivalent spread for the simple wager that the contestant will *fail to win*, so here one of the steps needed in Appendix B cannot be completed. We face similar difficulties if some spreads are *very* wide.

5. ADVANTAGEOUS BETS

In the notation of Section 3 above, a buy bet is favorable when $\mu > b$, and a sell bet is favorable when $\mu < a$. We will concentrate on the first case, noting that parallel considerations will apply in the second. Kelly (1956) considered how much a gambler should stake when a series of bets at favorable odds are on offer, and his work has been refined by successive investigators. (The account here is based on Haigh, 2000.)

The essence of Kelly's argument is that, in order to maximize the safe long-term growth rate of his fortune, the gambler should stake that fraction of his current fortune that corresponds to the size of his advantage. In spread betting, one immediate problem is to decide what is meant by the size of a stake; if the quantity of interest, X , is unbounded below when a gambler makes a buy bet, the potential loss is unlimited. To avoid this, we assume that X is bounded below, and we make this lower bound essentially zero by requiring that

1. $\Pr(X < 0) = 0$
2. For every $\varepsilon > 0$, $\Pr(X < \varepsilon) > 0$.

Suppose a bettor has initial capital V_0 , and fortune V_n after n bets. These bets are on the outcomes of a sequence of independent random variables $\{X_k : k \geq 1\}$, all distributed as the generic variable X . The bettor must select the unit stakes, $\{\beta_k : k \geq 1\}$, for a sequence of buy bets. The maximum possible loss on the k th bet is $\beta_k b$, so write $\beta_k = u_k V_k / b$, where $0 < u_k \leq 1$. Then plainly

$$\frac{V_n}{V_0} = \prod_{k=1}^n [1 + (X_k - b) u_k / b] \quad (7)$$

and the mean growth rate of the bettor's fortune after n bets is $G_n = \frac{1}{n} \ln \left(\frac{V_n}{V_0} \right)$.

Kelly's criterion of maximizing this long-term expected growth rate is not the only way to reach a decision. But it has the added attraction of leading to the optimal bet under other criteria, too. Suppose the bettor always risks the same proportion $u_k = u$ of his current fortune, and that $G(u) = E \{ \ln[1 + (X - b)u/b] \}$ is finite. Then the Strong Law of Large Numbers shows that $G_n \rightarrow G(u)$, almost surely. To maximize this long-term growth rate, we seek solutions of

$$G'(u) = E \left[\frac{X - b}{b + u(X - b)} \right] = 0. \quad (8)$$

Finkelstein and Whitley (1981) showed that there is some essentially unique choice $u = u^*$ that achieves this maximum. They also showed that, if V_n^* is the fortune after n plays using this fixed choice u^* , and V_n corresponds to any other permissible choice

of strategy, including allowing $\{u_k\}$ to vary, then using u^* each time is superior in the following senses:

1. V_n/V_n^* is a *supermartingale*, that is, $E(V_n/V_n^*|V_i/V_i^* \text{ for } i < n) \leq V_{n-1}/V_{n-1}^*$. Also $E(V_n/V_n^*) \leq 1$, V_n/V_n^* converges almost surely to some finite value, and $E[\lim(V_n/V_n^*)] \leq 1$.
2. V_n^*/V_n is a *submartingale*, that is, $E(V_n^*/V_n|V_i/V_i^* \text{ for } i < n) \geq V_{n-1}^*/V_{n-1}$. Also $E(V_n^*/V_n) \geq 1$, V_n^*/V_n converges almost surely to some real number (or to $+\infty$), and $E[\lim(V_n^*/V_n)] \geq 1$.

Ethier and Tavaré (1983) considered the behavior of this value u^* when the advantage $\mu - b = \varepsilon$ is small. By an expansion of Equation (7), they showed that

$$u^* \approx b\varepsilon/\sigma^2 \quad (9)$$

and that the corresponding growth rate is then

$$G(u^*) \approx \varepsilon^2/\sigma^2 \quad (10)$$

that is, the growth rate achieved by using the Kelly criterion is directly proportional to the *square* of the size of the punter's advantage, and inversely proportional to the variance of the quantity under consideration.

When there are several apparently favorable spread bets to choose between, and the advantages are small, the expression in Equation (10) could be a basis for selecting one of them.

Example 8. Suppose, as Dixon and Robinson (1998) indicate, it is reasonable (to a first approximation) to take the process of scoring goals at soccer as a homogeneous Poisson process of rate λ . Spread bets are offered on both X , the total number of goals in the match, and on T , the time until the first goal. Let $F (= 90)$ denote the duration, in minutes, of the match, so that T is formally recorded as F if the match is scoreless.

Then X has a Poisson distribution with parameter λF , and T has an exponential distribution, truncated at the value F . Then, with $\alpha = \exp(-\lambda F)$, the mean and variance of T are $(1 - \alpha)/\lambda$ and $(1 - \alpha^2)/\lambda^2 - 2F\alpha/\lambda$. If the spreads offered are consistent with the firm having underestimated λ , the buy price, b , for X will be too low, and the sell price, s , for T will be too high. On the basis of Equation (9), buying X is preferable to selling T , if and only if

$$(\lambda F - b) \sqrt{1 - \alpha^2 - 2F\alpha\lambda} > (\lambda s - 1 + \alpha) \sqrt{\lambda F}. \quad (11)$$

Spreads of (2.5, 2.8) for X and (35, 38) for T are common in soccer. Because goal times are rounded to the minute in which a goal occurs, the latter spread represents a width of *four* minutes. If the actual mean number of goals expected in the match is $\lambda F = 3$, then the left side of this inequality is 0.167 and the right side is 0.317. With these figures, selling T is the more attractive bet.

For a different consideration, suppose the distribution of X is governed by a parameter θ . We refer to $B_X = [\theta: a < E(X) < b] = (A, B)$, say, as the *firm's interval* for the spread (a, b) . Regarding θ as a random variable with distribution function $H(\theta)$, the difference $H(B) - H(A)$ is then the probability that both buy and sell bets favor the firms, and so measures their unwillingness to be exposed to risk of loss. If we are ignorant of $H(\theta)$, we can use the *widths* of the respective firm's intervals as proxies to compare spread bets offered on two different variables.

Example 9. In games such as basketball, American football, or rugby, the number of points scored by a team might reasonably be modeled by a normal distribution, with some particular mean and variance (see, e.g., Stern, 1994). Spreads are offered on both N , the total number of points scored in a match, and on S , the points superiority of one team over the other. Almost invariably, the widths of these two spreads are equal.

The values of N and S are governed by the scoring rates of each team, which in some circumstances would be expected to have a positive correlation (e.g., both are affected equally by weather and/or pitch conditions; if one team takes risks to score more points, that usually gives point-scoring opportunities to the opponents, etc.). Here, N would have a larger variance than S . Hence, the probability associated with an interval of given width around the mean value of N would be smaller than that for a similar interval of the same width for S . A punter is more likely to find a favorable bet on N .

Sackrowitz (2000) suggests that in a game of American football, there are about two dozen possessions during a game. The more possessions one team has, the fewer the other will have; in such circumstances, a negative correlation between the teams' scores might be expected. Here, S would have the larger variance, and so S would be more likely to be a source of a favorable bet.

Within the same match, some bets may be so closely associated that a direct comparison of the respective firms' intervals can be made.

Example 10. (Soccer, Barnsley vs. Stockport, September 10, 1999.) The firm William Hill offered, among other bets:

1. Total goals in the game, (2.7, 3.0).
2. Total goal minutes (i.e., sum the times at which all goals are scored), (135, 145).

Let's analyze this bet. The ratio of the mid-points of the two spreads, $140/2.85 = 49.12 = m$, is an estimate of the average time, in minutes, at which a randomly chosen goal is scored. The firm's interval for the scoring rate in (2) corresponds to an expected number of goals in the range $(134.5/m, 145.5/m) = (2.74, 2.96)$, which is contained within the spread for (1). *A bettor, buy or sell, can expect better value from bets on (2) than on (1).*

The notion of a firm's interval easily extends to more than one parameter.

Example 11. Consider the South Africa–England test match from Example 6. Writing $p = \Pr(\text{South Africa win})$ and $q = \Pr(\text{England win})$, so that $1 - p - q$ is the chance

of a draw, the mean scores to each team are $10 + 15p - 10q$ and $10 + 15q - 10p$. Let R be the region (a parallelogram) within the set $S = \{(p, q) : p \geq 0, q \geq 0, p + q \leq 1\}$ defined by $6 < 15p - 10q < 7.5$ and $-4 < 15q - 10p < -2.5$. So long as the point (p, q) is within R , all bets are unfavorable; thus R would be the *firm's region* for the spreads listed earlier.

At the same time, another firm, Sporting Index, offered spreads of (15, 16.5) and (6.5, 8) with the same reward structure. That firm's region is defined by $5 < 15p - 10q < 6.5$ and $-3.5 < 15q - 10p < -2$ (and $p \geq 0$). The small overlap in these regions suggests that most bettors could locate some favorable bet with one firm or the other.

One striking conclusion drawn by Kelly (1956) was that the optimal overall bet on a horse race with conventional odds might include some bets that were individually disadvantageous. Once this has been mathematically established, an intuitive explanation offers itself: the inclusion of such bets raises the total amount staked, so this increased chance of including the winning horse can enable a larger fraction of the gambler's current fortune to be actively used. It is thus not surprising that the same counterintuitive result also holds for spread bets on Performance Indices, where bets specifically *against* individual horses can be made.

A bettor whose current fortune is F will buy horse i at unit stake $u_i F/b_i$, and sell horse i at unit stake $v_i F/(R_1 - a_i)$, for $i = 1, 2, \dots, n$. It would be absurd to buy and sell the same horse, so at least one of u_i and v_i will be zero every time. Let $w = 1 - \sum_i u_i - \sum_i v_i$. The outcomes are correlated, so we need not insist that $w \geq 0$ to ensure that the bettor's fortune cannot become negative.

Provided the mean value of the reward for some horse lies outside that horse's spread interval, a favorable bet will exist. Identifying the *optimal* combination bet in general is described in Appendix C, which also includes an algorithm for the special case where the winner of a race scores R , and all others score zero.

Example 12. Suppose we have the position shown in Table 1, showing possible win probabilities and their respective spreads for a race with seven horses, with one reward $R = 100$. Use the steps in Appendix 3 to follow this example.

Using the notation of Appendix C, $X_1 = \{1, 2, 3\}$ and $Y_1 = \{6, 7\}$ are both non-empty, so that to buy any of $\{1, 2, 3\}$ or to sell either of $\{6, 7\}$ is a favorable bet. In Step 2, $t_{11} = 25/23$, but Step 3.1 fails for horse 3. Taking $X_u = \{1, 2\}$ and $Y_v = \{6, 7\}$ leads to $t_{uv} = 15/14$; here all conditions in Step 3 hold, so this is a possible optimum and $G_{uv} = 0.069988239$. With the same X_u , let $Y_v = \{5, 6, 7\}$; then $t = 25/23$ (again), and this time the conditions in Step 3 hold, giving $G_{uv} = 0.070045508$. Several other

TABLE 1 Data for Example 12

Horse	1	2	3	4	5	6	7
Win probability	0.25	0.05	0.20	0.05	0.20	0.15	0.10
Spread	(12, 14)	(2, 4)	(17, 19)	(3, 5)	(19, 21)	(25, 27)	(15, 17)

choices for $\{X_u, Y_v\}$ satisfy Step 3, but none of them gives a higher value for G_{uv} . The (unique) optimal bet uses $u_1 = 0.0978$, $u_2 = 0.00652$, $v_7 = 0.3572$, $v_6 = 0.3652$, and $v_5 = 0.0278$. Notice that it does *not* include the favorable option to buy horse 3, and *does* include the unfavorable option to sell horse 5.

Some commentators have pointed out that the size of stake indicated by a strict adherence to Kelly's formula may cause justifiable concern to many bettors. In Example 12, the optimal strategy calls for a bet that uses over 85% of the available capital. Thorp (1997, 2006) noted that although the Kelly strategy in a favorable game would maximize the probability that the gambler's fortune ever exceeded some given level before a given time, the probability that the gambler's fortune ever fell below some specified level during that time could also be uncomfortably high. He gave examples of "fractional Kelly" strategies, where the bet size is a fixed fraction of what strict Kelly would give. MacLean et al. (1992) also explored this alternative: they concluded that a gambler need forego only a small part of the expected growth rate of his fortune in order to have a greatly increased protection against substantial short-term loss.

6. REGULATION, TAXATION, AND BIASES IN SPREAD BETTING MARKETS

Spread betting markets often operate alongside fixed odds markets, so that gamblers may place bets about the same event in the two markets simultaneously. We have noted that spread betting in the UK is regulated by the Financial Services Authority, and hence is subject to the 1986 Financial Services Act, while fixed odds betting was regulated by the 1963 Betting, Gaming, and Lotteries Act (as subsequently amended) until the introduction of the 2005 Gambling Act.

Here we first consider the issue of the differing regulatory structures on the incidence of insider trading in the two markets, using data from 257 soccer matches from the 1996–1997 season. To do this, Paton et al. (1999) used the model developed by Shin (1993) to estimate the percentage of bettors with inside information in the market for football bets. Shin's model links the percentage of money wagered by insiders, z , to the bookmakers margin, or overround, calculated as in Equation (5) above.

By using the nonlinear estimation technique developed by Jullien and Salanié (1994), Paton et al. conclude that the level of insider trading is about 3.12% in the fixed odds market, and 1.51% in the spread markets. This difference may well be attributable to the more rigid regulatory framework applied to spread betting markets, as this may impose a greater cost on insider activity.

There is an array of evidence, for example, Thaler and Ziemba (1988), Vaughan Williams and Paton (1997), Sauer (1998), Vaughan Williams (1999), Cain et al. (2000), Deschamps and Gergaud (2007), and Sung and Johnson (2007), to support the contention that fixed odds markets are subject to systematic biases. The most notable of these is the favorite-longshot bias, the observed tendency for the expected return to bets placed at lower odds to exceed that of bets placed at higher odds.

Paton and Vaughan Williams (1998) compared the bias in fixed odds and spread betting markets, at different odds levels, using data from 265 soccer matches in the English Premier League over 1996–1997. For fixed odds, each match provides three data values, one each for a win for the favorite, a draw, and a win for the longshot. Taking a unit stake on each bet, the return is -1 for a losing bet, or $+FIXED$ for a winning bet at odds $FIXED:1$; the best-fitting linear regression is

$$RETURN = -1.874 + 0.065 \times FIXED$$

The equivalent spread bet is a supremacy bet. Returns to a unit spread bet are calculated as (actual supremacy–BUY) for buy bets and as (SELL–actual supremacy) for sell bets.

For fixed-odds bets, there are three match odds: favorite win, longshot win, and draw. These are pooled for each match and the following regression is performed:

$$RETf = a + a1.FIXED + u$$

Because returns are truncated at -1 , this will lead to biased results using OLS. For this reason, Tobit estimation is employed.

Spread bets can be divided into BUY and SELL bets. SELL bets are multiplied by -1 , and the data are pooled to give two observation points for each match. The following regression is now performed:

$$RETs = b + b1.BUY \text{ [or SELL]} + v$$

As both returns and losses are open-ended, OLS estimation is used. The results of the analysis confirm the conventional favorite-longshot bias for fixed-odds bets, but there is no significant bias found for spread bets.

A possible explanation for the difference is that transaction costs (implied in the overround) were significantly lower for spread bets than fixed-odds bets (as is the case with person-to-person betting exchange markets, with similar results—see Smith et al., 2006). Spread bets also faced (at that time) a lower tax incidence than fixed-odds bets. In particular, tax on spread bets was levied only on the marginal trading unit, or tick, and constituted a much lower proportion of the overall trading gains and losses than did the tax on fixed-odds bets. There is a theoretical case for believing that costs of trading may explain, at least in part, the existence of a favorite-longshot bias. To an extent, the study supports this theory.

More recent work that considers the forecasts and efficiency of spread betting markets includes Simmons et al. (2003) for rugby league games, Twomey (2005) for horse racing, and Vaughan Williams (2005) for political elections.

In the 2001 UK Budget, the taxation of spread betting (and fixed odds betting) was changed from a tax on turnover to a tax on gross profits (revenue minus payout). A switch to a rate of tax on gross profits that was, say, revenue neutral for fixed odds would, however, have suddenly and dramatically increased the tax burden on the spread companies. For this reason, tax rates of 10% for sports spread betting and 3% for

financial spread betting were implemented, compared with a proposed rate of 15% for fixed-odds betting. The lower rate for financial spreads compared to sports spreads is related to the substantial hedging costs that are characteristic of financial spread betting. The implication of these costs is that the apparent gross profit to the company in the financial spread sector might substantially overstate its true profit. A single tax rate on gross profits across all sectors might therefore, have seriously disadvantaged financial spreads. These and related betting taxation issues are considered further in Paton et al. (2000, 2003, 2004).

The potential for profitable exploitation of biases in spread markets at current tax levels was explored by Vaughan Williams (2000b, 2000c, 2002), following on from a study of the performance of forecasting services in fixed-odds markets (Vaughan Williams, 2000a).

The specific area of interest in the latest analysis is those cases where the spread offered by one market-maker is out of line with that offered by others. In particular, where one spread deviates in its entirety from another it is possible, in theory, to trade with both and make a risk-free profit, a so-called arbitrage. In practice, such circumstances are rare, or employed as a marketing exercise, and confined to small stakes. More commonly, the top end of one spread is coincident with the lower end of another, or the spread interval offered by at least one market-maker does not contain the mean mid-point of the spreads offered by all market-makers collectively. Vaughan Williams (2001) terms these two situations as quasi-arbitrages (simple quarbs and full quarbs, respectively).

In Vaughan Williams (2000b, 2000c), the outlying position is defined as the mid-point of the quote offered by the market-maker most out of line with the average mid-point of quotes. A set of such quotes was collected for the eight months from August 1999 to April 2000, in the market offered about disciplinary bookings in the English Premier League (soccer). In this market, 25 points are awarded for a red card (a player dismissed from the field of play by the referee) and 10 points for a yellow card (a player officially cautioned).

The results suggest that the mid-point of all quotes is a better forecast of the actual outcome in the bookings market than is the mid-point of the spread offered by the market outlier, a finding duplicated in Paton and Vaughan Williams (2005; see also Smith et al. 2005, for a consideration of quasi-arbitrage opportunities in fixed-odds betting markets). Indeed, it would have been possible in these samples to have made positive profits by trading upon a strategy based on exploiting these differentials. In view of the limited sample size, however, these findings should be treated with caution.

References

- Cain, M., D. Law, and D. Peel. 2000. The Favourite-Longshot Bias and Market Efficiency in UK Football Betting, *Scottish Journal of Political Economy* 47(1), 25–36.
- Deschamps, B., and O. Gergaud. 2007. Efficiency in Betting Markets: Evidence from English Football, *The Journal of Prediction Markets* 1(1), 61–73.
- Dixon, M. J., and M. E. Robinson. 1998. A Birth Process Model for Association Football Matches, *The Statistician* 47(3), 523–538.

- Ethier, S. N., and S. Tavaré. 1983. The Proportional Bettor's Return on Investment, *Journal of Applied Probability* 20(3), 563–573.
- Finkelstein, M., and R. Whitley. 1981. Optimal Strategies for Repeated Games, *Advances in Applied Probability* 13(2), 415–428.
- Haigh, J. 1999. (Performance) Index Betting and Fixed Odds, *The Statistician* 48(3), 425–434.
- Haigh, J. 2000. The Kelly Criterion and Bet Comparisons in Spread Betting, *The Statistician* 49(4), 531–539.
- Jackson, D. A. 1994. Index Betting on Sports, *The Statistician* 43(2), 309–315.
- Jullien, B., and B. Salanie. 1994. Measuring the Incidence of Insider Trading: A Comment, *Economic Journal* 104, 1418–1419.
- Kelly, J. L. Jr. 1956. A New Interpretation of Information Rate, *Bell Systems Technical Journal* 35, 917–926.
- Maclean, L. C., W. T. Ziemba, and G. Blazenko. 1992. Growth Versus Security in Dynamic Investment Analysis, *Management Science* 38, 1562–1585.
- Paton, D., and L. Vaughan Williams. 1998. Do Betting Costs Explain Betting Biases? *Applied Economics Letters* 5, 333–335.
- Paton, D., L. Vaughan Williams, and S. Fraser. 1999. Regulating Insider Trading in Betting Markets, *Bulletin of Economic Research* 51(3), 237–241.
- Paton, D., D. Siegel, and L. Vaughan Williams. 2000. *An Economic Analysis of the Options for Taxing Betting: A Report for HM Customs and Excise*.
- Paton, D., D. Siegel, and L. Vaughan Williams. 2002. A Policy Response to the E-Commerce Revolution: The Case of Betting Taxation in the UK, *Economic Journal* 112, 296–314.
- Paton, D., D. Siegel, and L. Vaughan Williams. 2003. The Demand for Gambling, in L. Vaughan Williams (ed.), *The Economics of Gambling*. Routledge, London, pp. 247–263.
- Paton, D., D. Siegel, and L. Vaughan Williams. 2004. Taxation and the Demand for Gambling: New Evidence from the United Kingdom, *National Tax Journal* 57(4), 847–861.
- Paton, D., and L. Vaughan Williams. 2005. Forecasting Outcomes in Spread Betting Markets: Can Bettors Use “Quarbs” to Beat the Book? *Journal of Forecasting* 24(2), 139–154.
- Sackrowitz, H. 2000. Refining the Points-After-Touchdown Decision, *Chance* 13(3), 29–34.
- Sauer, R. D. 1998. The Economics of Wagering Markets, *Journal of Economic Literature*, 36, 2021–2064.
- Shin, H. S. 1993. Measuring the Incidence of Insider Trading in a Market for State Contingent Claims, *Economic Journal* 103, 1141–1153.
- Simmons, R., D. Forrest, and A. Curran. 2003. Efficiency in the Handicap and Index Betting Markets for English Rugby League, in L. Vaughan Williams (ed.), *The Economics of Gambling*. Routledge, London, pp. 114–134.
- Smith, M., D. Paton, and L. Vaughan Williams. 2005. An Assessment of Quasi-Arbitrage Opportunities in Two Fixed-Odds Horse Race Betting Markets, in L. Vaughan Williams (ed.), *Information Efficiency in Financial and Betting Markets*, Cambridge University Press, Cambridge, 2005, pp. 159–171.
- Smith, M., D. Paton, and L. Vaughan Williams. 2006. Market Efficiency in Person-to-Person Betting Exchanges, *Economica* 73(292), 673–689.
- Stern, H. 1994. A Brownian Motion Model for the Progress of Sports Scores, *Journal of the American Statistical Association* 89(427), 1128–1134.
- Sung, M., and J. E. V. Johnson. 2007. The Influence of Market Ecology on Market Efficiency: Evidence from a Speculative Market, *Journal of Gambling Business and Economics* 1(3), 185–198.
- Thaler, R., and W. T. Ziemba. 1988. Parimutuel Betting Markets: Racetracks and Lotteries, *Journal of Economic Perspectives* 2, 161–174.
- Thorp, E. O. 1997. The Kelly Criterion in Blackjack, Sports Betting and the Stock Market. *The Tenth International Conference on Gambling and Risk-Taking*.
- Thorp, E. O. 2006. The Kelly Criterion in Blackjack, Sports Betting and the Stock Market, in *Handbook of Asset and Liability Management, Volume 1: Theory and Methodology*, S. A. Zenios and W. T. Ziemba (eds.), North Holland, Amsterdam, pp. 385–428.
- Twomey, P. M. 2005. Market Efficiency of the 50-30-20-10 Horse Race Spread Betting Market, in L. Vaughan Williams (ed.), *Information Efficiency in Financial and Betting Markets*. Cambridge University Press, pp. 277–286.

- Vaughan Williams, L., and D. Paton. 1997. Why Is There a Favourite-Longshot Bias in British Racetrack Betting Markets? *Economic Journal* 107(1), 150–158.
- Vaughan Williams, L., and D. Paton. 1998. Why are Some Favourite-Longshot Biases Positive and Others Negative? *Applied Economics* 30(11), 1505–1510.
- Vaughan Williams, L. 1999. Information Efficiency in Betting Markets: A Survey. *Bulletin of Economic Research* 51(1), 1–30.
- Vaughan Williams, L. 2000a. Can Forecasters Forecast Successfully? Evidence from UK Betting Markets, *Journal of Forecasting* 19, 505–515.
- Vaughan Williams, L. 2000b. Arbs, Quarbs and Market Efficiency: Some Findings. *NTU Occasional Papers in Economics*, 2000/7.
- Vaughan Williams, L. 2000c. Index Investment Markets and Information Efficiency: Some Evidence from the UK. *Global Business and Economics Review—Anthology 2000*, pp. 24–29.
- Vaughan Williams, L. 2001. Can Bettors Win? A Perspective on the Economics of Betting. *World Economics* 2(1), 31–48.
- Vaughan Williams, L. 2002. *Betting to Win: A Professional Guide to Profitable Betting*, High Stakes, London.
- Vaughan Williams, L. 2005. Models, Markets, Polls and Pundits: A Case Study of Information Efficiency, in L. Vaughan Williams (ed.), *Information Efficiency in Financial and Betting Markets*. Cambridge University Press, Cambridge, UK, pp. 193–214.

APPENDIX A

Consider a two-person contest with possible rewards $R = R_1 > R_2 \geq \dots \geq R_{n-1} > R_n = 0$, paired so that if one contestant scores R_i , then the other scores 0, and, if one scores R_i , then the other scores some definite value, S_i , for each $i = 2, 3, \dots, n-1$. Spreads (a, b) and (c, d) are offered for the two contestants.

Suppose that $b - a = d - c$, and that there exist values $\{\beta_1, \dots, \beta_n\}$ with $0 < \beta_i < 1$ with $\sum_{i=2}^{n-1} \beta_i R_i < b$, $\sum_{i=2}^{n-1} \beta_i (R - R_i) < R - a$ and $\sum_{i=2}^{n-1} \beta_i (R - R_i - S_i) = R - b - c$. Then there is a set of odds $\{(\alpha_i : 1) : i = 1, 2, \dots, n\}$ corresponding to the n distinct outcomes, so that for each of the four possible spread bets, a combination of odds bets can be constructed that have exactly the same result as the spread bet; and these odds bets have overround $(b - a)/R$.

Proof

Suppose odds of $\alpha_i : 1$ were to be offered against A obtaining R_i (and B obtaining S_i), for $i = 1, 2, \dots, n$. Then the outcome of a buy bet on A at unit stake will be the same as a series of stakes of w_i on outcome i for $i = 1, 2, \dots, n-1$, provided that $b = w_1 + \dots + w_{n-1}$ and, for all these values of i , $(\alpha_i + 1)w_i = R_i$.

Similarly, stakes x_i on outcome i for $i = 2, 3, \dots, n$, with $R - a = x_2 + x_3 + \dots + x_n$ and $(\alpha_i + 1)x_i = R - R_i$ are equivalent to a unit sell bet on A . For the buy bet on B , we need stakes y_i with $(\alpha_i + 1)y_i = S_i$ and $d = y_2 + \dots + y_n$. Finally, for sell bets on B , we need stakes z_i with $z_1 + \dots + z_{n-1} = R - c$ and $(\alpha_i + 1)z_i = R - S_i$. We must check that these conditions can simultaneously hold, with non-negative stakes and odds.

Write $\beta_i = 1/(\alpha_i + 1)$. Given positive values for $\beta_2, \dots, \beta_{n-1}$, choose $w_i = \beta_i R_i$ for $i = 2, 3, \dots, n-1$, and then let $w_1 = b - (\beta_2 R_2 + \dots + \beta_{n-1} R_{n-1})$. Let $\beta_1 = w_1/R_1$. Since $z_1 = \beta_1 R_1$, we must take $z_1 = w_1$ and then, for $i = 2, \dots, n-1$, we have $z_i = \beta_i (R - S_i)$. With these choices, $\sum_{i=1}^{n-1} z_i = b + \sum_{i=2}^{n-1} \beta_i (R - R_i - S_i)$, so we shall require $\sum_{i=2}^{n-1} \beta_i (R - R_i - S_i) = R - b - c$. Thus we need $\beta_n = [R - a - \sum_{i=2}^{n-1} \beta_i (R - R_i)]/R$ and, for $i = 2, 3, \dots, n$, write $x_i = \beta_i (R - R_i)$. For $i = 2, \dots, n$, let $y_i = \beta_i S_i$; we need $d = \sum_{i=2}^n y_i = \sum_{i=2}^n \beta_i S_i$, which means that

$$\sum_{i=2}^{n-1} \beta_i (R - R_i - S_i) = R - a - d.$$

These conditions can be met simultaneously for genuine odds bets so long as $b - a = d - c$, $\sum_{i=2}^{n-1} \beta_i (R - R_i - S_i) = R - b - c = R - a - d$, and the inequalities

$$\sum_{i=2}^{n-1} \beta_i R_i < b \quad \sum_{i=2}^{n-1} \beta_i (R - R_i) < R - a$$

can both be satisfied. When this occurs, the overround of this set of odds is

$$\sum_{i=1}^n \beta_i - 1 = \sum_{i=2}^{n-1} \beta_i + \left\{ b - \sum_{i=2}^{n-1} \beta_i R_i \right\} / R$$

$$+ \left\{ R - a - \sum_{i=2}^{n-1} \beta_i (R - R_i) \right\} / R - 1 = (b - a) / R.$$

There may be many sets of odds that yield bet combinations that can be exactly matched with the four spread bets on offer, but all have this same overround.

APPENDIX B

Suppose a race with n horses has a Performance Index with K distinct rewards $R_1 > R_2 > \dots > R_K$, where R_1 is awarded to the winning horse only. Denote the sum of all the rewards by R , and suppose the spread on horse i is (a_i, b_i) , where $\sum (a_i + b_i) / 2 = R$. Then, provided the values of $\{b_i - a_i\}$ are sufficiently small, a collection of win odds for each horse can be constructed with the properties

1. the bookies' margin on these win odds is $m = 0.5 \sum (b_i - a_i) / (R_1 - R_K)$
2. these win odds are derived to be consistent with the actual spreads on offer.

Proof

Fix attention on one particular horse, Alpha. To construct the spread (a, b) for Alpha, a firm may believe that the probabilities that Alpha achieves the different rewards are $\{p_j : j = 1, 2, \dots, K\}$, and so its mean reward is $u = \sum_j R_j p_j$. We have argued that we can expect (a, b) to take the form $(u - w, u + w)$.

For any $j = 1, 2, \dots, K - 1$, the firm could consider offering spreads where the reward is either R_j or R_{j+1} , according as Alpha would have scored at least R_j , or at most R_{j+1} , respectively, on the original Performance Index. On this new scale, Alpha's mean reward is

$$\theta_j = R_j \sum_{k=1}^j p_k + R_{j+1} \sum_{k=j+1}^K p_k \quad (\text{B1})$$

By continuity considerations, there will be some spread (c_j, d_j) so that the firm is indifferent whether a punter would buy or sell at this new spread, or at the original spread (a, b) . Similarly, a punter would be indifferent between buying the original

spread at unit amount β , or buying the j th new spread at unit amount x_j , for each j , provided that

$$\sum_{j=1}^{L-1} x_j(R_{j+1} - d_j) + \sum_{j=L}^{K-1} x_j(R_j - d_j) = \beta(R_L - b) \quad (\text{B2})$$

for each distinct reward L .

Subtracting Equation (B2) when $L = M + 1$ from that when $L = M$ leads to

$$x_M(R_M - R_{M+1}) = \beta(R_M - R_{M+1}) \quad (\text{B3})$$

and so, since the rewards are distinct, $x_M = \beta$ for all M . But then Equation (B2) collapses to

$$\sum_{j=1}^{K-1} d_j = b + \sum_{j=2}^{K-1} R_j \quad (\text{B4})$$

whatever the value of L .

Similarly, by considering the equivalence of a single sell bet to a series of sell bets on these new spreads, the same outcome can arise for the punter so long as

$$\sum_{j=1}^{K-1} c_j = a + \sum_{j=2}^{K-1} R_j \quad (\text{B5})$$

But in a two-horse race with spread (c_j, d_j) for rewards R_j or R_{j+1} , there is an equivalent odds bet with overround $\gamma_j = (d_j - c_j)/(R_j - R_{j+1})$. As all these new spreads are to be seen as equally attractive, set this ratio equal to γ for all j . Equations (B4) and (B5) lead to

$$b - a = \sum_{j=1}^{K-1} (d_j - c_j) = \sum_{j=1}^{K-1} \gamma(R_j - R_{j+1}) = \gamma(R_1 - R_K) \quad (\text{B6})$$

Thus

$$d_j - c_j = \gamma(R_j - R_{j+1}) = (b - a)(R_j - R_{j+1})/(R_1 - R_K) = 2w_j \quad (\text{B7})$$

and we expect $c_j = \theta_j - w_j$, $d_j = \theta_j + w_j$.

For this spread to be sensible, we require that

$$R_{j+1} < \theta_j - w_j \quad \text{and} \quad \theta_j + w_j < R_j,$$

that is, that both

$$w_j < (R_j - R_{j+1}) \sum_{k=1}^j p_k \text{ and } w_j < (R_j - R_{j+1}) \sum_{k=j+1}^K p_k.$$

Using Equation (B7), these are satisfied for all j if and only if

$$b - a < 2(R_1 - R_K) \text{Min}\{p_1, p_K\} \quad (\text{B8})$$

Suppose this condition is satisfied, so that the firm would be able to construct spreads (c_j, d_j) for rewards R_j and R_{j+1} , that are equivalent for all choices of j , and also to the actual spread (a, b) on offer. A punter who buys the spread on (R_2, R_1) is betting that Alpha will win, so a buy bet is the same as a win bet at odds of $R_1 - \theta_1 - w_1 : \theta_1 + w_1 - R_2$, where $2w_1 = (b - a)(R_1 - R_2)/(R_1 - R_K)$.

Repeat this procedure for every horse in the race, assuming that the condition in Equation (B8) holds each time. The condition in Equation (B1) can hold for every horse only if $\sum (a_i + b_i)/2 = R$, so we make this assumption. Let $(t_i - v_i, t_i + v_i)$ be the equivalent spread over (R_2, R_1) for horse i so that $2v_i = (b_i - a_i)(R_1 - R_2)/(R_1 - R_K)$. Then the overround for the full collection of win odds is

$$\sum_{i=1}^n \frac{t_i + v_i - R_2}{R_1 - R_2} - 1 \quad (\text{B9})$$

Now $t_i = q_i R_1 + (1 - q_i) R_2$, where q_i is the probability that horse i wins the race, so that $\sum_i q_i = 1$; hence, expression (B9) reduces to $0.5 \sum (b_i - a_i)/(R_1 - R_K)$, as claimed.

APPENDIX C

Let π be some permutation of the labels $\{1, 2, \dots, n\}$. If π corresponds to the finishing order of the horses, the bettor's fortune after the race will be

$$F^1 = F + \sum_i \frac{F(R_{\pi_i} - b_i)u_i}{b_i} + \sum_i \frac{F(a_i - R_{\pi_i})v_i}{R_1 - a_i}$$

which simplifies to

$$\frac{F^1}{F} = w + \sum_i \frac{R_{\pi_i} u_i}{b_i} + \sum_i \frac{(R_1 - R_{\pi_i}) v_i}{R_1 - a_i} = D_{\pi}.$$

Writing $t = w + \sum_j R_1 v_j / (R_1 - a_j)$, we have

$$D_{\pi} = t + \sum_i u_i R_{\pi_i} / b_i - \sum_i v_i R_{\pi_i} / (R_1 - a_i).$$

If a series of such bets are available, and p_π is the probability of this outcome, the Kelly strategy is to maximize

$$G = \sum_{\pi} p_{\pi} \ln(D_{\pi})$$

under the constraints that $u_i \geq 0$, $v_i \geq 0$ and $\sum u_i + \sum v_i + w = 1$. Write

$$H = G + \lambda(1 - \sum u_i - \sum v_i - w),$$

where λ is a Lagrange multiplier. Then

$$\frac{\partial H}{\partial u_i} = \sum_{\pi} \frac{p_{\pi} R_{\pi_i}}{D_{\pi} b_i} - \lambda \quad (C1)$$

$$\frac{\partial H}{\partial v_i} = \sum_{\pi} \frac{p_{\pi} (R_1 - R_{\pi_i})}{D_{\pi} (R_1 - a_i)} - \lambda \quad (C2)$$

$$\frac{\partial H}{\partial w} = \sum_{\pi} \frac{p_{\pi}}{D_{\pi}} - \lambda \quad (C3)$$

For an optimum, we require $\frac{\partial H}{\partial w} = 0$, and $\frac{\partial H}{\partial \theta} = 0$ when $\theta > 0$, $\frac{\partial H}{\partial \theta} \leq 0$ when $\theta = 0$, for $\theta = u_i, v_i$. Hence $w \frac{\partial H}{\partial w} + \sum_i \left(u_i \frac{\partial H}{\partial u_i} + v_i \frac{\partial H}{\partial v_i} \right) = 0$, from which $\lambda = 1$. After some simplification, we find that, at a possible optimum:

$$\text{if } u_i > 0, \text{ then } \sum_{\pi} p_{\pi} R_{\pi_i} / D_{\pi} = b_i \quad (C4)$$

$$\text{if } v_i > 0, \text{ then } \sum_{\pi} p_{\pi} R_{\pi_i} / D_{\pi} = a_i \quad (C5)$$

$$\text{if } u_i = 0 \text{ and } v_i = 0, \text{ then } a_i \leq \sum_{\pi} p_{\pi} R_{\pi_i} / D_{\pi} \leq b_i \quad (C6)$$

Solving Equations (C4)–(C6) for a general set of rewards normally requires an iterative process based on (C1) and (C2). But consider the special case when the winner scores R ($R > 0$) and all others score zero. Write $p_i = \text{Pr}(\text{Horse } i \text{ wins})$, $\mu_i = R p_i$ and $D_i = t + R u_i / b_i - R v_i / (R - a_i)$. Then Equations (C4) and (C5) simplify to $p_i / D_i = b_i / R$ and $p_i / D_i = a_i / R$. Thus, at an optimum,

$$\text{when } u_i > 0, \text{ we have } u_i = p_i - t b_i / R; \quad (C7)$$

$$\text{when } v_i > 0, \text{ we have } v_i = (R - a_i)(t a_i / R - p_i) / a_i \quad (C8)$$

Write $X = \{i: u_i > 0\}$, $Y = \{i: v_i > 0\}$ and $Z = \{i: u_i = 0, v_i = 0\}$. For all $i \in Z$, $D_i = t$ is constant. Using Equation (C3) at a possible optimum we have

$$1 = \sum p_i / D_i = \left\{ \sum_{i \in X} + \sum_{i \in Y} + \sum_{i \in Z} \right\} p_i / D_i = \sum_{i \in X} b_i / R + \sum_{i \in Y} a_i / R + \sum_{i \in Z} p_i / t$$

so that

$$t \left\{ 1 - \sum_{i \in X} b_i / R - \sum_{i \in Y} a_i / R \right\} = \sum_{i \in Z} p_i. \quad (C9)$$

We argue that if a favorable bet exists, there will always be an optimal bet that does not use every horse, that is, where Z is nonempty. To show this is the case, suppose that some optimal bet does use every horse. The “no arbitrage” condition ensures that neither X nor Y can consist of all the runners, hence both are nonempty. Equation (C9) shows that either $\sum_{i \in X} b_i + \sum_{i \in Y} a_i = R$, and the value of t is indeterminate, or $t = 0$. The latter is impossible, since Y is nonempty and if $i \in Y$, then $v_i = (R - a_i)(ta_i/R - p_i)/a_i$ would always be negative. If the former holds, select $j \in X$, and replace b_j by $b_j - \varepsilon$, where $\varepsilon > 0$ is so small that the “no arbitrage” condition still holds, and favorable bets still exist. At an optimum here, if the new Z , Z_ε , were empty, then Equation (C9) shows $i \in Y$ that $t_\varepsilon = 0$. Hence, as before, Y_ε is empty, X_ε would consist of all the horses and the “no arbitrage” condition is violated. Thus Z_ε is nonempty, and there is an optimum that does not use every horse. By continuity considerations, as $\varepsilon \rightarrow 0$, the optimum that uses $\{X_\varepsilon, Y_\varepsilon\}$ converges to an optimum for our problem, that indeed does not use every horse.

Thus an algorithm for this special case can be given. For any $s \geq 0$, write $X_s = \{i: Rp_i/b_i > s\}$, $Y_s = \{i: Rp_i/a_i < s\}$. Then follow the steps:

- Step 1 If both X_1 and Y_1 are empty, there is no favorable bet. Otherwise,
- Step 2 Select $u > 0$, $v > 0$ so that $X_u \cap Y_v = \emptyset$, and with $X_u \cup Y_v$ a proper subset of $\{1, 2, \dots, n\}$. Calculate $t_{uv} = \left\{ 1 - \sum_{i \in X_u} p_i - \sum_{i \in Y_v} p_i \right\} / \left\{ 1 - \sum_{i \in X_u} b_i / R - \sum_{i \in Y_v} a_i / R \right\}$
- Step 3 Check the conditions
 1. if $i \in X_u$, then $Rp_i/b_i > t_{uv}$.
 2. if $i \in Y_v$, then $Rp_i/a_i < t_{uv}$.
 3. if $i \notin X_u \cup Y_v$, then $Rp_i/a_i \geq t_{uv} \geq Rp_i/b_i$.

If any of these fail, this choice of $\{X_u, Y_v\}$ does not lead to the optimal bet. But if all of them hold, calculate $G_{uv} = \sum_{i \in X_u} p_i \ln(Rp_i/b_i) + \sum_{i \in Y_v} p_i \ln(Rp_i/a_i) + (\ln(t_{uv})) \{1 - \sum_{i \in X_u \cup Y_v} p_i\}$.

Step 4 Repeat Step 2 for all possible choices of $\{X_u, Y_v\}$, and identify the choice that gives the maximum value of G_{uv} . For this choice, let $t = t_{uv}$. Then $u_i = p_i - tb_i/R$, $v_i = (R - a_i)\{ta_i/R - p_i\}/a_i$, determine the buy and sell stakes in an optimal bet. The corresponding optimal growth rate is then

$$G = \sum_{i \in X} p_i \ln(Rp_i/b_i) + \sum_{i \in Y} p_i \ln(Rp_i/a_i) + \left\{ \sum_{i \in Z} p_i \right\} \ln(t).$$

DOWNLOAD FULL VERSION HERE: <https://bettingbooks.org/12>

This page intentionally left blank

Download More Sports Betting Books at: www.bettingbooks.org

CHAPTER 18

Prediction Markets: From Politics to Business (and Back)

Erik Snowberg

Division of Humanities and Social Sciences, California Institute of Technology, Pasadena, CA, USA.

Justin Wolfers

The Wharton School, University of Pennsylvania, Philadelphia, PA, USA.

Eric Zitzewitz

Dartmouth College, Hanover, NH, USA.

1. Overview	386
2. The First Prediction Markets	388
3. Markets in the Lab	391
4. Current Uses of Prediction Markets: Business and Policy	393
5. Future Directions: Decision Markets	397
6. Potential Pitfalls	398
7. Conclusion	400
<i>References</i>	401

Abstract

Prediction markets are the subject of a growing body of scholarly literature, and growing attention from the business community. These recent trends are often discussed without reference to the long history of these markets. Prediction markets started as simple wagers on political contests and have expanded through laboratory and field experiments. We survey this history, detailing the past and current uses of prediction markets. We conclude by examining some potential challenges that will need to be addressed as the uses of prediction markets expand.

1. OVERVIEW

In recent years, prediction markets have been used to accurately predict the outcome of political contests, sales figures of products, and the outcomes of policy decisions. The recent explosion in the utilization of these markets among leading corporations, policymakers, and academics often obscures their long history and humble roots. In this chapter, we trace the known arc of prediction markets from simple wagers on political contests to a promising business and policy tool. Along the way, we explore the accuracy and information aggregating properties of these markets, as well as potential problems.

We begin by surveying the basic concepts and design of prediction markets. As noted above, such markets began as organized gambling markets, but before they were recognizable as prediction markets they adopted many of the qualities of traditional financial markets. Indeed, the line between prediction markets and gambling or financial markets is often fuzzy, as all three types of markets deal with contingent claims on uncertain future events.

Loose distinctions are possible, however. Prediction markets, unlike financial markets, are not explicitly concerned with the diversification or pooling of risk.¹ Moreover, prediction market securities tend to be concerned with at most a small number of uncertain events, such as whether the Senate will change partisan hands in a given election, whereas the value of financial commodities are linked to a potentially infinite series of events, such as profits over an infinite time horizon.

The distinction between prediction and gambling markets is not in terms of structure, since traders in both are playing a negative sum game. Rather, prediction markets are distinct from gambling markets in that the former produce information externalities that can inform business and policy decisions.²

¹For a vision of how prediction markets, if they develop sufficient liquidity, may also prove useful for those wishing to hedge against specific risks, see the discussions in Athanassoulis et al. (1999) and Shiller (2003).

²Another distinction that has been proposed is that the holding of prediction market securities is not inherently enjoyable, as it is in sports betting and other gambling contests. However, since prediction market securities have negative expected returns as well, rational traders must have some outside utility benefit to buying and selling these securities. Another possibility is that prediction markets have a large number of noise traders so that trading is profitable for rational investors. For more on this point see Wolfers and Zitzewitz (2006).

The simplest, and most common, form of prediction market security is one that pays a fixed amount if a certain event, such as the election of a particular politician, occurs, and nothing otherwise. These “winner-take-all” securities reflect prediction markets’ early history as gambling markets featuring simple wagers. If the marginal trader in such a market is rational and risk-neutral, then the price at any given time (divided by the size of the fixed payoff) reflects the market’s perceived probability of that event occurring. Thus, if we observe a security that pays a dollar if a Republican candidate wins trading for \$0.57, we would infer that the market’s assessment of the probability of this event is 57%.

By modifying this simple contract slightly so that the payoff is linked to some continuous variable, such as a candidate’s vote share, we can instead gain the market’s estimate of the expected value of that variable. If such an “index” contract pays the Republican candidate’s vote share and the price of the contract is \$0.46, the market expects the Republican candidate to carry 46% of the popular vote.

A final type of contract is similar to spread betting on sports. Just as a spread bet in basketball asks the bettor to take a position on whether one team will beat the other by a specified number of points, in a prediction market contract it would ask a trader to take a position on whether a candidate will garner a particular share of the vote. If bets pay twice as much as they cost, then the spread that equalizes the amount of money on either side will reflect the median of the uncertain variable—as each side must win half the time. So, for example, if a contract costs \$1 and pays \$2 if the Republican’s vote share exceeds 52% (and there is equal money on each side of the wager) then the median of the market’s distribution of beliefs about the vote share is 52% accruing to the Republican candidate. These three types of contracts are summarized in Table 1.

These simple contracts reveal simple facts about the distribution of the market’s belief about a particular event. By using variants of these contracts, we can recover higher order moments. For example, an index contract on the vote share of a candidate,

TABLE 1 Contract Types: Estimating Uncertain Quantities or Probabilities

Contract	Example	Details	Reveals market expectation of . . .
Winner-take-all	Event y : Bush wins the popular vote.	Contract costs $\$p$. Pays \$1 if and only if event y occurs. Bid according to value of $\$p$.	Probability that event y occurs, $p(y)$
Index	Contract pays \$1 for every percentage point of the popular vote won by Bush.	Contract pays $\$y$.	Mean value of outcome y : $E[y]$
Spread	Contract pays even money if Bush wins more than y^* % of the popular vote.	Contract costs \$1. Pays \$2 if $y > y^*$. Pays \$0 otherwise. Bid according to the value of y^* .	Median value of y .

Source: Wolfers and Zitzewitz, 2004a.

squared (y^2) yields the market's expectation of this variable ($E[y^2]$). When coupled with an index contract on the vote share itself (y) we can infer the variance of the distribution of the market's belief of a candidate's vote share: $E[y^2] - E[y]^2$. Likewise, a series of index contracts designed to elicit the 48, 49, 50, 51, . . . percentile of the distribution, can recover nearly the entire probability distribution of the market's belief. Of course, having liquid markets for such a large variety of contracts might be a challenge, since it is unlikely that there are many people who would be interested in trading (or understand) a contract on the 48th percentile of the market's belief.³

With the basic design of prediction market contracts in hand, we proceed with our stated task of investigating the history of prediction markets.

2. THE FIRST PREDICTION MARKETS

The earliest data we have from prediction markets are those from organized markets for betting on the presidential election between 1868 and 1940. Data from these large and often well-organized markets have been collected and extensively studied by Rhode and Strumpf (2004, 2005) and much of the following section relies on their work. These markets are said to date back to the election of Washington, and in all likelihood wagering on political contests precedes historical records. However, the market in 1868 seems to be the first market we would call a prediction market in that its data was used to inform the public about the likelihood of a particular candidate winning and possibly used by financial asset traders.

Although many cities had prediction markets of this type in operation during the period from the Civil War to World War II, the most prominent were those operating out of New York. A rough estimate has the New York market accounting for half of all the money wagered on the presidential contest nationwide. From the 1880s to the 1910s, these markets were centered on the Curb Exchange, the informally organized predecessor to the AMEX. In closely contested races, such as those in 1916, the *New York Times*, *Sun*, and *World* provided nearly daily price quotations from early October until Election Day (Rhode and Strumpf, 2004). Indeed, 1916 provides a particularly astonishing example of the size and importance of these markets. Two days after the election there was still no sure winner, and large sums continued to be wagered on the smallest piece of information. The *New York Times* reported that between \$500,000 and \$1,000,000 were wagered at the curb that day (November 9, 1916—two days after the election) and that “oil stocks were almost forgotten.”⁴ Moreover, the total amount wagered in these markets in 1916 was \$165 million (in 2002 dollars), which is twice as much as both candidates combined spent on their election campaigns.

Figure 1 shows the betting odds from 1916 taken from Rhode and Strumpf (2004). These odds reflect a great deal of trading and a rapid incorporation of information

³Such a contract would be an index contract that costs \$29, and pays \$50 if the realized value of the variable (y) is greater than the proposed spread (y^*) and \$0 otherwise. The bet on the other side would cost \$31 and pay \$50 if $y < y^*$. See Wolfers and Zitzewitz (2006) for more on the trade-off between interest and contractibility.

⁴Election Doubts Stimulate Betting, the *New York Times*, November 10, 1916, p. 22.

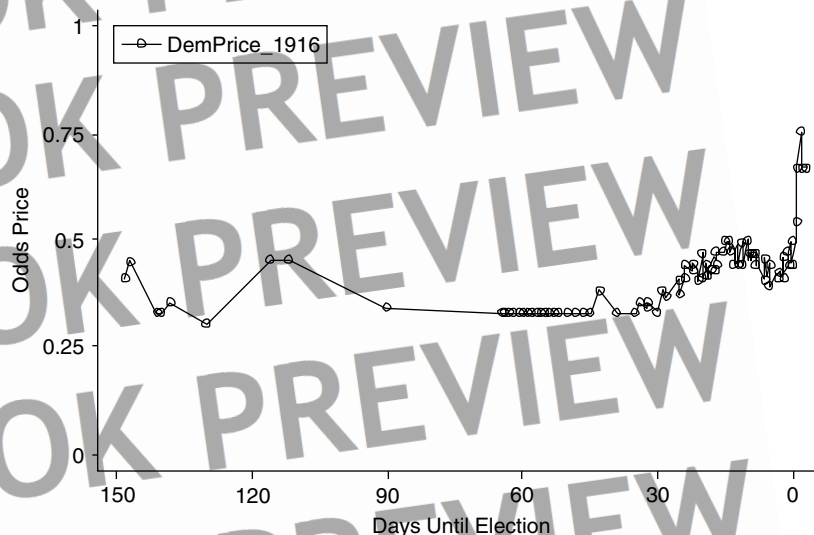


FIGURE 1 Prediction markets move quickly with new information. *Source:* Rhode and Strumpf, 2004.

as the election drew near and then afterwards as the votes were being counted. This rapid incorporation of information is a trait of prediction markets that has been noted elsewhere (see, e.g., Snowberg et al., 2005).

The data from these prediction markets allowed for accurate predictions about the outcome of elections long before the advent of scientific polling. While such data was clearly of use to the voting public at the time, these assessments are still useful to us today. Snowberg et al. (2006) use each presidential candidate's probability of victory the night before the election to describe the amount of *partisan shock* generated by a candidate's election. If, for example, the night before the election the Republican candidate had a 90% chance of winning, then his or her election would cause only a 10% partisan shock. This is clearly different than if the Republican candidate only had a 40% chance of winning—in which case the level of shock would be much higher: 60%. If a Democratic candidate wins, the shock would be in the opposite direction. Thus, if the Republican candidate had a 30% chance of winning and the Democratic won, then the shock would be -30% .

Snowberg et al. (2006) look at equity returns over the same period that the election was resolved (the entire period over which a given year's partisan shock accrued) and found that it was robustly correlated with partisan shock during the period 1880–2004. These results are different from an earlier study by Santa-Clara and Valkanov (2003) who found no correlation between the party that won and changes in equity prices on election night. The difference comes from Santa-Clara and Valkanov's assumption that the magnitude of the partisan shock was always the same, regardless of the market's expected pre-election probability of victory of each candidate. Table 2 shows the results of these regressions, and Figure 2 shows these results graphically.

TABLE 2 The Effect of a Republican on Value-Weighted Equity Returns

Sample	Dependent variable: stock returns from election-eve close to post-election close			
	1928–1996	1928–1996	1880–2004	1880–2004
I(GOP President)	0.0129			
(As in Santa-Clara and Valkanov)	(0.0089)			
Prob(GOP President)		0.0297**	0.0255***	0.0248***
		(0.118)	(0.0082)	(0.0084)
Prob(Incumbent party elected)				−0.0046
				(0.0084)
Constant	−0.0038	−0.0027	−0.0015	0.0014
	(0.0044)	(0.0040)	(0.0028)	(0.0028)

NOTE: *** and ** denote statistically significant at 1%, 5%, and 10%. (White standard errors in parentheses.) Source: Snowberg, Wolfers, and Zitzewitz, 2007.

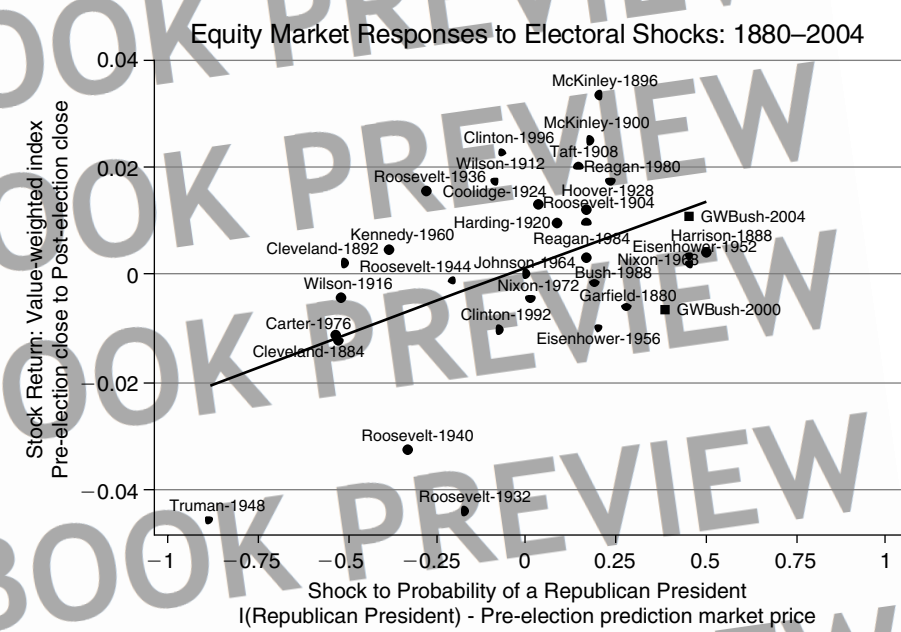


FIGURE 2 Equity markets have historically preferred Republican presidents. Source: Snowberg, Wolfers, and Zitzewitz, 2007.

Knowing the market’s assessment of pre-election probabilities of victory was key for estimating an accurate partisan effect on equity prices. It is likely that data from these first prediction markets will yield more valuable insights in the future.

3. MARKETS IN THE LAB

The first prediction markets faded as scientific polling became available and gambling on sports was legalized in New York State (Rhode and Strumpf, 2004). As these markets slipped into history, studies run by experimental economists laid the groundwork for modern prediction markets. The basic theory of prediction markets—that markets efficiently aggregate dispersed information—comes from the rational expectations (RE) model of Lucas (1972). Early laboratory tests of this theory laid the groundwork for prediction markets by demonstrating the feasibility of using markets to aggregate information.

Plott and Sunder (1982) tested the RE model in a laboratory setting. In their experimental market, participants traded a single-period security. The value of the security was determined by the realization of a single event. In each experimental design, traders differed in the marginal value of the security and the information they possessed about the state underlying the security's value. Plott and Sunder found that market prices could indeed convey information to uninformed traders. Further, they found that the RE model fit many other features of the data—such as the final allocation of securities among traders and the relative profits of informed and uninformed traders—better than other models.

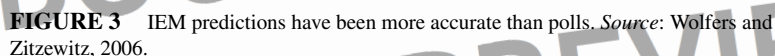
This study was extended by Plott and Sunder (1988) by allowing for heterogeneity in the information possessed by the informed traders. Once again, they found support for the RE model. Specifically, they found that the market could aggregate the different types of information possessed by traders into an accurate representation of the underlying state.

Beyond the overall strong results that markets could both aggregate and convey information to traders, these studies contain warnings for would-be prediction market designers. In each, there was an experimental setup in which RE theories predicted poorly both in an absolute sense and when compared with theories that stipulated that market prices would contain no information other than ex ante supply and demand. These experimental designs tended to be the most complex markets tested—in one case the information revealed to subjects was not easily interpreted, and in another there was wide divergence in the marginal value of the underlying securities. This suggests that more complex prediction markets may not work as well as the simpler ones that are generally the subject of study, a point echoed by Wolfers and Zitzewitz (2006). We will return to the subject of potential pitfalls of prediction market design in our penultimate section.

The Iowa Electronic Market (IEM), arguably the most famous prediction market, was launched in 1988 to study market dynamics in a quasi-experimental setting.⁵ Started as the Iowa Presidential Stock Market, the first set of contracts traded were index contracts with payoffs tied to the major presidential candidates in 1988 (Forsythe et al., 1992). The experiment was continued with contracts linked to vote shares in gubernatorial and congressional elections, as well as elections in countries other than the U.S. (Berg et al., 2001). In addition, winner-take-all contracts tied to the outcome of

⁵Participation in the IEM is open to anyone who opens an account—but there are still provisions that one might expect in an experimental setting, such as a cap on the total amount of assets any one trader is allowed to have.

Once again, there is intrinsic interest in predicting election outcomes, but the information from these markets can be used to answer more complex economic questions. For example, Herron et al. (1999) and Knight (2005) analyze the correlation of industry stock indices and individual stocks with movements in the 1992 and 2000 IEM U.S. Presidential election markets. This allows them to ascertain the relative impacts of the two candidates on various stocks and industries in the U.S. economy. Slemrod and Greimel (1999) use IEM data to examine the effect on municipal bond prices of changes in the probability of the Republicans nominating Steve Forbes, whose flat tax would have eliminated the tax exemption for municipal bond interest. Finally, Roberts (1990) analyzes changes in the betting odds posted by Ladbrokes, a British bookmaker, on Ronald Reagan's re-election and the returns to holding stocks in defense firms, inferring that Reagan led to more robust defense spending.



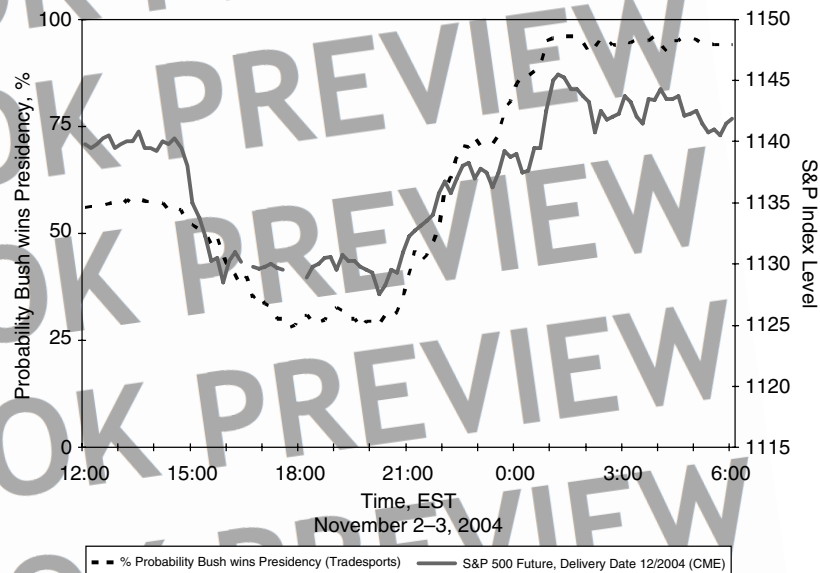


FIGURE 4 The S&P 500 is higher under a Bush vs. Kerry presidency. *Source:* Snowberg, Wolfers, and Zitzewitz, 2007.

Unfortunately, the methodology employed in the above studies cannot measure the impact of candidates on economic aggregates, such as the S&P 500, since changing economic aggregates reflect changes in economic fundamentals that have a direct effect on the probability of re-election of the incumbent party.⁶ Snowberg et al. (2007) sidestep this problem by examining the correlation between changes in the re-election probabilities inferred from prediction market prices and aggregates for the 2004 election, as illustrated in Figure 4. They find that a Bush Presidency led equity prices to be 1.5–2% higher versus a Kerry Presidency.

4. CURRENT USES OF PREDICTION MARKETS: BUSINESS AND POLICY

Having traced the history of prediction markets, we can focus on their current uses. Two characteristics typify the current state of prediction market usage. The first is their

⁶The findings in these studies may also be biased by unobserved factors affecting both the pricing of these portfolios and re-election prospects. For instance, suppose that an election features a pro- and anti-war candidate, and the pro-war candidate is viewed as being more capable of executing a war, should the need arise. If we observe prices of shares in defense contractors increasing in value when the pro-war candidate's electoral prospects increase, one might be tempted to conclude that the defense contractor's stocks are worth more *because* there is a higher chance of the pro-war candidate will be elected. However there may be that a third factor—such as threatening actions from a terrorist group or another nation—that have led both numbers to appreciate: the defense contractor's from their increased sales in an increasingly likely war, and the pro-war candidate's from his country's increased need of his leadership in wartime.

increasing application to understanding critical questions for the conduct of business and the making of policy. The second is the large scale of these markets enabled by the internet.

There are a growing number of web-based prediction markets, often run by companies that provide a range of trading and gambling services. Some prominent examples include Tradesports.com and Betfair.com, and pseudomarkets (in which participants trade virtual currency) such as Newsfutures.com and Ideosphere.com. The use of real money versus a virtual currency that can be redeemed for prizes seems to be of secondary concern as reported in Servan-Schreiber et al. (2004). The most popular emerging prediction markets are summarized in Table 3, which is reprinted from Wolfers and Zitzewitz (2004a).

The use of prediction markets in a business context began with Chen and Plott (2002). In this study, researchers constructed small prediction markets with 20–60

TABLE 3 Popular Prediction Markets

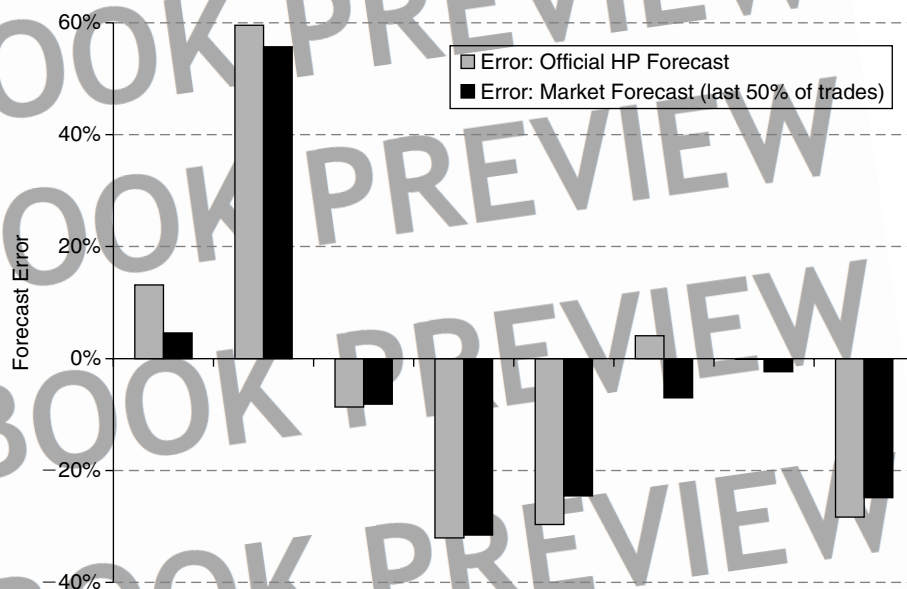
Market	Focus	Typical turnover on an event (\$US)
Iowa Electronic Markets http://www.biz.iowa.edu/iem run by the University of Iowa	Small-scale election markets. Similar markets are run by: UBC (Canada), http://www.esm.buc.ca , and TUW (Austria) http://ebweb.tuwien.ac.at/apsm/ .	Tens of thousands of dollars (traders limited to \$500 positions).
TradeSports http://www.tradesports.com run by for profit company	Trade in a rich set of political futures, financial contracts, current events, sports and entertainment.	Hundreds of thousands of dollars.
Economic Derivatives http://www.economicderivatives.com run by Goldman Sachs and Deutsche Bank	Large-scale financial market trading in the likely outcome of future economic data releases.	Hundreds of millions.
Newsfutures http://www.newsfutures.com run by for profit company	Political, finance, current events, and sports markets. Also technology and pharmaceutical futures for specific clients.	Virtual currency redeemable for monthly prizes (such as a TV).
Foresight Exchange http://www.ideosphere.com run by non-profit research group	Political, finance, current events, science, and technology events suggested by clients.	Virtual currency.
Hollywood Stock Exchange http://www.hsx.com run by Cantor Fitzgerald	Success of movies, movie stars, awards, including a related set of complex derivatives and futures. Data used for market research.	Virtual currency.

Source: Wolfers and Zitzewitz, 2004a.

employees from Hewlett-Packard (HP). The contracts in these markets were designed to capture information about uncertain events that had a direct impact on HP profits, such as the future sales of a given product. As can be seen from Figure 5, these markets outperformed the existing methodologies HP used to predict the same uncertain quantities.

Since this initial study, prediction markets have been used in a wide array of corporate and policy related contexts. Ortner (1998) reports that an internal market correctly predicted that a firm under study would definitely fail to deliver a software project on time, even when traditional planning tools suggested that the deadline could be met. The Hollywood Stock Exchange has generated forecasts of box office success and of Oscar winners, which have been more accurate than expert opinions (Pennock et al., 2001). Both real and play-money markets have generated more accurate forecasts of the likely winners of NFL football games than all but a handful among 2,000 self-professed experts (Servan-Schreiber et al., 2004).

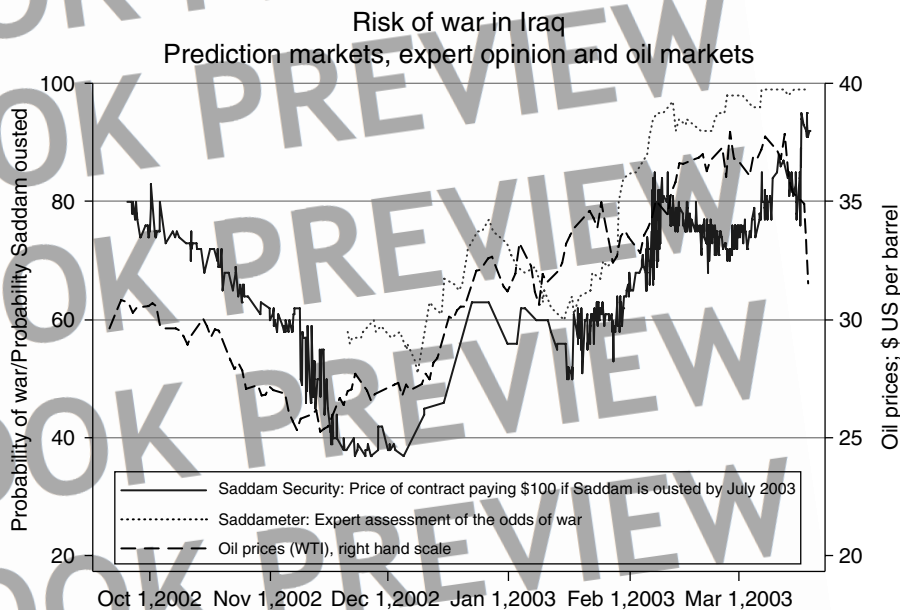
In the public sector, the Pentagon attempted to use markets designed to predict geopolitical risks, although negative publicity stopped the project (Hanson, 2006). An intriguing attempt to apply prediction markets to forecasting influenza outbreaks is detailed in Nelson et al. (2006). In a similar context of policy prediction, Wolfers and Zitzewitz (2005b) report on an ex-ante analysis of the co-movement of oil and equity prices with a contract tracking the probability of a U.S. attack on Iraq in 2002–2003



8 Different Forecasting Experiments: Collected from Chen and Plott (2002)

FIGURE 5 Chen and Plott's experimental market had better forecasts than HP's traditional.

Source: Chen and Plott, 2002.



Sources: Trade-by-trade Saddam Security data provided by Tradesports.com; Saddameter from Will Saletan's daily column in Slate.com

FIGURE 6 The Saddam security closely tracked other measures of Saddam's demise.

Source: Wolfers and Zitzewitz, 2005b.

(Figure 6). The results suggest that a substantial war premium was built into oil prices (and a discount built into equities).

Figure 6 also shows the high correlation between prediction market prices and other indicators of the probability of the same event. The figure contains the "Saddam Security," which was a contract offered on TradeSports paying \$100 if Saddam Hussein were ousted from power by the end of June 2003. The figure further shows that the price of this contract moved in lockstep with two other measures of the end of Hussein's tenure: expert opinion as shown by an expert journalist's estimate of the probability of the United States going to war with Iraq; and oil prices, an obvious barometer of political strife in the Middle East. The prices of all these securities move in lockstep.

This point is reinforced by Figure 7, which shows evidence collected by Gürkaynak and Wolfers (2005) on the relative performance of a prediction market (the Economic Derivatives market established by Goldman Sachs and Deutsche Bank), and a survey of economists, in predicting economic outcomes. They show that the market-based forecast encompasses the information in the survey-based forecasts. Moreover, the behavioral anomalies that have been noted in survey-based forecasts are not evident in the market-based forecasts.

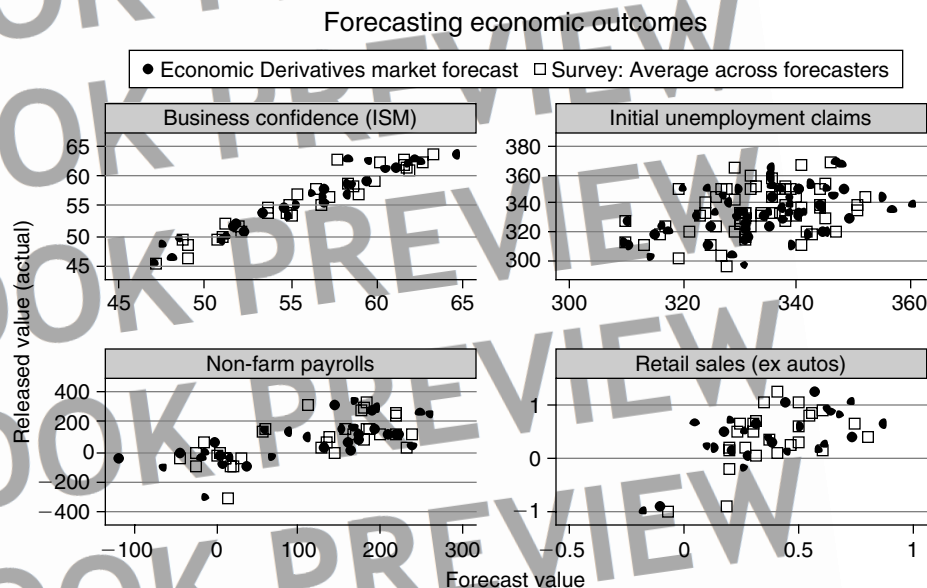


FIGURE 7 Prediction market prices reflect expert opinion and some additional information.
Source: Gürkaynak and Wolfers, 2005.

5. FUTURE DIRECTIONS: DECISION MARKETS

The contracts we have described thus far have depended on only one outcome. The same principles can be applied to contracts tied to the outcomes of more than one event. These contingent contracts potentially provide insight into the correlation between events. For instance, Wolfers and Zitzewitz (2004b) ran experimental markets on the online betting exchange Tradesports.com in the run-up to the 2004 presidential election. In one example, they ran markets linked to whether George W. Bush would be re-elected, whether Osama bin Laden would be captured prior to the election, and whether *both* events would occur. At a given point in time, these markets suggested a 91% chance of Bush being re-elected *if* Osama had been found, but a 67% unconditional probability. Berg and Reitz (2003) report on contracts whose payoff was linked to 1996 Democratic vote shares conditional on different potential Republican nominees; on the basis of these prices they argue that alternative nominees, such as Colin Powell, would have outperformed Bob Dole.

The potential to apply these markets to determine the consequences of a range of contingencies has led Hanson (1999) to term these “Decision Markets.” Indeed, Hanson (2000) has suggested that such markets could be used to remove technocratic policy implementation issues from the bureaucracy, a suggestion endorsed in Hahn and

Tetlock (2006). Moreover, while the previous example involves only one contingency, Hanson (2003) suggests that market scoring rules can allow traders to simultaneously predict many combinations of outcomes. The basic intuition of his proposal is that rather than betting on each contingency, traders bet that the sum of their errors over all predictions will be lower.

6. POTENTIAL PITFALLS

While we remain strongly optimistic about the current and future potential of prediction markets, we would like to draw attention to what we regard as issues that could potentially impede their development path. For example, in the discussion of contracts tied to more than one event immediately above, it is worth cautioning the reader that the standard econometric problem of separating correlation and causation remains. Taking the example of the 1996 Republican nomination one step further, it might be that, holding everything else the same, nominating Colin Powell would have garnered the Republicans a larger vote share than nominating Bob Dole. However, if everything else were the same, Colin Powell would not have been nominated. In order to win the nomination, Powell would have had to create a campaign organization and win primaries, which, in turn, would have revealed information about his political abilities. In a state of the world where Powell's abilities are good enough to win the Republican nomination, they might also be good enough to garner a large vote share against Bill Clinton, the Democratic nominee.

Although we have seen that prediction markets work quite well in practice, there are still theoretical gaps in our understanding of the dynamics of these markets. We know from the observation of markets in the real world and in the laboratory that they are prone to bubbles, information traps, false equilibria, and excess volatility (Plott, 2000; Camerer et al., 1999). Without a theory of the microstructure of these markets, it is difficult to predict when such undesirable properties will appear. Rather, we will only be able to come up with a list of one-off situations in which they have appeared. This makes the task of designing prediction markets more difficult and risky.

Theorists are slowly beginning to understand prediction markets. Manski (2004) notes an example where prediction market prices fail to aggregate information appropriately. In his model all traders are willing to risk exactly \$100. Thus if a contract paying \$1 if an event occurs, is selling for \$0.667, then buyers each purchase 150 contracts, while sellers can afford to sell 300 contracts (at a price of \$0.333). This can only be in equilibrium if there are twice as many buyers as sellers, implying that the market price must fall at the 33rd percentile of the belief distribution, rather than the mean. The same logic suggests that a prediction market price of π implies that $(1 - \pi)\%$ of the population believes that the event has less than a $\pi\%$ chance of occurring. Clearly, the driving force in this example is the assumption that all traders are willing to risk a fixed amount.

Wolfers and Zitzewitz (2005a) provide sufficient conditions under which prediction market prices coincide with average beliefs among traders (and hence aggregate all information). They consider individuals with log utility and initial wealth, y , who must

choose how many prediction market securities, x_j , to purchase at a price, π , given that they believe that the probability of winning their bet is q_j :

$$\text{Max } EU_j = q_j \text{Log}[y + x_j(1 - \pi)] + (1 - q_j)\text{Log}[y - x_j\pi]$$

yielding

$$x_j^* = y \frac{q_j - \pi}{\pi(1 - \pi)}.$$

The prediction market is in equilibrium when supply equals demand:

$$\int_{-\infty}^{\pi} y \frac{q - \pi}{\pi(1 - \pi)} f(q) dq = \int_{\pi}^{\infty} y \frac{\pi - q}{\pi(1 - \pi)} f(q) dq.$$

If beliefs (q) and wealth (y) are independent, then this implies:

$$\pi = \int_{-\infty}^{\infty} q f(q) dq = \bar{q}.$$

Thus under log utility, the prediction market price equals the mean belief among traders. If wealth is correlated with beliefs, then the prediction market price is equal to a wealth-weighted average belief. This finding is general in the sense that no assumptions are required about the distribution of beliefs, but it is also quite specific, in that it holds only under log utility. Experimenting with a range of alternative utility functions and distributions of beliefs typically yields prediction market prices that diverge from the mean of beliefs by only a small amount.

A particularly well-documented feature of issue in small markets should be familiar to those who study gambling. This so-called favorite-longshot bias, describes a tendency to overprice low probability events. Figure 8, from Snowberg and Wolfers (this volume), shows that the favorite-longshot bias is robust in gambling on horse races. A similar tendency has been documented in a range of other market contexts, suggesting that some caution is in order in interpreting the prices of low probability events.

A final potential issue that has garnered concern, especially in markets designed to inform policy, is the possibility of market manipulation. Camerer (1998) addresses this question in pure gambling markets by making and canceling large bets at a racetrack. He finds that his actions had little effect on the final odds after a brief transitory period.

Rhode and Strumpf (2005) study the possibility for manipulation in prediction markets and find little cause for concern. They document episodes in early prediction markets where agents of parties and candidates placed large wagers in order to try to build political momentum. These wagers were often reported in the press due to the public and often nonanonymous nature of early prediction markets. They find that these wagers had a transient effect on prices, but that they quickly returned to their pre-attack level. Rhode and Strumpf also study more recent attempts at market manipulation. For

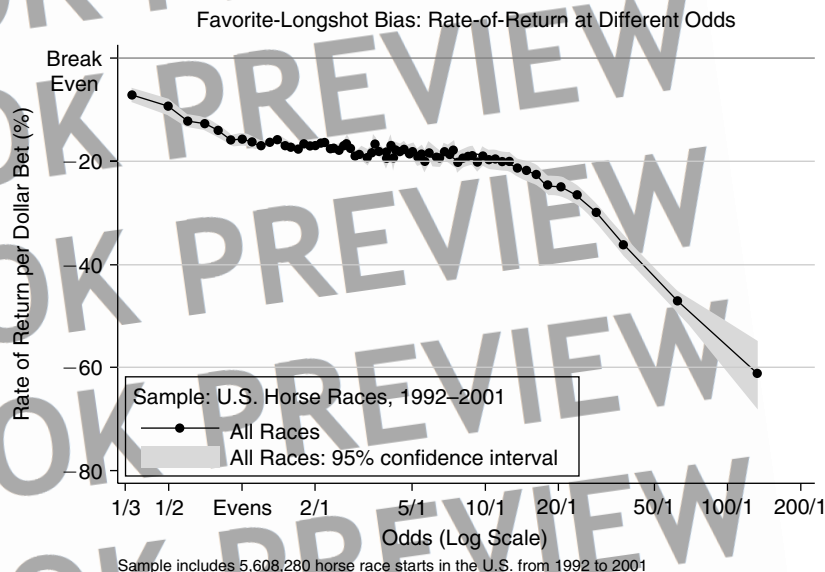


FIGURE 8 The favorite-longshot bias is caused by bettors misestimating small probabilities.
Source: Snowberg and Wolfers, 2008.

example, in 2000 they placed large random bets on the IEM accounting for 2% of the total volume and once again found small transient effects.

Finally, Rhode and Strumpf document the speculative attacks on TradeSports during the 2004 elections. These attacks are illustrated in Figure 9. These attacks intended to drive the price of the Bush security down, possibly in an effort to create panic selling and a lower perceived probability of Bush winning. As should be clear from the figure, these attacks had no long-term effects on the price of the security.

7. CONCLUSION

Prediction markets have proven extremely capable at eliciting simple features of the market's assessment of uncertain events. These capabilities were initially noted outside of academic circles. Laboratory and other experimental markets have brought new understanding of prediction markets, and have allowed for a robust academic agenda to improve their scope and precision. While innovations such as decision markets may be prone to some of the problems outlined in this essay, current research has already shown that some of our concerns, such as manipulation and a lack of a strong theoretical foundation, are not likely to be hindrances to future development of prediction markets. This gives us hope that whatever issues or limitations prediction markets may encounter in the future will be quickly solved by the large and expanding group of scholars working with these markets.

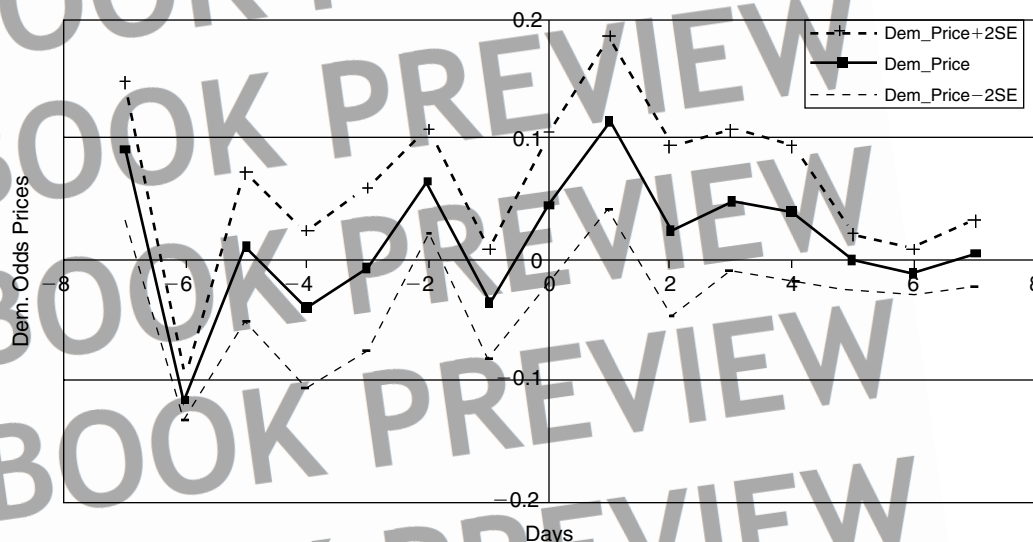


FIGURE 9 Price effects of “Democratic manipulation” in presidential races. *Source:* Rhode and Strumpf, 2005.

References

- Athanasoulis, S., R. Shiller, and E. van Wincoop. 1999. Macro Markets and Financial Security, *Economic Policy Review* 5(1), 21–39.
- Berg, J., R. Forsythe, F. Nelson, and T. Rietz. 2001. Results from a Dozen Years of Election Futures Markets Research, in *Handbook of Experimental Economic Results*. C. Plott and V. Smith (eds.), Elsevier, Amsterdam.
- Berg, J., F. Nelson, and T. Rietz. 2006. Accuracy and Forecast Standard Error in Prediction Markets, mimeo, University of Iowa.
- Berg, J., and T. Rietz. 2003. Prediction Markets as Decision Support Systems, in *Information System Frontiers* 5(1), 79–93.
- Camerer, C. 1998. Can Asset Markets be Manipulated? A Field Experiment with Racetrack Betting, *Journal of Political Economy* 106(3), 457–482.
- Camerer, C. F., M. Noeth, C. R. Plott, and M. Webber. 1999. Information Aggregation in Experimental Asset Markets: Traps and Mismatched Beliefs. CalTech Social Science Working Paper No. 1060.
- Chen, K.-Y., and C. Plott. 2002. Information Aggregation Mechanisms: Concept, Design and Implementation for a Sales Forecasting Problem. CalTech Social Science Working Paper No. 1131.
- Forsythe, R., F. Nelson, G. R. Neumann, and J. Wright. 1992. Anatomy of an Experimental Political Stock Market, *The American Economic Review* 82(5), 1142–1161.
- Gürkaynak, R., and J. Wolfers. 2005. Macroeconomic Derivatives: An Initial Analysis of Market-Based Macro Forecasts, Uncertainty, and Risk, in C. Pissarides and J. Frankel (eds.), *NBER International Seminar on Macroeconomics*, NBER, Cambridge, MA.
- Hahn, R., and P. Tetlock. 2006. Using Information Markets to Improve Decision-Making, *Harvard Journal of Law and Public Policy*, forthcoming.
- Hanson, R. 1999. Decision Markets, *IEEE Intelligent Systems* 14(3), 16–19.
- Hanson, R. 2000. Shall We Vote on Values, But Bet on Beliefs? Mimeo, George Mason University.
- Hanson, R. 2003. Combinatorial Information Market Design, *Information Systems Frontiers* 5(1), 105–119.
- Hanson, R. 2006. Designing Real Terrorism Futures, *Public Choice*, forthcoming.

- Herron, M., D. Cram, J. Lavin, and J. Silver. 1999. Measurement of Political Effects in the United States Economy: A Study of the 1992 Presidential Election. *Economics and Politics* 11(1), 51–81.
- Knight, B. 2005. Are Policy Platforms Capitalized into Equity Prices? Evidence from the Bush/Gore 2000 Presidential Election. *Journal of Public Economics*.
- Lucas, R. E., Jr. 1972. Expectations and the Neutrality of Money. *Journal of Economic Theory* 4, 103–124.
- Manski, C. 2004. Interpreting the Predictions of Prediction Markets. NBER Working Paper No. 10359, NBER, Cambridge, MA.
- Nelson, F., G. Neumann, and P. Polgreen. 2006. Operating With Doctors: Results from the 2004 and 2005 Influenza Markets. Mimeo, University of Iowa.
- Ortner, G. 1998. Forecasting Markets—An Industrial Application. Mimeo, Technical University of Vienna.
- Pennock, D., S. Lawrence, F. A. Nielsen, and C. L. Giles. 2001. Extracting Collective Probabilistic Forecasts from Web Games. *Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 174–183.
- Rhode, P. W., and K. S. Strumpf. 2004. Historical Presidential Betting Markets. *Journal of Economic Perspectives* 18(2), 127–142.
- Rhode, P. W., and K. S. Strumpf. 2005. Manipulation in Political Stock Markets. Mimeo, University of North Carolina.
- Roberts, B. 1990. Political Institutions, Policy Expectations, and the 1980 Election: A Financial Market Perspective. *American Journal of Political Science* 34(2), 289–310.
- Santa-Clara, P., and R. Valkanov. 2003. The Presidential Puzzle: Political Cycles and the Stock Market. *The Journal of Finance* 58(5), 1841–1872.
- Servan-Schreiber, E., J. Wolfers, D. Pennock, and B. Galebach. 2004. Prediction Markets: Does Money Matter? *Electronic Markets* 14(3), 243–251.
- Shiller, R. 2003. *The New Financial Order: Risk in the Twenty-First Century*. Princeton University Press, Princeton, NJ.
- Slemrod, J., and T. Greimel. 1999. Did Steve Forbes Scare the Municipal Bond Market? *Journal of Public Economics* 74, 81–96.
- Snowberg, E., and J. Wolfers. 2008. Examining Explanations of a Market Anomaly: Preferences or Perceptions? D. B. Hausch and W. T. Ziemba (eds.), *Handbook of Sports and Lottery Markets*. North-Holland, Amsterdam, pp. 103–136.
- Snowberg, E., J. Wolfers, and E. Zitzewitz. 2007. Partisan Impacts on the Stockmarket: Evidence from Prediction Markets and Close Elections. *The Quarterly Journal of Economics* 122(2), 807–829.
- Wolfers, J., and E. Zitzewitz. 2004a. Prediction Markets. *Journal of Economic Perspectives* 18(2), 107–126.
- Wolfers, J., and E. Zitzewitz. 2004b. Experimental Political Betting Markets and the 2004 Election. *The Economists' Voice* 1(2), article #1.
- Wolfers, J., and E. Zitzewitz. 2005a. Interpreting Prediction Market Prices as Probabilities. NBER Working Paper No. 10359, NBER, Cambridge, MA.
- Wolfers, J., and E. Zitzewitz. 2005b. Using Markets to Inform Policy: The Case of the Iraq War. Mimeo, University of Pennsylvania.
- Wolfers, J., and E. Zitzewitz. 2006. Five Open Questions About Prediction Markets, in Robert Hahn and Paul Tetlock (ed.), *Information Markets: A New Way of Making Decisions in the Public and Private Sectors*. AEI-Brookings Joint Center, Washington, DC.
- Wolfers, J., and E. Zitzewitz. 2006. Prediction Markets in Theory and Practice in *The New Palgrave Dictionary of Economics*, 2nd ed. Larry Blume and Steve Durlauf (eds.).

CHAPTER 19

Betting Exchanges: A Technological Revolution in Sports Betting

Michael A. Smith

Leeds Business School, Leeds Metropolitan University, Leeds, UK.

Leighton Vaughan Williams

Nottingham Business School, Nottingham Trent University, Nottingham, UK.

1. Introduction	404
2. The Operation of Betting Exchanges	404
3. Empirical Models and Evidence Concerning Weak-Form Information Efficiency in Betting Exchanges	409
4. New Evidence on the Degree of Bias in Betting Exchange Odds	412
5. Conclusions	416
<i>References</i>	417

Abstract

This chapter offers an introduction to person-to-person betting exchanges, a relatively new market format in sports betting. There is currently a debate concerning the integrity of exchanges, which has elicited a variety of policy responses in different countries, ranging from regulation to outright prohibition. To enable the debate to be conducted on a more informed basis, there is a role for economists in the measurement of the economic costs and benefits of this new type of betting.

The chapter is primarily concerned with the information efficiency of exchanges. Their operation and functionality are discussed, along with some of the initial impacts they have had on betting markets, including the hostile response from some traditional bookmakers. Previous empirical studies concerning information efficiency in exchanges are outlined, although they are few at this relatively early stage of the development of the exchanges. New evidence about the degree of odds bias in exchange horse race markets is then presented. This evidence is found to be consistent with previous studies showing a much smaller degree of favorite-longshot bias in the exchanges than exists in traditional bookmaker markets.

1. INTRODUCTION

We examine a relatively new and rapidly growing market format in sports betting, namely the betting exchange. The discussion is confined to horse race betting markets, as the number of events and turnover is greater than in other sports, enabling more reliable measurement of the impact of information relevant to outcomes.

An outline of the concept and operation of betting exchanges is given in Section 2. Section 3 reviews the previous empirical studies concerned with the information efficiency of betting exchanges, which are scant at this relatively early stage of their development. Section 4 presents some new evidence on the degree of bias exhibited by betting exchange markets, employing a trade weighted regression of objective probabilities on subjective probabilities. Section 5 concludes.

2. THE OPERATION OF BETTING EXCHANGES

It is now widely accepted that the technologies associated with the internet have spawned radical new business models and formats, with potentially far-reaching implications for market structures, behavior, and performance (Kim and Mauborgne, 1999, Amit and Zott, 2001). Aspects of the gambling industry are being transformed by these pervasive technologies, with the internet prompting the innovation of new retail formats and technological platforms to facilitate betting markets (Jones et al., 2006).

Arguably the most notable recent internet innovation in this respect came with the advent of the betting exchange, an interactive web-based platform for placing and laying

bets on sporting events. The commercial expertise required to create this business format was imported from the City (London's financial district).

Betting exchanges exist to match people who want to bet on the outcome of an event at a given price with others who are willing to offer that price. The advantage of this form of betting for the bettor is that, by allowing anyone with access to a betting exchange to offer or lay odds, it serves to reduce margins in the odds compared to the best odds on offer with traditional bookmakers. Exchanges allow clients to act as a backer (accepting odds) or layer (offering odds) at will, and indeed to back and lay the same outcome at different times during the course of the market.

The major betting exchanges present clients with the three best odds and stakes which other members of the exchange are offering or asking for. For example, for Take The Stand to win the Grand National, the best odds on offer might be 14 to 1, to a maximum stake of £80, 13.5 to 1, to a further stake of £100, and 12 to 1, to a further stake of £500. These odds, and the staking levels available, may have been offered by one or more other clients who believe that the true odds are longer than they have offered, or who are trying to hedge existing liabilities.

An alternative option available to potential backers is to enter the odds at which they would be willing to place a bet, together with the stake they are willing to wager at that odds level. This request (say £50 at 15 to 1) will then be shown on the request side of the exchange, and may be accommodated by a layer at any time until the event takes place. Every runner in the race will similarly have prices offered, prices requested, and explicit bet limits.

Table 1 illustrates an extract from a *Betfair* horse race market, as accessed via the internet.

TABLE 1 Partial Reproduction of *Betfair* Odds and Bet Limits for a UK Horse Race Betting Market

	BACK			LAY		
Air Wave	3.4	3.45	3.55	3.6	3.65	3.7
	£368	£87	£714	£84	£338	£930
Ratio	6.6	6.8	7	7.2	7.4	7.6
	£854	£623	£373	£312	£350	£5
Crystal Castle	6	6.2	6.4	6.8	7	
	£303	£300	£409	£195	£275	
Fayr Jag	9.2	9.4	9.6	10	10.5	11
	£295	£444	£693	£586	£284	£193
Acclamation	12	12.5	13	13.5	14	14.5
	£34	£115	£531	£811	£315	£196
Rudi's Pet	16	17.5	18	19	19.5	20.5
	£200	£140	£156	£249	£188	£150

Source: <http://www.Betfair.com>.

The “Back” side of the market indicates the best three odds available for runners, to stakes limits advised in each case. These have been offered by *Betfair* clients acting as layers; odds are expressed including stake, therefore the best odds against Air Wave are 3.55/1, or 2.55/1, to a maximum stake of £714. The “Lay” side of Table 1 shows odds which have been requested by traders, with similar interpretation of odds and limits as above. The margin between the best odds on offer and the best odds sought tends to narrow as more clients offer and lay bets, so that in popular markets the real margin against the bettor (or layer) tends toward the commission levied by the exchange. This commission varies up to 5% on a customer’s net winnings on a market, the exact rate depending on the history and volume of business conducted by the trader.

Clients can monitor price changes, which are frequent, on the internet Website pages of the betting exchange, and execute bets, lay bets, or request a price, instantly and interactively. It is also now routinely possible for members of major exchanges to monitor prices and place wagers from a mobile phone.

Exchanges also operate “in running” markets, characterized by rapid and frequent price changes as races progress. Such markets are facilitated by the ability of clients to employ joint technologies, observing televised races on the internet or by means of conventional television, with simultaneous access to the interactive internet based markets. The remainder of this chapter is concerned with the more conventional pre-event markets, as in running markets are largely unexplored in terms of empirical analysis.

There are a number of key differences between the betting exchanges and bookmaker markets, principal of which is that whereas the bookmaker sets nominal odds, the betting exchange operator acts merely as a broker and offers an information platform, whereby third parties can offer odds or accept odds, in return for which the exchange charges a commission. The exchange assumes no risk of its own and merely brings together bettors and layers. Odds offers are displayed by value and are anonymous and pooled for similar odds values—individual bettors and layers are not specifically matched.

Exchanges avoid settlement risk by transferring funds from the parties to a bet, for the amount of their respective potential liabilities, into a secure holding account at the time the bet is struck. This system ensures that confidence is maintained in the integrity of settlement arrangements.

Unlike the bookmaker markets, exchanges display bet limits for all horses, which are determined not by the exchange, but by the amounts clients who are acting as layers are prepared to stand at the various odds. The ability to lay odds, which also enables bettors to extinguish existing back (bet to win) liabilities, and to request prices, are also features of the exchange markets not available from bookmakers (for further commentary and details of betting exchange functionality see Vaughan Williams, 2002, 2005). A client may also pursue arbitrage activities, for example, backing at a high price early in the market and laying the same horse at a lower price later in the market, should its general price contract, thus locking in value.

A further attraction of the betting exchanges is their low rate commission structure relative to the profit margin implicit in bookmaker odds. For example, *Betfair* charges

a standard commission on net *race* winnings, not individual *horse* winnings—an important factor for bettors who wager on more than one horse in a race. Exchange commissions compare favorably with bookmakers' margins; although the latter make no explicit deductions for either their operation of the market, or for taxation (no tax on winning bets has been paid at the point of sale in the UK since October, 2001), a profit margin is implicit in nominal odds, as the associated sum of probabilities in a race typically exceeds one. The extent of this margin or overround varies from race to race and will typically increase directly with field size, and inversely with liquidity of the market and the extent of race specific competition with other bookmakers. To give an idea of the magnitude of the margin, bookmaker markets analyzed in one recent study (Smith et al., 2006) averaged approximately 20% overround. While betting exchange markets rarely lead to a "Dutch Book" (sum of odds probabilities less than unity), the degree of overround is usually much lower than in book odds.

By 2004–2005, betting exchanges accounted for £4,352,000,000 in sports betting turnover, arising from exponential market growth since their inception at the turn of the Millenium (Mintel, 2005). Approximately 90% of this turnover was attributable to *Betfair*, who by 2005 had 300,000 users and claimed to process over 1,000,000 bets per day at peak betting periods (Jones et al., 2006). An embryonic early major player in the industry, *Flutter*, was taken over by *Betfair* in 2002, consolidating its early position as market leader (*The Economist*, May 8, 2003). *Betfair*'s principal marketing pitch is that it claims the best odds it displays to be 20% better than bookmaker odds on average.¹

The remaining fragmented competition comes from a number of smaller exchanges such as *Betdaq*, *WBX* (*World Bet Exchange*), and *Betsson*. None of *Betfair*'s competitors currently have the market liquidity to seriously challenge the market leader, although the alternative exchange model of *Intrade* and *Tradesports*, for example, based on buying and selling of contracts in a binary framework (outcome = 0 and 100), is an interesting addition to the marketplace (see Oliver, 2007, for a fuller consideration of binary betting). Bookmakers have challenged the legitimacy of the operations of betting exchanges, claiming them to be illegal and representing unfair competition on a number of grounds, which were examined at some length by the Joint Committee on the Draft Gambling Bill (United Kingdom Parliament, 2004).

The arguments put forward to the Committee revolved around issues of integrity, equity, transparency, and market stability (Jones et al., 2006). The integrity of betting exchanges was called into question by bookmakers, who claimed that following a number of high-profile instances of betting irregularities relating to specific races, betting exchanges were subject to an unacceptable degree of insider activity, notably where stable connections can lay horses, then manipulate the result by taking steps to restrain the relevant runner(s) in the race. *Betfair* responded by establishing a code of practice that commits it to full cooperation with the Police, the Fraud Squad, and the racing

¹The odds in the *Betfair* dataset utilized in Smith et al. (2006) actually exceeded the mean of an array of matching bookmaker odds by 18.12%, close enough to make the claim credible. In relation to bookmaker outlier odds, however, the corresponding differential was only 5.39%, approximately equal to the commission rate charged by the exchange, suggesting that bookmaker outlier and exchange odds are mutual benchmarks.

authorities in sharing information that might lead to a subsequent arrest for manipulation of markets through insider activity (*Betfair*, 2007), a criminal offense in the UK.² This code breaches the client confidentiality that is also part of the betting exchange professional code, but only by exception in the interest of the public and where criminal activity is suspected. Betting exchange operators were able to turn the argument against bookmakers, claiming the same duty of care standards were not currently applied by bookmakers themselves.

The equity or fairness issue concerned the tax treatment of bookmakers and betting exchanges, and the related issue of the Racing Levy by which the sport of horse racing is funded in the UK. Since 2001, bookmakers pay a gross profits tax, whereas betting exchanges pay tax on their commission. Bookmakers claimed this to be unequal treatment; layers on the exchanges are effectively acting as bookmakers, they argued, yet pay only commission. The counter-argument from the exchanges was that the tax applied to the gross profits (commission) of the exchange and not to individual clients of the exchange, whether they were backers or layers.

The exchanges claim to contribute more to consumer (bettor) welfare than bookmakers, with greater market transparency. For example, volumes traded for all horses in a race are explicitly stated on betting exchanges, unlike bookmaker markets. They also claim to offer flexibility to clients by giving them the opportunity in part or in whole to negate betting decisions by facilitating offsetting or hedging opportunities. Other benefits suggested by the exchanges in their evidence to the Joint Committee include some of those already discussed above: namely, that they (the exchanges) offer more generous odds, and a level of commission that is more modest than the bookmaker overround, and which is reduced for those bettors executing a high volume of wagers. Moreover, the commission structure is based on winnings per race rather than the winning horse. In short, the betting exchanges implied that bookmakers' objections were really an attempt to prevent further loss of market share to the exchanges (Jones et al., 2006).

On the grounds of greater flexibility, lower transaction costs, and the injection of new competition from the exchanges, the economist might hold an a priori expectation of greater market efficiency in the betting exchanges than in bookmaker markets, perhaps followed by convergence of the two over time (Carlton and Perloff, 2005, pp. 259–267). If this is the case, governments might be well advised to create a favorable regulatory and fiscal climate for the exchanges to develop, in the interests of consumer welfare. However, the influences leading to an expectation of greater efficiency must be seen in the context of the claims (justified or otherwise) of increased insider trading inspired by exchange betting, and the issue of funding the sport of horse racing in the UK needs to be addressed.

²A recent UK court case, in which a number of jockeys (including ex-champion Kieren Fallon) were accused of race fixing to facilitate insider bets on the exchanges, collapsed in December, 2007.

3. EMPIRICAL MODELS AND EVIDENCE CONCERNING WEAK-FORM INFORMATION EFFICIENCY IN BETTING EXCHANGES

This section is limited to a discussion of the extent and sources of the favorite-longshot bias in betting exchanges. There have been few published studies of this type, as betting exchanges have only existed since 2000 as a significant market format.

In a recent empirical analysis using matched data for 700 UK horse races from betting exchanges and from traditional betting media, evidence of the favorite-longshot bias in exchange markets was found, but significantly less so than in the corresponding bookmaker odds (Smith et al., 2006). We further suggested that, based on an application of the methodology of Sobel and Raines (2003), an information-based model explains the favorite-longshot bias more convincingly than earlier explanations based on bettors' risk preferences.

Smith et al. (2006) set out to test the Hurley and McDonough (1995) cost-based explanation for the bias, which states that the higher are transactions costs, the greater will be the favorite-longshot bias. As transactions costs are observed to be lower in exchange markets than those implicit in the overround characteristic of bookmaker markets, the favorite-longshot bias should be less pronounced in exchanges than in bookmaker markets.

Information costs are integral to the Hurley and McDonough explanation of odds bias. High transaction costs imply that better information is required to achieve positive returns; this is a variation on the assertion that where transaction costs exist, it is impossible to discount all information, since a financial incentive must exist for information search (Grossman and Stiglitz, 1990). A further implication in the current market context is therefore that for races where there is less public information one would expect a higher proportion of "casual bettors" and a greater degree of favorite-longshot bias. Vaughan Williams and Paton (1997) and Sobel and Raines (2003) independently found empirical evidence to this effect.

Smith et al. measured the degree of bias employing the Shin's z measure of insider trading, a proxy for bias (Shin, 1991, 1992, 1993). Shin explains the favorite-longshot bias observed in bookmaking markets as the consequence of bookmakers' response to asymmetric information, where some bettors have privileged information concerning the true probability of one or more horses winning a race. The bookmaker response is modeled by Shin as an adverse selection problem, with the empirical consequence that bookmaker odds are depressed below true odds to preserve margins in the face of insider activity.

Shin derives a functional relationship between the sum of odds probabilities, D , associated with a race, such that:

$$D = z(n-1) + \sum_{k=0}^K a_k n^k \text{Var}(\mathbf{p}) + \sum_{k=0}^K b_k n^k [\text{Var}(\mathbf{p})]^2 \quad (1)$$

where $\text{Var}(p)$ is the variance of true probabilities of runners in the race,³ n is the number of runners, and k is the order to which the expression is expanded as a polynomial; for Shin's sample the best fit was achieved by a quadratic function. The coefficient z in Equation (1) is Shin's measure of the proportion of turnover attributable to insider trading, and given the nature of the model, can also be interpreted as a proxy for the degree of odds bias. Shin's estimate of z was 2.46% (calculated as $z \times 100$ to give a percentage) for his sample. Comparable values have subsequently been found for much larger samples of UK races, using bookmaker *SP* data. Vaughan Williams and Paton (1997) and Law and Peel (2002), for example, estimated values of 2.03% (481 races) and 2.7% (971 races), respectively. Coleman (2007) uses an alternative methodology to suggest that at the Melbourne racetrack about 2% of betting is by insiders.

In Smith et al. (2006), we employed Shin's z measure of bias in relation to our betting exchange and bookmaker data. We were interested to establish the relative bias across the two datasets and to test the Hurley and McDonough hypothesis. Our sampled races were categorized by information levels or "classes." Class 1 races were those where publicly available information concerning runners was least, and at the other extreme Class 4 races were those where a great deal of race specific information was available. The results are reproduced in Figure 1.

Figure 1 shows that odds bias, proxied by Shin's z , was lower in the betting exchange odds than in bookmaker markets, a result consistent with the Hurley and McDonough contention that the regular favorite-longshot bias will be greater in those markets where transaction costs are greatest. Further, in both exchange and traditional betting markets, the level of bias was lower the greater the amount of public information that is available to traders, again consistent with the Hurley and McDonough theoretical model. The Shin's z coefficients were confirmed to be significant and independent using the Breusch Pagan test. These results are therefore consistent with an information based model of the favorite-longshot bias, explaining the structural characteristics within and between the two markets in terms of available information and transaction costs.

Aside from identifying bias, a further question of interest is whether the structural differences in odds between betting exchanges and bookmaker or other market odds can be exploited by employing arbitrage or quasi-arbitrage strategies (Paton and Vaughan Williams, 2005; see also Smith et al., 2005; Vaughan Williams, 2000, 2001). Exploitability is important since if a trading profit cannot be made in relation to an imperfection, then markets are operating efficiently to the point where the benefits of further arbitrage are offset by transaction costs, reflecting the Grossman and Stiglitz hypothesis.

Studies conducted prior to the advent of widespread bettor access to the internet show mixed evidence concerning the possibility of profitable arbitrage between markets.

In a study of parlay⁴ markets, Ali (1979) found that returns to the parlay were not significantly different from those in the win markets of the constituent races, suggesting

³Shin's term $\text{Var}(p)$ is a measure of distance of vector p from the vector $1/n$, as opposed to variance in the normal statistical sense.

⁴A parlay is a double win bet that succeeds only if the two nominated runners, competing in separate races, both win.

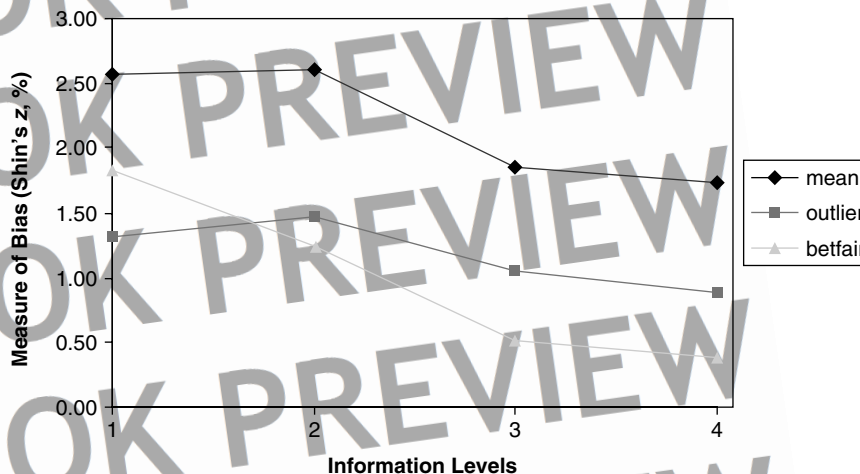


FIGURE 1 Degree of bias for mean, outlier, and *Betfair* prices for matched data in relation to 700 horse races run in the UK during 2002.

NOTE: (i) The y axis shows the coefficient of $n - 1$, or Shin's z , multiplied by 100. The interpretation of this value is that it indicates the percentage of insider trading volume in the market concerned, and also acts as a direct proxy measure of the degree of bias. (ii) Odds were derived from internet arrays of competitive bookmaker prices for 700 races acquired at 10:30 AM on race days, yielding mean odds and outlier odds values for each horse. Matching *Betfair* prices were acquired at the same time, being the best price available to non-trivial stakes. (iii) Class 1 = least public information; Class 4 = most public information. (iv) Mean = mean bookmaker odds from a competitive array of prices per horse per race; outlier = the corresponding bookmaker outlier odds per horse per race; *Betfair* = the corresponding best odds to non-trivial bet limits on the exchange; all matched data. Source: Smith et al. (2006).

weak-form efficiency. Hausch and Ziemba (1985), on the other hand, found persistent weak-form inefficiencies in North American place, show, and exotic betting markets, which are related to win betting markets in complex ways. They were able to demonstrate profits to an arbitrage strategy in these pools, known as "Dr. Z's system," based on estimations of true probabilities from win odds, with wagers subsequently made in the place and show pools.

Pope and Peel (1989) conducted the first study of efficiency in betting markets characterized by an array of odds rather than singular values, in relation to bookmaker odds on English football (called soccer in America). They found that more efficient forecasts could be made by pooling competing odds values, but not to the extent that positive returns could be generated.

Smith et al. (2007) measured the relative accuracy of exchange relative to bookmaker odds for all horses in the 700 race sample outlined above, applying conditional logistic regression to establish which set of odds were the best predictors of race outcomes.

Our principal finding was that, after adjusting for the favorite-longshot bias in both sets of odds, and with the exception of low liquidity markets, exchange odds were significantly better predictors than bookmaker odds. This may be because betting exchanges offer opportunities for the most skilled or informed bettors not available in

bookmaker markets. For example, skilled traders, insiders, and bettors seeking hedging opportunities are all able to lay odds on the exchanges that may as a result reflect the chances of the horses concerned more accurately than those offered by bookmakers. In the main, therefore, this result would suggest that an arbitrage strategy based on differences in parallel odds between the two markets should employ exchange odds as a proxy for true odds, with the exception of low liquidity races.

4. NEW EVIDENCE ON THE DEGREE OF BIAS IN BETTING EXCHANGE ODDS

An alternative to the Shin method of identifying odds bias is the regression of objective probabilities (derived from results) on subjective probabilities (as contained in the odds), or the equivalent analysis of returns distribution by odds value (e.g., Snyder, 1978, Bruce and Johnson, 2000). The virtue of this method is that it allows clear exposition of odds bias through a comparison of results or returns against their expected values across the full range of odds. It further permits comparison of results with empirical studies conducted prior to the advent of the Shin methodology.

Regression of objective on subjective probabilities was adopted for the current study using trade weighted betting exchange data to shed further light on the degree of bias implied by the Shin methodology employed in Smith et al. (2006) in relation to nominal exchange odds.

The specific estimation procedure adopted here is that used by Bruce and Johnson (2000). For their dataset of starting prices (*SP*) for 2,109 races run in 1996, Bruce and Johnson estimated a polynomial function:

$$\ln(pre.b) = \alpha + \beta \ln(o^b) + \chi [\ln(o^b)]^2 + \dots + \delta [\ln(o^b)]^n \quad (2)$$

where the dependent variable *pre.b* is the predicted or expected value of the objective probability of winning for an odds class, and the independent variable o^b is a vector of odds. Weighted least squares regression was used to estimate the values of coefficients α , β , χ , and so on, such that all terms in the polynomial expression were significant. The log of subjective probabilities corresponding to o^b was used as a reference class, against which $\ln(pre.b)$ could be compared to establish the existence of any bias. Bruce and Johnson found that a quadratic function best fitted their *SP* data, while a cubic function fitted a separate Tote dataset.

For our study of exchanges, price-volume data for approximately 6,000 UK horse races run between August 2001 and April 2002 were acquired from the market leading betting exchange, *Betfair*. The number of winners and total runners in each category was not available.

The betting exchange prices are expressed on a decimal odds scale calibrated to as little as 0.01 odds points at the short end of the odds scale, 0.1 intervals in the middle

TABLE 2 Extract of *Betfair* Trading Data

Odds	Total traded	Profit/loss
1.01	261,262.86	-61,371.715
1.02	192,049.4	-54,493.033
1.03	131,869.3	-24,464.527
1.04	125,528.3	-1,706.982
1.05	102,851.83	-12,420.479
1.06	14,553.03	-1,010.142
1.07	7,058.38	371.59
1.08	4,132.4	-157.563
1.09	7,593.95	398.615
1.1	104,705.51	-21,670.373
1.11	72,559.54	701.648
1.12	225,130.48	-368.582
1.13	254,993.53	22,609.789
1.14	311,237.27	37,526.379
1.15	334,900.49	25,356.147
1.16	404,864.35	23,850.829
1.17	439,536.15	51,681.652
1.18	380,297.83	34,160.065
1.19	347,850.95	48,919.749
1.2	446,096.03	17,301.557

Source: *Betfair*, 2003.

odds, and 1 point intervals at long odds. Thus, categories exist not only for 2.0 to 1, but also 2.1, 2.2, and so on.

The betting exchange data contained fields for odds, total traded, and profit/loss before commission (aggregated across all races for which there were markets in the eight month period covered). An extract from the *Betfair* trading information is in Table 2.

The data continues in this way for 7,954 odds ticks, with increases of 0.01, until the odds become large, when the ticks frequently increase with odds intervals of 1 or more, for all values at which bets were struck, up to an odds value of 1,000.

As prices on the exchanges include a unit stake, the initial step in preparing the data for use was to subtract one from the advertised price to obtain an odds value. For example, the odds of 1.01 are actually odds of 0.01.

Objective odds probabilities, the dependent variable, were derived for a range of odds categories from the observed values of exchange “Total traded” and “Profit/loss” as follows.

As the distribution of bet sizes, the incidence and proportion of winners and losers within each odds category were all unknown, an initial assumption that all bets within

each odds category are of equal value was made; the alternative assumption that winners at a given level of odds are distributed evenly among bets of different sizes is equivalent. Either assumption enables the following procedure to be employed to estimate true probabilities.

Let p_{nv} be the probability corresponding to a *Betfair* nominal odds value, o_v , less an adjustment for the exchange operator's rate of commission on net winnings, c , here applied at *Betfair*'s standard rate of 5%, so that $c = 0.05$. This adjustment is made to ensure comparability with starting prices, where bookmakers' margins are implicit in the odds. Then

$$p_{nv} = \frac{1}{o_v(1 - c) + 1}. \quad (3)$$

For example, for the odds of 1.01 above, the actual odds value was recorded as 0.01, and by substitution in Equation (3), $p_{nv} = 0.9906$.

Now let p_{iv} be the probability corresponding to an odds value that would yield zero profits; this is a proxy for the true probability of a horse winning at the nominal odds corresponding to p_{nv} .

Let π_v = profit/loss at the *Betfair* odds value, for example, -£61,371 at odds of 0.01.

Let ϕ_v = total traded at the *Betfair* odds value, for example, £261,262.86 at odds of 0.01. Then

$$p_{iv} = p_{nv} \left(1 + \frac{\pi_v}{\phi_v} \right) \quad (4)$$

Equation (4) derives the probability equivalent to odds that would adjust actual profit/loss to zero. These odds are by implication the true probability of a horse winning in the odds category v .

The next step is to compute a weighted average of p_{iv} for a range of odds categories to enable a regression, as specified in Equation (2) above, of these computed values against average nominal odds probabilities associated with specified odds categories. The precise method of odds classification used is traceable to Weitzman (1965), whereby normalized odds probabilities are categorized according to a measure of the monetary return to a nominal winner at given odds to a unit bet, with categories increasing in one unit increments. This largely solves the problem of classes having an insignificant number of runners, especially in the shortest odds categories, and gives a rational basis for choice of odds boundaries.

Monetary returns, as defined by Weitzman, in the range 1–120 (inclusive of unit stake) were adopted for the purpose of the regression. Let j be the Weitzman category and $n = 120$.

Then the weighted true odds probability values, P_{ij} for Weitzman categories are computed as:

$$P_{ij} = \frac{\sum_{v=1}^n \phi_v p_{iv}}{\sum_{v=1}^n \phi_v} \quad (5)$$

P_{ij} becomes the dependent variable in Equation (2) for the betting exchange data. Average nominal odds corresponding to the Weitzman categories, adjusted for commission, are generated in the same way to give values of the independent variable. Let O_j be the weighted average nominal odds for Weitzman category j . Then:

$$O_j = \frac{\sum_{v=1}^n \phi_v o_v}{\sum_{v=1}^n \phi_v} \quad (6)$$

The results of the ensuing regression are shown in Table 3 and Figure 2. In Figure 2 the betting exchange odds suggested by the reference curve of log subjective probabilities closely matches the expected values of the objective probabilities. There does not appear to be an appreciable bias in the exchange across the range of odds values, with the possible exception of the highest prices, where the odds are more generous than is warranted by the true chances of such horses. This bias at long odds is modeled by the cubic term in the estimation of Equation (2) with respect to the exchange data. Subjective and objective probabilities converge at about 3/1.

These results, based on exchange odds data weighted by trading volume, are largely consistent with the earlier findings in Smith et al. (2006) based on nominal odds for a much smaller sample of races; if anything, the current result indicates a lower degree of bias than that based on Shin methodology in our previous study.

TABLE 3 Coefficient Estimates for Equation (2)

Coefficient	Value	Standard error	<i>t</i>	Significance
α	-0.6838	0.0056	-122.4803	0.0000
β	-0.5360	0.0069	-77.6772	0.0000
χ	-0.1480	0.0072	-20.6242	0.0000
δ	0.0189	0.0028	6.8530	0.0000
R	0.9990			
R^2	0.9980			
Durbin-Watson	1.962			

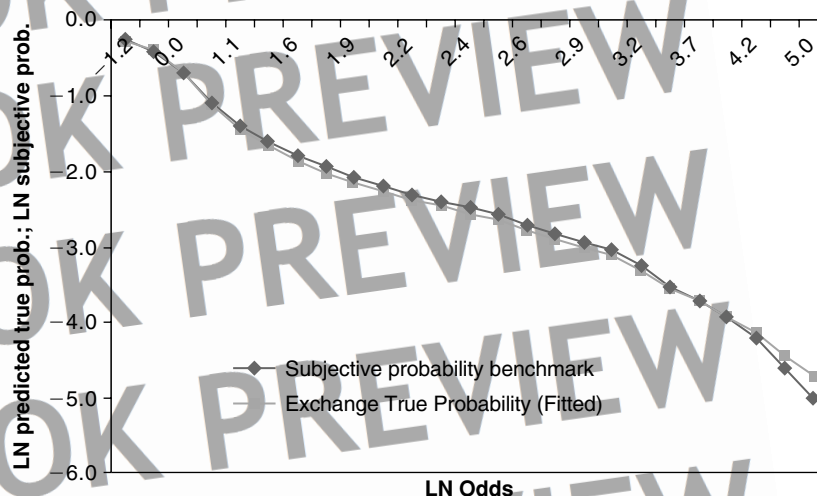


FIGURE 2 Favorite-longshot bias, betting exchange.

5. CONCLUSIONS

Established gambling operators have argued that person-to-person wagering on internet betting exchanges represents unfair competition, on the grounds of differential treatment of the two media for tax purposes, because the exchanges undermine the bookmaker licensing system and because exchange clients are allegedly often exploited by insiders. Prior empirical results and the new evidence presented here suggest that betting exchanges have brought about significant efficiency gains by lowering transaction costs for consumers. In contrast to traditional betting media, it was found that betting exchanges exhibited significantly lower levels of bias and therefore approximated more closely the conditions of weak-form market efficiency. Exchanges contribute to reducing the information barrier; they increase the incentive to process and act on race relevant information by reducing transaction costs and increasing bet flexibility.

The low value of Shin's z found by Smith et al. (2006) in relation to exchange data suggests that only a small proportion of turnover is attributable to "bet to win" insiders. The main controversies surrounding the extent of insider trading in the exchanges, however, relate to "lay" type insider betting. Whether the Shin measure adequately captures such activity is debatable and further research is required in this respect.

Whatever the truth of the inequities and insider abuse claimed by bookmakers to be the result of the exchanges, policy makers contemplating intervention in the industry need to weigh the ensuing economic costs (so far unproven) against the empirically demonstrated welfare gains derived from the exchanges. Betting exchanges are clearly vulnerable to insider activity, especially where laying of "non-triers" or "performance retarded" horses are concerned; the nature of the bet and lay formats would make it incredible for them not to be subject to such influences. Ironically it is the very features

that make the exchanges relatively efficient markets that make them subject to potential abuse from insiders. This is not, however, an argument for the prohibition of exchanges, a policy adopted in some countries; rather it is an argument for adequate policing of exchange client activities, and a strong regulatory framework, in order to preserve the welfare gains accruing to mainstream bettors.

References

- Ali, M. M. 1979. Some Evidence of the Efficiency of a Speculative Market, *Econometrica* 47(2), 387–392.
- Amit, R., and Zott, C. 2001. Value Creation in e-Business, *Strategic Management Journal* 22(6), 493–520.
- Betfair. 2007. Code of Practice for Betting Exchanges. Available at: http://content.betfair.com/aboutus/?region=GBR.IRL&locale=en_GB&brand=betfair. Accessed April 27, 2007.
- Bruce, A. C., and J. E. V. Johnson. 2000. Investigating the Roots of the Favourite-Longshot Bias: An Analysis of Decision Making by Supply- and Demand-Side Agents, *Journal of Behavioral Decision Making* 13, 413–430.
- Carlton, D. W., and J. M. Perloff. 2005. *Modern Industrial Organization*, 4th ed. New York: Pearson Addison Wesley.
- Coleman, L. 2007. Just How Serious is Insider Trading? An Evaluation Using Thoroughbred Wagering Markets, *The Journal of Gambling Business and Economics* 1(1), 31–55.
- The Economist. 2003. Flutter Away. May 8. Available at http://www.economist.com/displaystory.cfm?story_id=S%27%29H0%2ARQ%2F%2B%21%20%20%40%0A&ppv=1.
- Grossman, S., and J. E. Stiglitz. 1980. On the Impossibility of Informationally Efficient Markets, *American Economic Review* 70(3), 393–408.
- Hausch, D. B., and W. T. Ziemba. 1985. Transaction Costs, Extent of Inefficiencies, Entries and Multiple Wagers in a Racetrack Betting Model, *Management Science* 31(4), 381–394.
- Hurley, W., and L. McDonough. 1995. A Note on the Hayek Hypothesis and the Favourite-Longshot Bias in Parimutuel Betting, *American Economic Review* 85(4), 949–955.
- Jones, P., D. Turner, D. Hillier, and D. Comfort. 2006. New Business Models and the Regulatory State: A Retail Case Study of Betting Exchanges, *Innovative Marketing* 3, 112–119.
- Kim, C., and R. Mauborgne. 1999. Creating New Market Space, *Harvard Business Review* January-February, 83–93.
- Law, D., and D. A. Peel. 2002. Insider Trading, Herding Behaviour, and Market Plungers in the British Horse Race Betting Market, *Economica* 69, 327–338.
- Mintel. 2005. Remote Betting-UK-August 2005, London: Mintel.
- Oliver, P. 2007. Financial Binary Betting, Styles, Valuations and Deductions from Data. *The Journal of Prediction Markets* 1(1), 127–146.
- Paton, D., and L. Vaughan Williams. 2005. Forecasting Outcomes in Spread Betting Markets: Can Bettors use “Quarbs” to Beat the Book? *Journal of Forecasting* 24(2), 139–154.
- Pope, P. F., and D. A. Peel. 1989. Information, Prices and Efficiency in a Fixed-Odds Betting Market, *Economica* 56, 323–341.
- Shin, H. S. 1991. Optimal Betting Odds Against Insider Traders, *Economic Journal* 101, 1179–1185.
- Shin, H. S. 1992. Prices of State Contingent Claims with Insider Traders, and the Favorite-Longshot Bias, *Economic Journal* 102, 426–435.
- Shin, H. S. 1993. Measuring the Incidence of Insider Trading in a Market for State-contingent Claims, *Economic Journal* 103, 1141–1153.
- Smith, M. A., D. Paton, and L. Vaughan Williams. 2005. An Assessment of Quasi-Arbitrage Opportunities in Two Fixed-Odds Horse-Race Betting Markets, in L. Vaughan Williams (ed.), *Information Efficiency in Financial and Betting Markets*. Cambridge University Press, Cambridge, UK, pp. 159–171.
- Smith, M. A., D. Paton, and L. Vaughan Williams. 2006. Market Efficiency in Person-to-Person Betting, *Economica* 73, 673–689.

- Smith, M. A., D. Paton, and L. Vaughan Williams. 2007. Do Bookmakers Possess Superior Skills to Bettors in Predicting Outcomes? Working paper, Nottingham Trent University Discussion Papers in Applied Economics and Policy, 2006/7.
- Snyder, W. W. 1978. Horse Racing: Testing the Efficient Markets Model, *Journal of Finance* 33(4), 1109–1118.
- Sobel, R. S., and Raines, S. T. 2003. An Examination of the Empirical Derivatives of the Favourite-Longshot Bias in Racetrack Betting, *Applied Economics* 35, 371–385.
- United Kingdom Parliament. 2004. Joint Committee on the Draft Gambling Bill-First Report. Available at: <http://www.parliament.the-stationeryoffice.co.uk/pa/jt200304/jtselect/jtgamb/63/6302.htm>.
- Vaughan Williams, L., and D. Paton. 1997. Why Is There a Favourite-Longshot Bias in British Racetrack Betting Markets? *Economic Journal* 107, 150–158.
- Vaughan Williams, L. 2000. Index Investment Markets and Information Efficiency: Some Evidence from the UK, *Global Business and Economics Review—Anthology 2000*, 24–29.
- Vaughan Williams, L. 2001. Can Bettors Win? A Perspective on the Economics of Betting, *World Economics* 2(1), 31–48.
- Vaughan Williams, L. 2002. *Betting to Win: A Professional Guide to Profitable Betting*. High Stakes Publishing, London.
- Vaughan Williams, L. (ed.). 2005. Information Efficiency in Financial and Betting Markets. Cambridge University Press, Cambridge, MA, pp. 101–102.
- Weitzman, M. 1965. Utility Analysis and Group Behaviour: An Empirical Study, *Journal of Political Economy* 73(1), 18–26.

DOWNLOAD FULL VERSION HERE: <https://bettingbooks.org/12>

PART VII: Soccer

Download More Sports Betting Books at: www.bettingbooks.org

DOWNLOAD FULL VERSION HERE: <https://bettingbooks.org/12>

This page intentionally left blank

Download More Sports Betting Books at: www.bettingbooks.org

CHAPTER 20

Soccer Betting in Britain

David Forrest

Centre for the Study of Gambling, University of Salford, Salford, UK.

1. Introduction	422
2. Development of Soccer Betting	423
3. Transactions Costs in the Fixed Odds Market	426
4. Early Study of Market Efficiency	429
5. Tipsters	431
6. Fundamental Analysis as an Aid to Soccer Betting	432
7. Technical Analysis	437
8. Sentiment	440
9. The Future of Research on Soccer Betting	443
<i>References</i>	444

Abstract

A distinctive feature of soccer is that odds remain fixed throughout the (lengthy) betting period. Any apparent advantage this offers to bettors has, until recently, been offset by exceptionally high transactions costs. These have fallen substantially since 1999. Nevertheless, partly because of increased efficiency in odds setting, opportunities for bettors to earn reliable positive returns from application of fundamental or technical analysis remain elusive. Some anomalies have been noted, for example, less unfair odds are quoted for more popular teams, but none has been marked enough to be exploited for betting gain. Future research is likely to be oriented away from bookmaker markets to betting exchanges where in-play betting offers opportunities to assess how efficiently prices move in response to fast breaking news.

1. INTRODUCTION

This chapter reviews the surprisingly small, though growing, academic literature on soccer betting. In the past, the themes of the literature have been those most heavily emphasized in studies of the efficiency of sports betting markets in the U.S. principally the questions asked have centered on whether technical and/or fundamental analysis can suggest successful trading strategies for bettors (success being defined variously by positive profits before tax, positive profits after tax, or smaller losses than those that accrue to random wagering). Issues covered under these headings include bias in odds in the favorite-longshot and home/away dimensions, sentiment bias where odds are distorted to reflect the popularity or glamor of particular teams, the ability of statistical models to exploit team strength and form variables to select good value bets, and the usefulness (if any) of professional tipsters. More novel current research exploits the increasing popularity of in-play betting via betting exchanges to evaluate the response of the market to fast breaking news as incidents in a game unfold.

That soccer betting has been relatively little studied compared with the NFL betting market is surprising on a number of counts. First, soccer appears to be the most universally popular of spectator sports. Second, the wagering market is large and growing. Third, betting is organized in an intriguingly different way compared with much studied horse and dog betting markets (worldwide) and team sports betting markets in the U.S. in that bookmakers offer literally fixed-odds betting. Odds on match outcomes are announced several days in advance and remain available throughout the betting period to the start of the game regardless of the emergence of fresh information and regardless of betting volumes. This novel feature seems likely to make the achievement of market efficiency more problematic, since, in contrast to pari-mutuel and conventional bookmaker markets, prices (odds) are not allowed to be influenced by the behavior of large numbers of bettors who may bring to the market fresh information unavailable to odds-setters.

Section 2 briefly examines the history of betting on soccer. Section 3 outlines how and why transaction costs have varied over time and implications for the likelihood of

market efficiency being achieved. Sections 4–8 embody a critical literature survey and discuss how the literature might develop in the future. Section 9 concludes by looking forward to a new generation of studies that will employ betting exchange rather than bookmaker data.

2. DEVELOPMENT OF SOCCER BETTING

Officially tolerated betting on soccer developed in the inter-war period despite the high costs of circumventing legal restrictions (e.g., the companies could not collect stakes prior to the event on which the wager was based) and despite hostility from the sport's governing body (Munting, 1996). Obstructionism by the latter, based on fears that betting might taint the integrity of the competition on the field, was indeed so extreme that, for a short period, Football League fixtures were not published in advance: the idea was that, while this was inconvenient to fans, the betting industry would be prevented from compiling and printing lists of matches on which their customers could bet (Sharpe, 1997).

The approximately 250 firms in the market by the mid-1940s were not, strictly, bookmakers but operated football pools, that is, the product was pari-mutuel and involved betting on outcomes within small groups of games with low stakes and low prizes. The industry was, however, refashioned as a result of the innovation of a new long odds/high prize game, The Treble Chance. This required, and still requires, bettors to select eight matches from a list of some fifty (currently 49) fixtures; to secure a share of the grand prize, the eight games all had to end in draws (ties). This game rapidly became synonymous with the term Football Pools and caught the public imagination as the first long odds/high prize gambling opportunity made available in Britain since the abandonment of public lotteries in 1826. In fact, this game was essentially a lottery itself. Draws appear to be close to random events in soccer (bookmaker odds on the draw show little variance across matches) and no amount of skill was likely to be of assistance in selecting as many as eight matches on one day that would end with this result. So the companies were essentially offering an 8/50 numbers game and there is evidence (Forrest, 1999) that most bettors treated it as such by entering the same sets of numbers each week, regardless of which games corresponded to these numbers.

There is an element of natural monopoly in games like lotto to the extent that bettors are unlikely to support smaller operators who can, for a given takeout, offer a much less attractive grand prize. Hence, the invention of the treble choice format, which is so similar to lotto, initiated a rapid dynamic process that permitted the ultimate survival of only three firms.

If bettors treated the Football Pools like a lottery game, so too did the government, which steadily increased the tax rate to 42.5% of turnover, comparable to the government take-out from state lotteries elsewhere. The total take-out, in excess of 70%, made the pools a conspicuously unfair bet for clients; but this was not due particularly to high profits but rather to the extraordinarily high operating costs (about 30% of turnover; Munting, 1996, p. 228) of the pools companies that failed to embrace new technology

and still, by the 1990s, employed 96,000 agents to collect entries from clients' homes each week (Munting, 1996, p. 135).

In its heyday, the Football Pools was a prominent feature of British culture with over one-third of all adults playing (Sharpe, 1997, p. 72). However, the introduction of an official state lottery in 1994 began a period of rapid decline (documented in Forrest, 1999). The new lotto game employed on-line technology whose low operating costs permitted significantly larger jackpots than the Football Pools. Despite several government tax concessions to the Football Pools industry in the face of declining employment, there is no sign that this type of betting on football can ever recover the mass-market appeal it once held. By the time of the 2007 British Gambling Prevalence Survey (National Centre for Social Research, 2007), only 3% of adults reported having played the Football Pools at all in the preceding year.¹

Conventional bookmaker betting was legalized in Britain in 1960.² The soccer betting this new sector offered probably appealed to a significantly different market than the Football Pools to the extent that it invited bettors to assess the value implicit in quoted odds on each of the three possible outcomes in a soccer game. It was therefore skills-based and appealed especially to soccer fans. Nevertheless, the bookmakers evidently saw themselves in competition with the Football Pools because they organized betting around entry forms, known as coupons just as in the Football Pools sector. It is possible that the distinctive feature of offering fixed odds with respect to soccer was merely the result of wagers having to be made on printed forms and the odds needing to appear on these forms. Whatever the reason, the British system of fixed-odds betting on soccer has been imitated by the betting industries of other European countries.

The new licensed betting offices were initially very restricted in number and location and their main business was horse and dog betting. But easing of regulation, combined with the marked upturn in the popularity of soccer as a spectator sport from 1986 (Dobson and Goddard, 2001, Chap. 7), led to a period of rapid expansion in bookmaker betting on soccer. Global Betting and Gaming Consultants (2001) cited market research findings that, by 2000, nearly 4,000,000 adults in the United Kingdom were betting on sports on a weekly basis and claimed annual UK turnover in sports betting at that time of about £2 billion; they noted that soccer accounted for the huge majority of this market. These findings were consistent with more recent data from the British Gambling Prevalence Survey (National Centre for Social Research, 2007) in which 10% of adult males said they bet with a bookmaker on events other than horse and dog racing and 7% responded positively to a question on whether they had bet on the last World Cup (participation was, however, much lower for females). In addition to domestic demand,

¹An intriguing feature of the Football Pools industry since 1963 has been that, when soccer games are called off (typically due to adverse weather), it does not declare bets void. Rather it bases payouts on the decisions of its own "Pools Panel," a group of experts who predict what the outcome of postponed matches would have been had they taken place. The decision-making process of the Pools Panel is analyzed in Forrest and Simmons (2000a). It appears unlikely that such a process will become widespread in sports wagering markets!

²The new legislation permitted bookmakers to operate retail premises. Previously, only telephone/mail betting on credit was lawful, though there was an extensive illegal sector devoted mainly to horse betting (Miers, 2004, chap. 11).

there is worldwide interest in betting on English matches and much of this is channeled through UK bookmakers.

At bookmakers, product diversification is a notable trend with increased promotion, for high profile matches, of bets on aspects of the game other than the outcome; for example, a double bet involving the prediction of both the exact final score and the identity of the individual player scoring the first goal is reported to be the fastest growing product at bookmakers' shops. Except for Cain et al. (2000) who considered exact score betting (for season 1990–1991), no academic study has yet analyzed the market in these more exotic bets but their emergence appears significant and their operation a topic for useful future research. Classical favorite-longshot bias has almost never been detected in team sports betting markets (though it is documented for individual sports such as golf, Shmanske, 2005, and tennis, Forrest and McHale, 2007) and this is possibly because, given that there are only three possible results (two in some sports) and given that leagues are consciously organized to preserve relatively close balance between teams, the range of odds is almost always very narrow in a set of sports matches compared with a card of horse or dog races. The exotic types of bet may appeal to a new market of relatively risk-loving bettors who can now choose from a wide array of odds with respect to, for example, the identity of the first scorer in a game. The width of odds could be such as to permit classical favorite-longshot bias to emerge and indeed Cain et al. found such a bias in the exact score market. The issue is further addressed below.

The most important development in British betting since the 1990s has been a large fall in transactions costs. This is linked directly and indirectly to the dramatic increase in remote betting opportunities, principally accessed through the internet. The market has become much more competitive because it has effectively been globalized as bettors have gained access to bookmakers located in other countries. Further, the internet has facilitated the emergence of betting exchanges as an alternative to bookmakers; effectively, since an exchange permits traders to lay as well as to back betting propositions, it allows new entry into bookmaking without the barriers of having to secure premises and comply with licensing requirements. The resulting greater competition has been reflected in falls in bookmaker overround on soccer (Deschamps and Gergaud, 2007) and also led to the gradual abandonment of restrictions against wagering on single matches that had effectively kept transaction costs very high. Finally, and most dramatically, the betting tax payable from stakes by bettors with UK bookmakers was abolished in 2001. Eager to compete more effectively in the worldwide market for betting on English Premier League soccer, major UK bookmakers had exploited the emergence of the internet by establishing offshore arms in jurisdictions such as Gibraltar and Guernsey in time for the 1999 soccer season, and these offered the same odds as in the domestic market without the disadvantage of tax deducted from the stake. Tax became essentially optional for bettors at home and abroad as access to the offshore offices of familiar High Street bookmakers was readily available by internet (or telephone). The majority of bookmakers repatriated these offshore operations in October, 2001, when betting duty was abolished following an agreement with the government that the industry would accept a new style of tax, termed a profits tax but essentially a tax on take-out. Competitive forces appear to have prevented bookmakers from passing

on the new tax to clients as overround has decreased rather than increased since the tax on stakes ended.

The fall in transaction costs is of course likely to have strong implications for the achievement of efficiency in betting odds. There is now much more scope for professional bettors to discipline the odds to serve as near unbiased predictors of game outcomes.

3. TRANSACTIONS COSTS IN THE FIXED ODDS MARKET

The British domestic bookmaking market is served by three large national firms with a significant number of regional chains and local one-shop operations. Each of the national bookmakers determines its own set of odds for soccer games and virtually all the independents purchase odds from an odds-setting service, Super Soccer. In all cases, odds for the following weekend's matches are determined on Monday or Tuesday (Sharpe, 1997) and published by midweek. The odds of the largest firms, and of Super Soccer, are printed in the specialist daily betting newspaper, *The Racing Post*, and are also available on the internet. Coupons can be collected from shops in midweek and bets placed any time up to the start of whichever matches are chosen by the client (matches are mainly on Saturday but with a small number of games scheduled for Friday or Monday evenings or Sunday afternoon). Naturally, the sets of odds are highly correlated with each other.

Fixed odds are therefore available on all professional games in England and Scotland (typically 60 matches each weekend) during a betting period of three to six days. That bookmakers are willing to fix odds in this way is, according to Crafts (1985), evidence that soccer is a sport with inherently little scope for useful insider information to exist. Kuypers (2000) argues similarly that the large audiences for soccer and its widespread media coverage make it unlikely that significant information will remain private. However, U.S. professional basketball (for example) enjoys at least as high a profile as soccer, yet Gandar et al. (1998) still find evidence (as Crafts had for British horse racing) that, where odds are permitted to move, those movements are strong predictors of event outcomes. This may be interpreted as demonstration that insider trading has a relevant role even in high profile team sports (and over half of all British soccer leagues on which fixed odds bets are offered are decidedly less than high profile) or, alternatively, that the average assessment of thousands of bettors is likely to be more accurate than even the most skilled of individual odds-setters.

In any case, British bookmakers for long took care to protect themselves against the exploitation of the fixity of odds by those possessing information unavailable to odds-setters at the time odds were set. Traditionally, they did not accept wagers on the outcome of an individual match but rather insisted on a combination bet on three, or even five, fixtures. Bookmakers pointed to Football Association regulations as the origin of this arrangement (the football authorities were concerned about the danger of match-fixing) but variation in the rules over time and across bookmakers suggest that they were just part of a business strategy rather than something imposed exogenously.

Rules against singles bets constituted an important protective device against those with private information. Suppose one had known, as if with certainty, the outcome of a particular match. If the bookmaker accepted only “treble and upwards” bets, one would have been sure to win only if one had chosen two other games and made a series of six treble bets across which the known result from one’s own match was linked to each of the six possible pairs of results from the other two games. Given multiple bookmaker commissions, a positive expected return was unlikely to be available despite the possession of sure insider information.

An alternative and cheaper way of wagering on a single fixture emerged, however. Bookmakers permitted a double bet on any individual match where the two parts of the bet related to the outcome (home lead, draw, or away lead) at half-time and at full-time. The odds for this sort of combination bet were not individually set on a match-by-match basis, but were linked to the odds offered in the main coupon on the team one was backing to win. For example, suppose the main coupon quoted odds of 6/4 against team A defeating team B³ and one wished to support team A. A bookmaker offered the following odds with respect to combinations of half/full-time outcomes:

- A leads at half-time, A leads at full-time—7/2
- Sides level at half-time, A leads at full-time—9/2
- B leads at half-time, A leads at full-time—22/1

Had a single bet been permitted on A at 6/4, one would have needed to have spent £0.40 to be able to collect £1 following an A victory. In the absence of singles being permitted, one could instead have made three double bets as above, investing respectively £0.222, £0.182, and £0.044, so as to cover all combinations ending in an A victory. But this would have meant spending £0.447 to achieve the end of receiving £1 at the bookmaker window in the event of success. The true odds were therefore not the quoted 6/4 (1.5 to 1) but only 1.24 to 1. This shading of the odds was of very similar size for all possible odds of an A win and could be understood as an insurance premium to protect bookmakers against insider dealing: one could bet on a single match but had then to accept odds approximately 17.7% less favorable than published odds.

Restrictions against singles betting that existed until well after 2000 were often mentioned by writers on market efficiency in soccer betting but were then generally ignored. Published odds were always employed in evaluations of possible profitable wagering strategies in this period even though these were not actually available on a single game. Dobson and Goddard (2001, p. 414) justified this by pointing out that “the requirement to place bets on several matches simultaneously does not affect the expected return on any bet.” While this is true, the rule did raise the transaction cost of placing a bet

³In the British domestic market, odds are always quoted in the form a/b which signifies that a successful wager of b on an event occurring will win the better a (and in addition the stake of b is returned to the successful bettor). Odds may, of course, be expressed in alternative ways. Decimal odds, used in the international soccer betting market, quote the return (including stake) paid to the bettor for a unit stake. Hence, the odds 6/4 would be quoted in the decimal system as 2.5. Probability-odds corresponding to a/b would be $b/(a+b)$; for example, 6/4 becomes 0.40, signifying that an expenditure of £0.40 is required to receive a £1 payout (return of stake included) if the event occurs.

with respect to the one fixture for which one may have had specialist knowledge, either because of access to private information or because one supported (and therefore readily absorbed public information concerning) a particular team. It also reduced, of course, the feasibility of bettors enjoying a positive expected return if they simply spotted a mistake by odds setters regarding the outcome of one particular match.

The rules against singles could, then, be circumvented, but only at high cost. These rules raised transaction costs to very high levels relative to sports betting markets in other parts of the world or at least did so until increased competition emerged in about 1999. Transaction costs had three elements. First was the bookmaker take, 10.5% of match handle, assuming a perfectly balanced book (i.e., the same payout in a given match regardless of its outcome).⁴ Second was the 17.7% premium for betting on a single match. Third was the tax of 9%. The total take-out rate from all three deductions was $\{[1 - (0.91 \times 0.823)] + 0.105\}$ or 35.6%. This level of transaction cost was so high (not far short of the 50% take-out on the National Lottery) that one could envision considerable biases in odds on individual games existing but failing to generate profitable trading opportunities. The role for professional bettors must have been severely circumscribed at this level of transaction costs.

Terrell and Farmer (1996) present a model exploring the roles of professional and pleasure bettors in a pari-mutuel market. They use it to account for the well-documented favorite-longshot bias in dog and horse racing. In isolation, pleasure bettors generate bias in the array of odds because of lack of knowledge of the true win probabilities attached to each runner (one could suggest that their wagers are influenced by random extraneous factors such as the name or color of the horse or dog). The professional bettors wager whenever the expected value of a bet is positive and, in the absence of transaction costs, their activity will eliminate biases in the array of odds. However, given positive transaction costs, correction of odds will remain incomplete and inefficiency will be preserved in the market in the sense of some bets offering less bad returns than other bets. High transaction costs therefore spawn inefficiency in the odds. It can be argued similarly that, in a bookmaker market, bookmakers may quote prices that take account of bettor preferences as well as underlying true probabilities and that systematic inefficiency is sustainable if transaction costs are sufficiently high that professional bettors can still not identify any strategies that would generate positive returns.

However, as noted above, transactions costs fell substantially about soccer season 1999–2000. There was downward pressure on bookmaker overround; the government removed betting duty; and restrictions against singles bets were relaxed so that it became possible, first for Premier League games and then for any professional match, to place singles bets. These falls in all the elements of transaction cost should have induced a reduction in, or elimination of, biases in the betting odds because they opened up possibilities for professional bettors to earn positive returns at the expense of bookmakers if the odds were biased even to a relatively small extent.

⁴The calculation of take-out is based on the size of overround. In a typical game, the sum of probability-odds was 1.115, which implies that bettors would have to spend £111.50 to generate a certain payout of £100 after the result was known. Take-out was thus £11.50/£111.50 or 10.5%. An interesting feature of the market but one unexplained in the literature is the very low variance in the overroundness of the book across matches.

4. EARLY STUDY OF MARKET EFFICIENCY

For some years, Peel and Pope (1989) remained the only published analysis of efficiency in the fixed odds market. While the literature review here is generally organized thematically, this paper is considered first because it covered a number of aspects of efficiency and, to some extent, provided the template for subsequent studies. Despite its date of publication, the analysis relates to odds for matches played in the season of 1981–1982 when betting volumes and regulatory background were substantially different from even 1989. Peel and Pope note that their study of the odds actually revealed one match where, even accounting for the tax on stakes, a risk-free arbitrage return had been possible because of differences across bookmakers in relative odds on each outcome. Generally, they established in a series of regressions that there were systematic differences in the patterns of odds set by each of four bookmakers such that pure arbitrage opportunities would have been numerous in the absence of a (then) 10% tax. By contrast, later writers found a close relationship between the odds of different bookmakers and (Dixon and Pope, 1996) no arbitrage possibilities. This is to be expected given that every High Street now has several competing betting offices and prices are readily compared on internet sites. In 1981, regulation of the gambling industry permitted only one bookmaker shop in a given location and even telephone betting was underdeveloped; hence there was little competitive pressure on bookmakers.

Nevertheless, Pope and Peel found little evidence of inefficiency in individual bookmaker odds. Their lead result reports estimation of a linear probability model which investigates the relationship between the incidence of a named outcome in a game (e.g., home win) and the probability that that outcome will occur according to the published bookmaker odds. As was standard in subsequent articles, the implicit probability of a home win (bookprobH) was obtained from the ratio of the probability-odds of a home win to the sum of the probability-odds of all three outcomes (i.e., published odds were scaled to account for overround).⁵ Their full model is thus:

$$\text{prob (home win)} = a_0 + b_0 \text{bookprobH} \quad (1)$$

$$\text{prob (draw)} = a_1 + b_1 \text{bookprobD} \quad (2)$$

$$\text{prob (away win)} = a_2 + b_2 \text{bookprobA} \quad (3)$$

Weighted least squares was employed to account for heteroscedaseity. Strictly, the three equations comprised a system in that the observed dependent variable was binary and the three events were mutually exclusive in respect of each observation. Nevertheless, Peel and Pope estimated the three equations singly. They used 1,066 observations. The null hypothesis of efficiency required $a_0, a_1, a_2 = 0$ and $b_0, b_1, b_2 = 1$.

In fact, for home and away equations at each of the four bookmakers, estimates of a and b were never significantly different from zero and one, respectively, indicating no

⁵The notation throughout what follows (e.g., bookprobH to signify implied bookmaker probability of a home win) is my own rather than that used in the original, so as to preserve consistency of notation throughout the chapter.

evidence of inefficiency. For draws, b was never significantly different from zero. The last result is explainable by very low variance in bookmaker odds with respect to the draw. To Pope and Peel, and to subsequent authors, it was anticipated that profits would accrue to anyone who could produce a model that explained draws (since bookmakers treat the probability of a draw as a near constant) but it may of course actually be the case that draws are close to being truly random events.

The species of model employed to test for efficiency was to be adopted or adapted by later analysts of soccer betting and was very similar to the model typically used in early studies of betting on America's National Football League. However, the choice of model is problematic in two respects.

First, U.S. studies related to a point spread betting market where the test for efficiency focused on estimated coefficients in the regression of the actual margin of victory for a team on that predicted by the bookmaker spread. In that context, linear regression is entirely appropriate. However, in the soccer market, the outcome of a match is defined by a binary variable (the team wins or not). So a straightforward linear model in the American literature is translated to a linear probability model by analysts of the British market. But the linear probability model described by Greene (2000, p. 813) is rarely used, not the least of its problems being that fitted values for probability may be outside the unit interval. It might seem more natural, therefore, to adopt non-linear estimation, such as probit. However, the legitimacy of employing probit here could be disputed. In our context, and unusually in economics, the null hypothesis is defined not only by the significance, or the value of the marginal effect, of an independent variable but also by the functional form of the relationship. Specifically, if *bookprob* is measured on the one axis and true probability on the other, the relationship between the two should be linear (and described by a 45° line). A probit model *imposes* a specified non-linear relationship and cannot therefore be used to test a null hypothesis that includes linearity. Presumably (though no one ever discussed it) this is why many subsequent authors continued to imitate the Pope-Peel testing procedure and used linear probability for estimation. Of course, competitive balance in soccer is quite high, and probabilities close to zero or one for any match outcome are therefore so rare that it is not very practically relevant that linear probability models could, in principle, fit values outside these boundaries.

The second limitation of the test employed in Pope and Peel is that it has low power in that it essentially fails to test for specific biases (that may offset each other). For example, suppose one suspected that, of two possible biases (home team/away team, favorite/outsider), either or both may exist here. A non-zero constant would result from *either* bias being present and the test cannot therefore distinguish between them. Worse, both biases may be present but, because home/away and favorite/outsider status are correlated (home advantage is very strong in soccer and home teams are therefore much more often than not the shorter odds team) they may be sufficiently offsetting for the efficiency test still to be satisfied. Possible solutions to this problem are discussed in Section 7 below.

As a supplement to their principal results, Pope and Peel present estimation of similar regressions in which a logit model is employed and the values for *bookprob* as a

regressor are given by the geometric mean of the *bookprob* values at each of the four bookmakers. While the non-linearity of the model would make it inappropriate for testing efficiency directly, they in fact use it to generate fitted values that are then compared with the bookmaker odds themselves to identify bets offering superior value. The results are erratic but at two of the bookmakers a strategy based on the model appears to yield improved returns (though at one it worsened returns by a huge amount). What is being picked up in these findings is unclear but may be related to a possibility that there would be a reward to betting at bookmakers that offer odds in a particular game that are relatively far from the consensual odds on that game: when a bookmaker offers best value odds on a particular result, those odds may be liable to be too generous.

Deschamps and Gergaud (2007) pursued this idea for a later period by measuring the return to a strategy of betting at best odds on matches between 2002 and 2006 where the variance in odds across bookmakers was particularly large. They report that the strategy was capable of improving returns but these remained negative; and there was no out-of-sample testing.

Finally, Pope and Peel examine another familiar issue, whether tipsters can provide useful guidance to bettors. If they can, odds must be inefficient in failing to incorporate whatever information underlies tipsters' calls. The testing method is to assess returns to placing bets according to whatever is the majority verdict among six tipsters in national newspapers. In the main sample period, there is no evidence that such a strategy can reduce bettor losses. However, in a sample of 225 games in the final five weeks of the season, following the tipsters would have yielded a positive (pre-tax) return at three of the four bookmakers. This finding finds an echo much later in that an elaborate statistical model for forecasting match results (Dobson and Goddard, 2001, and Goddard and Asimakopoulos, 2004) was found to provide successful guidance to bettors only in the concluding weeks of the season (April/May). These are only two isolated pieces of evidence, but could suggest that bookmakers do not sufficiently modify their odds to take account of factors relevant at the end of the season such as whether or not a particular game has championship, promotion, or relegation significance for one team in a match (but not for the other team).

The most serious weakness of Pope and Peel was that their data related to one season only. This was forgivable in a pioneering contribution but it was not as excusable that most subsequent authors continued to base analysis on only one or two seasons of data. Sauer (1998) documents how unstable findings can be in team sports betting markets over time and it is unfortunate that it has, until recently, been common in the soccer literature to base efficiency tests on one or two seasons only rather than on a run of seasons as in several NFL studies.

5. TIPSTERS

Subsequent to Pope and Peel, there have been no further direct tests of the ability of soccer tipsters to point bettors in the direction of superior returns. However, Forrest and Simmons (2000) compare tipsters' performances with that of a very simple forecasting

model based on a small number of team strengths and form variables. Generally, they confirm the low information content embodied in newspaper tips.

Their testing methodology was based on an ordered logit model in which the three possible outcomes of a match were presumed to have a natural ordering from home win though draw to away win. Three equations employed as regressors (1) team strength and form indicators alone, (2) tipsters' forecasts alone, and (3) team strength and form indicators and tipsters' forecasts together. A series of nested hypothesis tests demonstrated that when team strength and form variables were added to the model including only tipsters' forecasts, the performance of that model improved. Hence, tipsters appeared not to take proper account of even the very obvious indicators used in the particular, and crude, statistical model. When each tipster's forecasts were added to the statistical model, its performance failed to improve (except in the case of one newspaper). Tipsters appeared therefore generally not to possess any useful information extraneous to the very basic statistical model. The resoundingly negative assessment of tipsters' capabilities was qualified only by the finding that some small improvement in the performance of the statistical model was evident when it was supplemented by an indicator variable signifying that the tipsters displayed consensus in their forecasts. Ordered logit analysis designed to explain tipsters' decisions was used to explore why tipsters perform poorly. A comparison of models that determine tipster and real results indicated that tipsters place undue weight on recent match results relative to season-long indicators of strength and so perhaps tend to mistake noise for signal.

The lack of insight displayed by newspaper tipsters and commentators is a result familiar from studies for other sports (see, e.g., Song et al., 2007). But it is not necessarily an indication that expertise cannot exist in forecasting soccer results. It appears unlikely that newspapers will find it worthwhile to hire high grade talent, because football tips probably comprise a trivial part of their output. However, high grade talent, if it exists, will be very valuable to bookmakers, since mistakes in odds-setting can have serious financial consequences, especially under a regime where odds are fixed through the betting period. Testing the professionalism of odds-setters by examining whether statistical models can help bettors secure positive returns has yielded very different results from tests of the professionalism of tipsters.

6. FUNDAMENTAL ANALYSIS AS AN AID TO SOCCER BETTING

Several attempts have been made to model the process by which soccer match results are determined to test whether fundamental information on teams can be used to raise bettor returns. Their potential for success depends either on odds being set by reference to an implicit model that is inferior to the authors' statistical model or on bookmakers deliberately building biases into the odds for commercial reasons (for example to exploit cognitive biases among the generality of bettors or to take account of preferences for betting on particular teams). Generally these exercises have tended to yield only limited

potential for better profit. The implication is that bookmaker odds do not depart from efficient odds by a sufficient amount to allow informed bettors to overcome transactions costs.

The first studies were based on a strand in the statistical literature (traceable to Maher, 1982) which proposes that the scores of two teams in a soccer game may be modeled as independent Poisson distributions with means reflecting the goal-scoring record of the team and the goal-conceding record of its opponent. Dixon and Coles (1997) and Rue and Salvesen (2000) operationalized the model to allow it to generate estimated *ex ante* probabilities of match outcomes that may be compared with bookmaker odds. If the statistical model can thereby identify a set of bets that will earn superior returns, the market may be termed inefficient with respect to fundamental information (the goal-scoring and conceding records of the teams in a game).

Dixon and Coles tested for efficiency by evaluating the returns to strategies based on placing a bet on an outcome in a game whenever the ratio of the model probability to *bookprob* exceeded some predetermined figure. With a filter setting the required ratio at 1.2, the return to such betting was reported as “borderline significantly different from -0.11 , the expected return to a random betting strategy.” This is evidence against efficiency in the odds to the extent that some bets appear to yield higher returns than others. Whether there was any possibility to find a potentially profitable betting strategy is much more open to question. From the authors’ Figure 4, allowing for the extra commission charged at that time for betting on a single match via the half-time/full-time route would have required the filter-ratio to be set at approximately 1.3 for returns to become positive. Dixon and Coles did not present details of the number of bets identified at different values of the filter-ratio but it is a familiar feature of horse betting that winning strategies require considerable effort and data analysis while identifying frustratingly few occasions on which a bet should be placed (Crafts, 1994).

Rue and Salvesen employed their variant of Poisson modeling to assess whether they could have made positive returns from a strategy which sought to place a number of bets such that expected profit was maximized subject to a constraint on its level of variance. The experiment was for the top two divisions of the English leagues in the season of 1997–1998 and betting was presumed to be with a large internet bookmaker based in the Caribbean. The model is applied only in the second half of the season to permit adequate current season data to be available for the statistical model to exploit. Claimed profits were emphatically and startlingly positive for both the Premier Leagues (39.6%) and Division One (54%) though the number of bets made in the simulation was small (112 in the two divisions in the season) and the return in the Premier League case depended substantially on a single winning bet on a team with exceptionally long odds (12.8/1). The experiment on its own was therefore unconvincing as a formal test of efficiency but it would be interesting if it were to be implemented for a longer period. It would also be of interest to compare returns from a less restrictive strategy that incorporated risk-neutrality in the bettor’s utility function.

From the economic rather than the statistical tradition has emerged a computationally less demanding methodology for modeling match outcomes, that of discrete choice models, ordered logit or probit. These techniques model match outcomes directly rather

than indirectly according to the probability distributions of the scores by each team. Ordered logit or probit models are employed by Forrest and Simmons (2000b) in a study of the efficiency of tipsters and by Koning (2000) in a paper on competitive balance while Kuypers (1999), Dobson and Goddard (2001), Goddard and Asimakopoulou (2004), and Graham and Stott (2008) develop ordered probit models to assess their efficacy as a source of profitable betting strategies. Goddard (2005) demonstrates that the approach of modeling results directly through such regression models yields very similar results when compared with the older tradition of deriving probabilistic forecasts of outcome from estimates of probability distributions of goals scored by each team.

Kuypers (2000) used an ordered probit model⁶ based on 1,773 matches in season 1993–1994. Unfortunately, presumably because of space limitations, he does not present estimates of the model itself but only a list of variables, which are obvious team strengths and form indicators. He used his estimates to select bets in the following season according to whether the ratio of model probability of success to the probability implicit in bookmaker odds exceeded 1.1, 1.2, and so on. He comments that this procedure yielded neither evidence of inefficiency nor of potential profits for the bettors. However, when the model was stripped down to include only one team strength indicator (difference in league position) and supplemented by the inclusion of odds as a regressor (specifically the odds of an away win), positive post-tax returns were obtained in the hold-out sample season, 1994–1995. Moreover, large numbers of bets were identified (over 1,000, even with the filter set at 1.3). The author remarks that implementation would have been complicated, and returns compromised, by the restriction against single bets but does not calculate returns actually available given the then bookmaker rules. Nonetheless, it would appear from his raw results that the application of a 1.2 filter would have generated at least a break-even outcome for a set of 1,681 bets. This is an impressive result as evidence of inefficiency in the market. My reservation is that the final simplified model uses only a small proportion of the variables included initially: the degree of ad hocery in selecting one from the many possible combinations of variables implies that a longer-run test of the efficiency of the specified model would be desirable. This is particularly true since the final model includes not only fundamentals but also odds data and examples of inefficiency derived from patterns in the odds themselves have frequently been noted (Sauer, 1998) as merely transitory phenomena in sports betting markets.

Dobson and Goddard (2001) and Goddard and Asimakopoulou (2004) present an ordered probit forecasting model estimated from over 50,000 matches in a 10 year period. The variety of data captured by the regressors is impressively wide and includes, for each team, league points won in the previous 12 months and the year before that (adjusted according to the division in which it was playing) and the result of the most recent home match (or away match as appropriate), the second most recent home (away) match and so on (the number of lags being determined by statistical significance). A necessarily ad hoc but influential variable captures whether the game had been significant to one (but not the other) of the teams in terms of championship, promotion,

⁶He remarked that estimation by ordered logit had yielded very similar results to those he reports.

or relegation issues. The estimated model appears successful in identifying key indicators to take into account in predicting likely results. Formal tests of whether the model yields forecasts (defined as the outcome—home win, draw, or away win—with the highest model probability) superior to random forecasts (constrained to have the same home-draw-away frequency distribution as the model's forecasts) confirm its informational content. Nevertheless, the inherent difficulty of forecasting soccer games is underlined if one calculates, from the table on p. 183 of the Dobson–Goddard presentation, that the model yields only a 43.06% success rate. Simply calling a home win every single match would have converted the success rate to 46.66%! Explaining draws appears a particular difficulty for the model (as it is indeed for bookmakers) and these are remarked by the authors to be “near-random” events.⁷ Of course, for betting purposes, the key output of the model is not which outcome has the highest probability but which offers the best value bet when model and bookmaker probabilities are compared.

Goddard and Asimakopoulous offer both “regression” and “economic” tests of betting market efficiency based on the results of their model estimated over 10 years and then applied to games in seasons 1999–2000 and 2000–2001. They estimate the Pope-Peel (weighted least squares, linear probability) model, adding a variable as follows:

$$\text{prob (home win)} = a_0 + b_0 \text{bookprob}H + c_0 (\text{modelprob}H - \text{bookprob}H) \quad (4)$$

$$\text{prob (draw)} = a_1 + b_1 \text{bookprob}D + c_1 (\text{modelprob}D - \text{bookprob}D) \quad (5)$$

$$\text{prob (away win)} = a_2 + b_2 \text{bookprob}A + c_2 (\text{modelprob}A - \text{bookprob}A) \quad (6)$$

where *modelprob* is the probability of a particular outcome (*H*, *D*, or *A*) according to their estimated forecasting model. Efficiency conditions are now $a_0, a_1, a_2 = 0$, $b_0, b_1, b_2 = 1$, $c_0, c_1, c_2 = 0$. The coefficients c_0, c_1 , and c_2 will be zero if the bookmaker odds (of which *bookprob* is a transformation) already take proper account of all the fundamental information exploited in the statistical model. That the estimate of c is significantly positive in all three equations is *prima facie* evidence of inefficiency. However, on splitting the sample into four segments by month (August–October, etc.) c retains its significance only for the last quarter of the season. The failure of bookmakers to properly use the information in the model is therefore restricted to the period from March onwards, the closing weeks of the season. Results of economic tests where returns to a betting strategy (in this case based on the statistical model) are calculated directly, point in the same direction. Placing a bet on every match according to whether a home, draw, or away bet has the highest expected value (according to the forecasting model) delivers mixed results across the two seasons though returns appear superior to those that would be associated with random betting. However, in April and May of each season, pre-tax returns of 8% would have followed from the strategy. This is fairly convincing evidence of market inefficiency in the late season. One can but speculate as to the source of the

⁷ If it is true that draws are random events, this undermines the claim that ordered, rather than multinomial, probit is the appropriate modeling technique.

inefficiency. One possibility is that the model is more effective late in the season because it has a larger volume of relevant current season data to process. Another is that many games at that time of year have particular significance for one team but not the other; this situation of asymmetric incentives is represented in the statistical model but may not be given due weight in odds-setting. It would be interesting to repeat the experiment while omitting the significance variable from the ordered probit to see whether or not superior returns were still associated with model-based betting late in the season.

Notwithstanding the apparent potential for employing a statistical model to secure positive returns late in the season, the literature reviewed so far has tended to find difficulty in establishing potential for using statistical modeling to secure positive as opposed to merely less negative returns. However, all these contributions relate to a period when transactions costs in soccer betting were very high compared with most forms of wagering. As noted above, transaction costs have fallen markedly in the current decade and this would appear to open up the possibility that less negative returns could now be converted to positive returns with an adequately data rich statistical model. However, bookmakers themselves now have greater incentive to set efficient odds given that transaction costs are less of an impediment to professional bettors.

Forrest et al. (2005) investigate whether odds were in fact set more efficiently as transaction costs fell. For the five seasons leading to 2003, they compare the forecasting performance of a richer version of the model used by Goddard and Asimakopoulou (2004) to that of a model where the only predictor variable is bookmaker odds. Results (consistent across tests for five different bookmakers) are that, based on standard evaluation criteria such as Brier scores, the fundamentals model is superior early in the period but inferior late in the period. In a series of likelihood-ratio tests, adding probabilities generated from the fundamentals model to the model that utilized only odds improved its performance but only in the first three seasons: by 2003, odds appeared to incorporate all relevant statistical information. In further likelihood-ratio tests, adding bookmaker odds to the fundamentals model improved its forecasting performance, demonstrating that odds-setters made effective use of information not available to the fundamentals model.

The set of results indicates that bookmakers responded to greater pressure to set efficient odds by adopting approaches that enabled them to do so (e.g., by themselves employing statistical modeling, to supplement their intuition). When Forrest et al. assessed the rate of return from using the statistical model to identify value bets, they found that, by placing such bets at the best odds offered across the five bookmakers, they could generate a superior outcome to random betting but this was never better than close to break-even. Graham and Stott (2008) analyzed matches from 2001 to 2006 and use of their model also failed to generate positive returns (indeed losses were large from their strategy: they speculate that identifying any biases in odds-setting cannot be exploited for betting gain because bookmakers have the offsetting advantage of incorporating information, on injuries and so on, not in a statistical model). The fall in transactions costs has therefore failed to transform the prospects for employing statistical modeling to beat the bookie. This appears to be because odds-setting itself has become more refined.

It is a common finding in the study of sports and non-sports prediction markets that betting odds better forecast events than experts or statistical models. But this is usually explained by the fact that closing odds reflect diverse information from large numbers of participants, induced by the prospect of financial gain to reveal their assessment of a future event. However, soccer betting is different because odds do not change as new money arrives in the market. The accuracy of odds here depends solely on the expertise of odds-setters. Nevertheless, the standard result roughly holds (employment of statistical modeling does not yield reliable positive returns), presumably because bookmakers have a sufficiently strong financial incentive to gather and process information accurately into the odds they post. This incentive has increased over time as transaction costs have fallen and odds have consequently become more efficient.

7. TECHNICAL ANALYSIS

The authors who have examined efficiency with respect to fundamental information on teams' playing records have also tended to present analysis designed to test whether historical patterns in the odds themselves, and the success of corresponding bets, can be exploited to generate superior returns to wagering. Themes emphasized have been potential inefficiency in the favorite-longshot and home/away dimensions. The issue of favorite-longshot bias is perhaps of particular interest given that the direction of bias has been reported to be opposite in American team sports compared with horse and dog racing (Woodland and Woodland, 1994, 2001, 2003).

The extent of any favorite-longshot bias in team sports markets in fact proves difficult to disentangle from any home/away bias that may be present. American studies (e.g., Golec and Tamarkin, 1991, and Vergin and Sosik, 1999) have claimed betting on home teams to be a superior or even profitable betting strategy in the case of football.⁸ Various commentators on the soccer market have also recommended placing bets in favor of home teams. But the evidence offered often conflates favorite-longshot and home/away bias and is based on statistical analysis of relatively short periods.

Cain et al. (2000) consider just one season, 1991–1992, and point to superior (but still negative) returns to bets on teams that are strong favorites; but, since strong favorites are almost always playing at home,⁹ any bias they capture could in fact result from how bookmakers choose to price home advantage into the odds. By contrast, Dixon and Pope (1996) employ data from two bookmakers over 1992–1996 and find that backing longer odds teams yielded better results; they also report that backing home teams (or draws) yielded lower losses than betting on away teams. Their results are suggestive but any findings of systematically different returns as between *either* home and away bets *or* short-odds and long-odds bets could result from the presence of either or both

⁸Gandar et al. (2001) find no such evidence for betting on basketball over the period of 1981–1997 or baseball over the period of 1989–1999.

⁹Home advantage is very strong in English soccer with about twice as many wins recorded by home as by visiting teams.

types of bias as discussed above. Multivariate analysis is required if the goal is to test for the presence or size of those two distinct biases.

Kuypers' (2000) summary statistics (his Tables 2 and 4) of returns to possible bets on a sample of over 3,000 English games in seasons 1993–1994 and 1994–1995 give no indication of systematically different returns from betting in different parts of the odds distribution though overall losses were smaller for home/draw than for away bets. His formal analysis regresses implied probability (i.e., *bookprob*) on outcome probability (proportion of successes in each of 24 odds categories) for home wins, draws, and away wins separately and reports slope coefficients never significantly different from one. This is, though, fragile evidence of his weak efficiency finding because the omission of a constant from the regression equations biases the slope coefficients towards one. From his summary data, bias, particularly in the home/away dimension, requires further evaluation as to statistical significance.

Dobson and Goddard (2001, pp. 407–410) replicate the regression test proposed by Peel and Pope [1989, Equations (1–3) above] for data from a single season (1998–1999, 1,568 matches) with odds from one leading bookmaker. Estimates of the coefficients on *bookprob* are insignificantly different from one in the home and draw equations but the point estimate of the coefficient on *bookprob A* is 0.81, which is significantly different from one at the 10% level (two-tailed test).¹⁰ This could be taken as indicative of reverse favorite-longshot bias within the subset of possible bets, away. The significance for betting is illustrated by a calculation that betting on the 500 away teams with the longest odds (31.9% of the sample) would have generated a positive pre-tax return of 8% (in sample). Betting long on away teams was therefore recommended and the authors noted optimistically that, subsequent to the season to which their analysis was applied, tax-free betting had in fact become feasible on the internet.

Unfortunately, the exhortation to bet long on away teams turned out to be one of those strategy rules which, once spotted, promptly disappear (Sauer, 1998). In a sample of 1,489 matches from the following season (with odds from Super Soccer), I calculated the return to betting on the 30% of cases where away odds exceeded 3/1. The loss was more than what would have accrued due to random betting! Further checks confirmed that the trading rule would have yielded a profit with Super Soccer odds in the season to which the Dobson-Goddard results pertained. Presumably, if profitable trading rules tend, in general, to be ephemeral, this is because there is some learning process within markets. However, in this case, the availability of tax-free betting would have necessitated the removal of the bias if bookmakers were not to face large losses from the activities of alert professional bettors.

A limitation of the Pope-Peel/Dobson-Goddard model is that, while the data are partitioned to consider home, draw, and away bets separately, any inefficiency in *either* the favorite-longshot or home/away dimensions will tend to shift both the constant and the slope coefficient away from the null hypothesis values of zero and one, respectively. Golec and Tamarkin (1991) recommended in the context of handicap (point spread)

¹⁰When Dobson and Goddard adapted the Pope-Peel equations to include information from their forecasting model, Equations (4), (5), and (6) above, slope coefficients were never significantly different from one.

betting that favorite-longshot and home team bias could be evaluated separately by selecting randomly from each match a focus team i so that the observed value of the dependent variable (point spread achieved in their context, win or not in our soccer market) is defined from the perspective of i . This permits an indicator variable for home advantage to be employed as a regressor and permits in principle the separate evaluation of home/away bias. This would lead to the following specification of a Pope-Peel type model:

$$\text{prob}(\text{win}_i) = a + b_1 \text{bookprob}_i + d \text{home}_i \quad (7)$$

where bookprob_i is the probability of a win for i implied by the odds and home_i is a dummy variable set equal to one if team i enjoys home advantage.¹¹

Yet this procedure raises a problem of its own because, in addition to sampling matches from a population, it further samples a subset of bets (i.e., those picked out in the randomization). This makes the standard errors unreliable for hypothesis testing.

Forrest and Simmons (2001) therefore specify an equation similar to Equation (7) but estimate it in repeated trials (in their case 20 trials). The statistical significance of the variable may then be assessed by counting the number of individual trials in which the coefficient appears as significant and comparing the figure with a critical value determined as in Snedecor and Cochran (1967). This methodology has been followed by Kossmeier and Weinberger (this volume) in their study of odds available from Austrian-based internet bookmakers.

Compared with Equation (7), Forrest and Simmons' equation added an interaction term between bookprob and home and they also included a variable to capture any sentiment bias. Their analysis pertains to four seasons from 1997–1998 to 2000–2001. There was no evidence of either favorite-longshot or home/away bias except in one season, 1998–1999, where they identified the same apparent advantage to backing away longshots as was noted by Dobson and Goddard.

Deschamps and Gergaud (2007) offer economic rather than regression testing. They tabulate returns to betting in different odds categories for matches between 2002 and 2006, partitioned into home, draw, and away bets. For home and away bets, they claim classic favorite-longshot bias as in horse and dog racing but the opposite result is noted for draw bets (where, however, they omit to mention that the variance in odds across matches is very low). Their findings are not subject to significance tests but it is clear from their charts of returns on home and away win bets across 20 odds categories (with equal numbers of observations) that there is no clear relationship between odds and returns except for there being very poor returns in the two most extreme categories at the longshot end of the scale. This is similar to Cain et al. (2000) where a conclusion of favorite-longshot bias is based only on returns in matches with extreme odds. Deschamps and Gergaud also find no evidence of differential returns as between home and away bets, though they note superior (but still negative) returns to draw

¹¹Bias over the home/draw/away dimension could equally be tested by randomizing between home team, draw, and away team in choice of focus for each observation and adding a draw dummy variable as a further regressor.

bets in their final two seasons. This was because of an increase in the number of draws generated in English football (i.e., soccer) in those seasons. It is hard to see how this could have been foreseen, so this should not be interpreted as evidence of inefficiency.

The literature on favorite-longshot and home/away bias has benefited from authors using longer runs of data in more recent contributions. Collectively, the studies cover seasons over a wide range of time. They continue to find very limited evidence of systematic bias that could be employed to secure superior, let alone positive, returns even with much lower transaction costs than in the past.

8. SENTIMENT

The language and approach of the literature on sports betting reflects the extent to which its paradigm has been provided by the disciplines of finance and financial economics. There is almost a presumption that prices (odds) *ought* to be efficient.¹² Pope and Peel (1989), who originated the literature on soccer betting, stated it to be “a reasonable assumption that the bookmakers’ posted odds reflect their expectations” (p. 323) and go on to motivate their study of odds as an opportunity to evaluate rationality through “directly observed expectations of expert agents.”

But one cannot take it for granted that odds reveal bookmaker expectations as to event probabilities.¹³ Bookmakers in our particular market are solely responsible for setting (fixed) odds but their behavior is likely to be motivated by expected profit; and efficient odds do not necessarily follow from profit maximization.

The bulk of the literature in sports betting markets fails to acknowledge a key defining feature of team sports (as opposed to horse racing or individual sports). Those who take an interest in a team sport almost always identify with a particular club as “their” team. Part of the market for betting on team sports events will comprise stakeholders, that is, those who care (often very passionately in the case of soccer) about the outcome of the event. In principle, fan bettors could typically wager against their team as an insurance against the disappointment of the team losing the match. However, I will proceed on the assumption that typical fans would bet, if at all, only on their own team because betting on the team permits them to become stakeholders to a greater degree and is a way of displaying loyalty, just as might be the purchase of other, more tangible items of club-related merchandise.

The presence of such bettors in the market will have to be taken into account by a profit-maximizing bookmaker. This is a striking difference from horse betting where (a horse’s connections apart) it is rare for many bettors to have a non-financial interest in the outcome of a race. This difference could underlie contrasts in findings between

¹²In another context (that of his “Critique of Welfare Economics”), Little (1957) discussed how a condition defined by the word “efficient” comes to be viewed as a desirable goal just because efficient is a loaded word with positive connotations. He identifies a problem in economics of using terms with “persuasive definitions.”

¹³Shin (1991) demonstrated in a theoretical model of horse betting that favorite-longshot bias may be a rational response by bookmakers to risks posed by insider trading.

the team sports and horse betting markets.¹⁴ So long as transaction costs are sufficiently high to inhibit participation by professional bettors, persistent bias in sports betting odds may be explained by reference to demand from fan bettors.

How will the presence of fan bettors in the market modify odds-setting? Kuypers (2000) was the first to try to build a model where a profit-maximizing bookmaker would set inefficient odds. The context is a market where one group of bettors is neutral and, like the bookmaker, has an objectively correct assessment of outcome probabilities. A second group of bettors is committed to one particular team in the match; in his example they are all supporters of Manchester United but the model could just as well specify that there is a level of net support in favor of Manchester United because it happens to be the larger club. Committed bettors are assumed to be over-optimistic with regard to their own team's chances. To maximize expected profit, the bookmaker exploits their naiveté by moving the odds such that backing Manchester United is (objectively) a less fair bet than backing their opponents. Hence, Kuypers claims, inefficiency in odds is to be expected in the market.

Four comments may be made on Kuypers' contribution. First, it was novel and welcome, in the context of the previous literature, to put price setting in a profit-maximizing framework. Second, it is not necessary to assume a failure of rationality by Manchester United supporters; the model would be unaffected by assuming committed fans to have objectively correct expectations about a game but to derive additional utility from bets on their particular team because such bets are an expression of loyalty. Third, the model is highly restrictive in that the number of bettors on an event is predetermined and it is only the distribution of betting turnover between outcomes that is affected by relative movements in odds; according to the model, committed bettors would switch to betting on their opponent if the odds varied sufficiently. Fourth, Kuypers, in an article which reports many statistical tests (discussed above), presents no empirical testing whatsoever of this particular hypothesis.

Forrest and Simmons (2001) raise the possibility that fan sentiment will have the opposite effect from that in Kuypers' model. Suppose fan bettors would only ever bet in favor of their particular team but that whether or not they would bet depends on the odds offered. Essentially, they are purchasing a consumer good (akin to club merchandise) for which the price is the expected loss from the bet. If the demand curve is downward sloping, the particular bet on team A winning will sell more as it becomes cheaper. There is a separate product, a bet on "their" team, sold to fans of opposing club B. As a practical matter, the prices set in the two markets are constrained to equal a constant (given by the bookmaker overround).¹⁵ If team A has more fans than team B, and if demand is elastic at efficient odds, the bookmaker will maximize profit by offering a cheaper product in the larger market, that is, superior returns to those backing the bigger club. Whether

¹⁴More studies of betting on individual sports such as running would be a useful addition to the literature. Those for which analysis has been reported, golf and tennis, appear to be characterized by similar bias as in horse and dog racing (Shmanske, 2005; Forrest and McHale, 2007).

¹⁵All authors note the near constancy of the overround across matches but none accounts for why this constant appears to have been binding over many years. One possibility is that it is an expression of informal collusive understanding between bookmakers.

this happens in practice is testable as relative levels of support for different clubs can be readily measured, for example, by typical home-game attendances.

American evidence for the influence of sentiment in a sports betting market has been provided by Avery and Chevalier (1999) and Strumpf (2003). Avery and Chevalier found inferior returns on bets in favor of glamorous teams in the National Football League, glamour being proxied by recent championship success and the division in which the team played. This suggests that Las Vegas sports books act as if they were running a pari-mutuel betting process: relatively large numbers of bettors wish to bet on high profile teams and the casino-bookmakers balance their books by offering less fair bets in respect of success by those teams (i.e., point spreads are moved to be more challenging to the high profile teams). Avery and Chevalier recognize that this may not be (expected) profit-maximizing but explore regulatory reasons why sports books *may* feel the need to hold a fully rounded book on each individual match.¹⁶ However, there is no evidence that this is how Las Vegas in fact behaves and the apparent influence of club popularity on odds is more plausibly interpreted as monopolistic price discrimination. Of course there are several sports books in Las Vegas but all save one contract their odds-setting to the same agency.

Strumpf's innovative (2003) paper examines the behavior of the large illegal book-making sector in the U.S. by analyzing records seized by the police in the New York area. The industry is characterized by local monopolies and Strumpf provides convincing evidence that unfavorable odds are offered on local teams and to particular clients with a record of regularly backing the same club. As in Avery and Chevalier, this can be interpreted as evidence of price discrimination that works to the disadvantage of those wishing to place bets on popular outcomes. The result is consistent with the theoretical contribution by Levitt (2004) who demonstrates that a bookmaker maximizing expected client losses may offer inefficient odds to take account of bettor preferences (in his illustration, a preference for betting on favorites rather than underdogs leads to less fair terms being offered for favorite bets).

As noted, Forrest and Simmons thought the opposite result possible in soccer where the market is more competitive than in the contexts examined by Avery and Chevalier and Strumpf. They estimate the repeat trials model specified as follows:

$$\text{prob}(\text{win}_i) = a + b \text{bookprob}_i + c \text{home}_i + d (\text{bookprob}_i)(\text{home}_i) + e \text{diffattend} \quad (8)$$

Here, *diffattend* is the mean attendance, in thousands, at team *i* home games in the preceding season *minus* the corresponding statistic for the opposing team. The interaction term takes note of the possibility that the degree of longshot bias may differ as between home and away teams. Efficiency requires $b = 1$ and all other coefficient estimates to be zero.

While they found no evidence for other types of inefficiency across the period, they found *diffattend* to be positive and significant in the whole period estimation and in

¹⁶For example, there may be concern from the viewpoint of avoiding suspicion of corruption, that a betting firm should have no particular interest in the outcome of a sporting contest.

three of the four seasons considered individually. So there was a clear tendency for bettors to be offered superior value when wagering on whichever was the bigger club in a match.¹⁷

Forrest and Simmons (2008) returned to the issue of sentiment bias to evaluate whether it was present more recently and in an internet betting market. They chose to examine betting on matches in two European leagues, those of Spain and Scotland. The former enjoys a much higher international profile than the latter but the two share the characteristic that the competition is contested by clubs with very different levels of support (i.e., the variance of *diffattend* is high). Working with data from four seasons, they found that, as before, there was significant sentiment bias (in each league) with less unfair odds tending to be quoted for more popular teams. Whether awareness of this bias could be converted to a profitable betting rule was problematic, however. Out of sample betting based on allowing for the big team effect on odds proved profitable; but this could not have been predicted from the in-sample results where the strategy would have yielded superior but still negative returns. Nevertheless, it is of interest that a degree of sentiment bias remains in a globalized betting market. International bettors may of course be swayed by club popularity and reputation just as domestic bettors are. Indeed premier league clubs in Europe have worldwide support and this may be reflected in the demand to bet on certain teams.

9. THE FUTURE OF RESEARCH ON SOCCER BETTING

Notwithstanding the introduction of sentiment bias to the debate, the themes of the published literature on soccer betting have remained fairly stable since its inception nearly 20 years ago. Despite falls in transactions costs, findings also continue to be similar: odds are efficient enough that it is hard to formulate any strategy, whether through fundamental or technical analysis, that yields reliable positive returns.

The literature as it has been constituted has perhaps run its course. But new opportunities present themselves and will potentially generate findings that can offer new insights into the behavior of conventional financial markets. One promising opportunity is that provided by the growth of betting exchanges where in-play trading continues throughout a game. These markets generate rich datasets in a context that is more similar to other types of financial market (than bookmaking) in that it is possible for any participant to take a position on either side of a question by choosing to buy (back) or sell (lay) a particular proposition. Transaction costs are low relative to bookmaker markets. News arrives during a game, most dramatically in soccer in the form of goals. The news

¹⁷The variable *diffattend* is also used in Forrest and Simmons (2002b). They test for biases in soccer odds in a single English season (1997–1998) employing a three-equation linear probability model as proposed by Pope and Peel but with estimation by seemingly unrelated regression. They include the difference between home and away club mean attendance as an additional regressor and find it strongly significant in the home- and away-win equations. Their paper goes on to employ bias-corrected odds to proxy the degree of uncertainty of outcome in soccer games and examines the impact of this uncertainty of outcome on individual match attendances.

reaches all participants simultaneously through television relay, and it is then possible to assess the speed and accuracy with which the price responds. Moreover, archives for some exchanges identify which traders are making which offers, permitting an unusual opportunity for analysts to gain understanding of how prices move toward new equilibria. The principal practical difficulty for researchers is relating the time recorded for incidents in the sporting record for a match to the time recorded in the trading records.

Studies of this sort have already appeared for cricket (Easton and Ulyangco, 2007) and American football (Borghesi, 2007) markets. An early contribution for soccer is Gil and Levitt (2007). They set out a compelling agenda for research but their own evidence is inadequate because they have a small dataset and it pertains to a somewhat illiquid market. Their sample relates to matches played in the 2002 World Cup. The tournament featured 64 games but the analysis is based on only 50 since the remainder generated no trades at all on the particular exchange. Moreover, matches included in the sample stimulated trading volumes as low as \$2,000; and \$50,000 was reached in only three of them. This limits what can be learned from analysis of minute-by-minute trading since the market will typically have been very thin even following news such as a goal. Future researchers will, however, have access to archives from exchanges such as <http://www.betfair.com>, where the number of matches available for study is high as is the level of liquidity.

Gil and Levitt pursue highly interesting issues. One is how the market responds to a goal. There is no indication that the market anticipates the event (as it appears to do for events in cricket and American football where the nature of the sports is that the contest is divided into discrete plays rather than based on continuous action) but price of course moves strongly when it occurs. Gil and Levitt focus on their finding that the price of a contract on the team scoring does not typically change in a single step but drifts upward for 10–15 min. This should not be taken as evidence of inefficiency in itself, since the initial price movement may merely take into account that a significant minority of goals in soccer are matched by a quick counter goal by the other team. However, it would be evidence of inefficiency if their proposition, that it would be a profitable strategy to buy on a goal and sell a few minutes later, were to be supported in larger datasets from more liquid markets. Among their other interesting questions is whether market makers who engage in a high proportion of trades earn positive returns. Surprisingly, perhaps, they lost money in this particular market.

At the time of writing, several researchers are engaged in analysis of newly released archives recording exchange transactions during English Premier League games, for many of which the volume of transactions is high. The findings from this analysis will make up the next generation of studies of betting on soccer.

References

- Avery, C., and J. Chevalier. 1999. Identifying Investor Sentiment from Price Paths: The Case of Football Betting. *Journal of Business* 72, 493–521.
- Borghesi, R. 2007. Price Biases in Prediction Markets: NFL Contracts on Tradesports. *Journal of Prediction Markets*, 1, 233–253.

- Cain, M., D. Law, and D. Peel. 2000. The Favourite-Longshot Bias and Market Efficiency in Football Betting, *Scottish Journal of Political Economy* 47, 25–36.
- Crafts, N. 1985. Some Evidence of Insider Knowledge in Horse Racing Betting in Britain, *Economica* 52, 295–304.
- Crafts, N. 1994. Winning-Systems? Some Further Evidence on Insiders and Outsiders in British Horse Race Betting in Britain, in D. B. Hausch, V. S. Lo, and W. T. Ziemba (eds.), *Efficiency of Racetrack Betting Markets*, pp. 545–549. Academic Press, San Diego, CA.
- Deschamps, B., and O. Gergaud. 2007. Efficiency in Betting Markets: Evidence from English Football, *Journal of Prediction Markets* 1, 61–73.
- Dixon, M., and S. Coles. 1997. Modelling Association Football Scores and Inefficiencies in the UK Football Betting Market, *Journal of the Royal Statistical Society, Series C* 46, 265–280.
- Dixon, M., and P. Pope. 1996. Inefficiency and Bias in the UK Association Football Betting Market. Working Paper, Lancaster University.
- Dobson, S., and J. Goddard. 2001. *The Economics of Football*. Cambridge University Press, Cambridge, UK.
- Easton, S., and K. Ulyangco. 2007. An Examination of In-Play Sports Betting Markets Using One-Day Cricket Matches, *Journal of Prediction Markets* 1, 93–109.
- Forrest, D. 1999. The Rise and Fall of the British Football Pools, *Journal of Gambling Studies* 15, 161–176.
- Forrest, D., J. Goddard, and R. Simmons. 2005. Odds-Setters as Forecasters: The Case of English Football, *International Journal of Forecasting* 16, 317–331.
- Forrest, D., and I. McHale. 2007. Anyone for Tennis (Betting)? *European Journal of Finance* 13, 751–768.
- Forrest, D., and R. Simmons. 2000a. Making Up the Results: The Work of the Football Pools Panel 1963–1997, *Journal of the Royal Statistical Society: Series D (The Statistician)* 49, 253–260.
- Forrest, D., and R. Simmons. 2000b. Forecasting Sport: The Behaviour and Performance of Football Tipsters, *International Journal of Forecasting* 16, 317–331.
- Forrest, D., and R. Simmons. 2001. Globalisation and Efficiency in the Fixed-Odds Soccer Betting Market. Working Paper, University of Salford.
- Forrest, D., and R. Simmons. 2002. Outcome Uncertainty and Attendance Demand in Sport: The Case of English Soccer, *Journal of the Royal Statistical Society, Series D (The Statistician)* 51, 229–241.
- Forrest, D., and R. Simmons. 2008. Sentiment in the Betting Market on Spanish Football, *Applied Economics* 40, 119–126.
- Gandar, J., W. Dare, C. Brown, and R. Zuber. 1998. Informed Traders and Price Variations in the Betting Market for Professional Basketball Games, *Journal of Finance* 53, 385–401.
- Gil, R., and S. Levitt. 2007. Testing the Efficiency of Markets in the 2002 World Cup, *Journal of Prediction Markets*, 1, 255–270.
- Global Betting and Gaming Consultants. 2001. First Annual Review of the Global Betting and Gaming Market, 2001. Global Betting and Gaming Consultants, West Bromwich, UK.
- Goddard, J. 2005. Regression Models for Forecasting Goals and Match Results in Association Football, *International Journal of Forecasting* 21, 331–340.
- Goddard, J., and I. Asimakopoulous. 2004. Forecasting Football Results and the Efficiency of Fixed Odds Betting, *Journal of Forecasting* 23, 51–66.
- Golec, J., and M. Tamarkin. 1991. The Degree of Inefficiency in the Football Betting Market, *Journal of Financial Economics* 30, 311–323.
- Graham, I., and H. Stott. 2008. Predicting Bookmaker Odds and Efficiency for UK Football, *Applied Economics* 40, 99–109.
- Greene, W. 2000. *Econometric Analysis*, 4th ed. Prentice Hall, Upper Saddle River, NJ.
- Koning, R. 2000. Balance in Competition in Dutch Soccer, *Journal of the Royal Statistical Society, Series D (The Statistician)* 49, 419–431.
- Kossmeier, S., and S. Weinberger. 2008. Efficiency of Soccer Betting Odds—Evidence from a Pan-European Electronic Market, in D. B. Hausch and W. T. Ziemba (eds.), *Handbook of Sports and Lottery Markets*. North-Holland, Amsterdam, pp. 447–456.
- Kuypers, T. 2000. Information Efficiency: An Empirical Study of a Fixed Odds Betting Market, *Applied Economics* 32, 1353–1363.

- Levitt, S. 2004. Why Are Gambling Markets Organized so Differently from Financial Markets? *Economic Journal* 114, 223–246.
- Little, I. 1957. *A Critique of Welfare Economics*, 2nd ed. Clarendon Press, Oxford, UK.
- Maier, M. 1982. Modelling Association Football Scores, *Statistica Neerlandica* 36, 109–118.
- Miers, D. 2004. *Regulating Commercial Gambling: Past, Present and Future*. Oxford University Press, Oxford, UK.
- Munting, R. 1996. *An Economic and Social History of Gambling in Britain and the USA*. Manchester University Press, Manchester, UK.
- National Centre for Social Research. 2007. *British Gambling Prevalence Survey 2007*. The Stationery Office, London.
- Pope, P., and D. Peel. 1989. Information, Prices and Efficiency in a Fixed-Odds Betting Market, *Economica* 56, 322–341.
- Rue, H., and O. Salvesen. 2000. Prediction and Retrospective Analysis of Soccer Matches in a League, *Journal of the Royal Statistical Society, Series D (The Statistician)* 49, 399–418.
- Sauer, R. 1998. The Economics of Wagering Markets, *Journal of Economic Literature* 36, 2025–2064.
- Sharpe, G. 1997. *Gambling on Goals: A Century of Football Betting*. Mainstream Publishing Company, Edinburgh, UK.
- Shin, D. 1991. Optimal Betting Odds Against Insider Traders, *Economic Journal* 101, 1179–1185.
- Shmanske, S. 2005. Odds-Setting Efficiency in Gambling Markets: Evidence from the PGA Tour, *Journal of Economics and Finance* 29, 391–402.
- Snedecor, G., and W. Cochran. 1967. *Statistical Methods*, 6th ed. Iowa State University Press, Ames, IA.
- Song, C., B. Boulier, and H. Stekler. 2007. Comparative Accuracy of Judgmental and Model Forecasts of American Football Games, *International Journal of Forecasting* 23, 405–413.
- Strumpf, K. 2003. Illegal Sports Bookmakers. Working Paper, University of North Carolina at Chapel Hill.
- Terrell, D., and A. Farmer. 1996. Optimal Betting and Efficiency in Pari Mutuel Betting Markets with Information Costs, *Economic Journal* 107, 150–158.
- Vaughan Williams, L. 1999. The Cost of Insider Trading: Evidence from Defined Markets, *Economic Issues* 4, 19–30.
- Vergin, R., and J. Sosik. 1999. No Place Like Home: An Examination of the Home Field Advantage in Gambling Strategies on NFL Football, *Journal of Economics and Business* 57, 21–31.
- Woodland, L., and B. Woodland. 1994. Market Efficiency and the Favourite-Longshot Bias: The Baseball Betting Market, *Journal of Finance* 49, 269–280.
- Woodland, L., and B. Woodland. 2001. Market Efficiency and Profitable Wagering in the National Hockey League: Can Bettors Score on Longshots? *Southern Economic Journal*, 67, 983–995.
- Woodland, L., and B. Woodland. 2003. The Reverse Favourite-Longshot Bias and Market Efficiency in Major League Baseball: An Update, *Bulletin of Economic Research* 55, 113–123.

CHAPTER 21

Efficiency of Soccer Betting Odds—Evidence from a Pan-European Electronic Market

Stephan Kossmeier

Institute for Advanced Studies, Vienna, Austria.

Simon Weinberger

Institute for Advanced Studies, Vienna, Austria.

1. Introduction	448
2. The Soccer Betting Market	449
3. Data Description	450
4. Efficiency Tests	451
4.1. Statistical Tests	451
4.2. Economic Tests	454
5. Conclusion	455
References	456

Acknowledgments

We thank David Forrest for highly instructive comments. We are also grateful to Ursula Hauser-Rethaller, Michael Jeckle, and Gabriel Lee for useful comments and discussions as well as research assistants at IHS for their data collection efforts.

Abstract

We analyze the efficiency of soccer betting odds quoted by several Austrian internet bookmakers. Due to tax and legal issues, Austria has become a major market place for online betting. Quoted odds cover all major European soccer leagues as well as international games. We use regression analysis and simulations of betting strategies to examine weak-form market efficiency. Similar to other betting markets, we find evidence for a home bias as well as a favorite-longshot bias.

JEL Classification: L83

1. INTRODUCTION

A bet can be considered a derivative security with a payoff depending on the uncertain occurrence of a future event. It is, for instance, similar to an option on a stock in the sense that both contracts have a fixed point of termination. However, while the distribution of the stock price at maturity is likely to be influenced by expectations and beliefs of market participants, the probability law governing the outcome of a soccer match can be safely assumed to be independent of bettors' beliefs. This makes betting markets a convenient model economy with reduced complexity to study market participants' information processing abilities and risk preferences. Consequently, betting markets have been widely studied in the literature on finance and economics.

The contribution of this chapter is twofold: First, to our best knowledge it is the first empirical analysis of an online-betting market. Analysis of traditional brick-and-mortar bookmakers has revealed two major biases in odds: the favorite-longshot bias, which refers to the phenomenon that favorites are overpriced relative to underdogs, and the home bias, that is, the home advantage is not sufficiently reflected in quoted odds. One could expect that these biases were reduced or even nonexistent in an electronic betting market, because of its electronic nature, in that odds are available to a large and possibly well informed public. Moreover, comparing odds of different bookmakers is virtually costless via the internet. The second contribution of the chapter is that it represents the first study of continental European soccer betting markets. In that respect, it helps to make the empirical evidence on betting market efficiency more complete.

Most work on the subject of efficiency of betting markets has been done on horse racing and to a lesser extent on major sports in the U.S., such as baseball, football, or basketball. However, in Europe, soccer clearly exceeds the previously mentioned sports in popularity. Consequently, the market for betting on soccer games provides an interesting platform to analyze market efficiency. Nevertheless, examinations of soccer betting markets are scarce. All of the previous studies work with odds from British bookmakers quoted for matches from the English and Scottish leagues. Pope and Peel (1989) analyze odds from the 1981–1982 season. They cannot reject weak-form efficiency and also do not find profitable gambling strategies. Dixon and Pope (1996) use odds from 1992 through 1996. Their findings indicate that odds-implied-probabilities

for home wins are too low (home bias) and they also find a reverse favorite-longshot bias (probabilities for the win of the underdog are underestimated). In contrast to this, Cain et al. (2000), using data from the 1991–1992 season, identify a favorite-longshot bias. In a recent paper, Forrest and Simmons (2001) examine the odds from 1997–1998 through 2000–2001. Odds in the first two seasons reveal a reverse favorite-longshot bias and a home bias as well as an additional bias, which relates to differences in fan support. However, the authors find no evidence for inefficiency in the last season of the sample period, which they attribute to increased competition by offshore bookmakers.

Our sample is made of odds quoted by 12 Austrian bookmakers and outcomes of roughly 1,800 games played in major European leagues and international cups. Our findings suggest that weak-form efficiency of the market in question can be rejected. We provide evidence that biases that have been reported for the traditional betting markets for various sports around the world are also prevalent in this electronic market. First, quoted odds show the typical favorite-longshot bias and home bias for all bookmakers. Accordingly, odds-implied probabilities of favorite and home wins are significantly biased downwards. Second, while one single bookmaker does not allow profitable betting strategies this changes when we look at all of them simultaneously.

The chapter proceeds in the following way. Section 2 introduces the soccer betting market. Section 3 describes the sample data and the conversion of odds into corresponding probabilities. The subsequent two sections test the efficiency of the soccer betting market, first statistically and then economically. Finally, Section 6 concludes.

2. THE SOCCER BETTING MARKET

Apart from numerous bricks-and-mortar bookmakers, recent years are characterized by the emergence of new firms entering the market for sports betting, in particular via the internet. These young firms often do not provide in-house betting anymore, but are completely focused on the electronic market. The market is very transparent as everybody is able to compare odds over various bookmakers by browsing the internet. Furthermore, firms have been founded that provide clients with an odds comparison service.

Due to legal and tax issues, Austria has become a major market in the business of online betting, attracting clients from all over the world. Contrary to many other jurisdictions, there is no betting tax, and the bettor can place bets on single matches.¹ Austrian bookmakers offer a wide range of games, including all major national leagues and international games. Accordingly, one would assume that this market is highly competitive and also not necessarily restricted to single countries, as anyone, without nationality restriction, can open up an account with the bookmaker and participate in the market. We have identified 12 bookmakers based in Austria providing odds for soccer games of the major European leagues. The business of online sports betting had

¹ In the UK, for example, bettors are restricted to combining bets on different matches or half-time and full-time results. Presumably, this is about to change due to increased competition from offshore bookmakers.

a turnover of 110,000,000 Euro in the year 2000 in Austria and has taken a share of approximately one-third of the total betting business. This share, as well as the industry as a whole, are expected to further increase significantly over the next years.

Sports betting is organized either as a pari-mutuel (e.g., horse races) or a fixed-odds market. With the latter, the bookmaker fixes odds a few days prior to the match and therefore faces an uncertain profit margin. European soccer betting is organized in a special form of fixed-odds market. Once bookmakers have posted the odds, they stick with them until the end of the betting period.² Therefore bookmakers bear the additional risk of an unbalanced book. Furthermore, bettors may utilize new information that arises during the betting period. Consequently, it is all the more important for bookmakers to accurately assess match outcome probabilities and anticipate bettors' behavior.

3. DATA DESCRIPTION

We have collected odds for games of six European national leagues (Austria, England, France, Germany, Italy, and Spain) as well as of international games (World Cup qualifying, Champions League, and UEFA-Cup). Overall our database comprises odds on the outcomes win of home team, win of away team and tie for about 1,800 games, played in the season 2000–2001. The sample was chosen according to the range supplied by most of the bookmakers. While some additional leagues are sometimes also available, they often only allow bets combining at least three games. In order to facilitate further analysis these games were not considered.

In a first step, the odds are used to compute implied probabilities of the three possible outcomes and the take-out rate in the following way. Assuming an equal margin for the three outcomes, we arrive at a system of four equations with four unknowns.

$$P_i = \frac{(1-h)}{Q_i} \quad (1)$$

$$\sum_i P_i = 1 \quad (2)$$

In the equations above, h denotes the take-out rate and $Q_i(P_i)$ the odds (implied probability) of outcome i (win home, win away, or tie) for the respective game. Therefore we hold the take-out rate constant over the three outcomes, but not over the series of games.

The average game offers an odds system of 2.0/3.2/3.8 for the outcomes win home team, tie, and win away team, respectively. The average take-out rate over all games and all bookmakers amounts to about 14.9%, but a large variation among the firms can be observed, ranging from 10.9% to 18.0% (see column two in Table 1). We can observe a clear tendency to underestimate the probability of a home win, because while the average implied probability amounts to 0.463, a win for the home team occurred in 50.6

²Betting rules give the bookmaker the option to change quoted odds afterwards, but this is hardly ever done.

TABLE 1 Descriptive Statistics

Bookmaker	h	P_1	P_X	P_2
Bookmaker A	0.145	0.458 (0.117)	0.269 (0.028)	0.273 (0.098)
Bookmaker B	0.109	0.470 (0.131)	0.267 (0.032)	0.263 (0.112)
Bookmaker C	0.156	0.462 (0.122)	0.273 (0.029)	0.266 (0.105)
Bookmaker D	0.148	0.460 (0.121)	0.272 (0.030)	0.268 (0.102)
Bookmaker E	0.144	0.476 (0.130)	0.261 (0.029)	0.263 (0.113)
Bookmaker F	0.123	0.467 (0.129)	0.271 (0.032)	0.262 (0.110)
Bookmaker G	0.148	0.459 (0.119)	0.274 (0.032)	0.266 (0.097)
Bookmaker H	0.160	0.462 (0.114)	0.270 (0.026)	0.268 (0.098)
Bookmaker I	0.180	0.450 (0.108)	0.276 (0.022)	0.274 (0.093)
Bookmaker J	0.150	0.469 (0.124)	0.268 (0.029)	0.263 (0.106)
Bookmaker K	0.162	0.463 (0.123)	0.272 (0.032)	0.265 (0.104)
Bookmaker L	0.151	0.460 (0.117)	0.274 (0.029)	0.265 (0.099)
Ex-post		0.506	0.260	0.233

NOTE: The numbers represent the average odds-implied probability of a home win, a tie, and an away win. In parentheses we quote the corresponding standard deviation.

out of 100 games. On the other hand, probabilities of a win for the away team and a tie are overestimated (0.267 vs. 23.3 away wins and 0.270 vs. 26 draws). This finding provides first evidence for the stylized fact of home bias, observed in other betting markets as well. The standard deviation of the implied probabilities varies over the possible outcomes and is remarkably lower for tie probabilities. While the standard deviation for tie probabilities is about 0.03, the variability of home and away probabilities is roughly four and three times larger, respectively. Subsuming, all considered bookmakers exhibit similar patterns concerning probability estimates, but differ significantly in the take-out rate.

4. EFFICIENCY TESTS

4.1. Statistical Tests

The classical rational expectations test involves a regression of the forecasting variable on the forecasted variable. If the intercept of this linear regression is significantly different from zero, then the prediction is biased, that is, the prediction is on average too high or too low. If the slope significantly differs from one then the variance of the prediction is too high or too low relative to the variance of the predicted variable. In the context of soccer betting, the forecasting variable is the bookmaker's odds implied probability and the forecasted variable is a binary variable indicating if team one has actually won the game. Due to the fact that the dependent variable is

discrete with only two possible outcomes, the linear model seems to be inappropriate. Instead, binary choice models seem to be more suitable. For instance, Lo (1994) applies such a model to test efficiency. In these models, the fitted values are transformed by a nonlinear function to ensure that they are in the interval between zero and one (see e.g., Greene, 2000). The usual choice for this function is the normal or the logistic probability function, corresponding to the probit or logit model. However, the nonlinear functional form of the probit or logit model causes another problem. Our hypothesis is formulated as a linear relationship between forecast and forecasted variable. Therefore, the relevant null-hypothesis cannot be formulated in terms of a nonlinear model. Probit/logit parameters are usually converted to marginal effects and test statistics are computed based on the average marginal effects. It can be shown in simulation experiments that the average marginal effect is biased downward even for an efficient forecast. Intuitively, the reason for this is that the tails of the normal and logistic distribution function flatten out at very low and very high probabilities, which is not necessarily supported by the data, but enforced by the functional form of the model.

Given the drawbacks of alternative specifications, we choose to use a linear model. We believe that the mis-specification does not do much harm, given that bookmakers' odds are capped at some minimum and maximum and therefore we do not have data points at the extreme ends of the unit interval. This view is supported by the observation that fitted values for our models never fall outside of the unit interval.

Another problem associated with odds for team sport events as pointed out by Golec and Tamarkin (1991) is that in team sports the home team is more often the favorite than the away team (due to the home advantage) and therefore both effects are correlated. At the same time, odds are usually quoted with the home team being team 1. To control for this effect, we apply the test strategy of Forrest and Simmons (2001). First, we randomly select the focus team for each game. Then in this random sample we estimate the following equation:

$$I_{Wi} = \beta_0 + \beta_1 \times PWIN_i + \beta_2 \times I_{Hi} \quad (3)$$

In Equation (3), I_{Wi} denotes a dummy variable indicating the win of the focus team, I_{Hi} refers to a dummy variable indicating whether the selected team is actually the home team, and $PWIN_i$ refers to the probability of the home team winning as implied by the bookmaker's odds in game i . We account for heteroscedasticity by using robust standard errors (White, 1980). We test the parameter restrictions $\beta_1 = 1$ and $\beta_2 = 0$ using t -tests and compute an F -test statistic for the joint hypothesis of $\beta_0 = 0$, $\beta_1 = 1$ and $\beta_2 = 0$.

Now, the question remains if the results of the hypothesis tests are artifacts of the particular random sample. To test for that, we perform the random selection 80 times and record the proportion of significant relationships at the 5% significance level. Under the null hypothesis that the true relationship is marginally significant at the 5% level, we should find not significantly more than 5% significant relationships. The observed

TABLE 2 Regression Results

Bookmaker	$\hat{\theta}\beta_1$	$\hat{\theta}\beta_2$	$\hat{\theta}F$ -Statistic
Bookmaker A	1.138**	0.055**	6.15***
Bookmaker B	1.161**	0.052**	6.56***
Bookmaker C	1.129*	0.043**	4.44**
Bookmaker D	1.160**	0.045**	5.82***
Bookmaker E	1.157**	0.050**	6.05***
Bookmaker F	1.144**	0.048**	5.59***
Bookmaker G	1.195***	0.043**	6.53***
Bookmaker H	1.149**	0.050**	5.95***
Bookmaker I	1.170**	0.044*	5.81***
Bookmaker J	1.151**	0.035*	4.05**
Bookmaker K	1.150**	0.054***	6.45***
Bookmaker L	1.142**	0.046**	5.23***

NOTE: ***/**/*, significant on 1%/5%/10% level.

proportion of significant relationships, θ , is normally distributed and therefore we compute the following test statistic.

$$t = \frac{|\theta - 0.05| - \frac{1}{2n}}{\sqrt{\frac{\theta \times (1-\theta)}{n}}} \quad (4)$$

Table 2 shows the result for all bookmakers vis-à-vis Equation (3) and corresponding statistical tests. The numbers denote the average coefficient over the 80 trials and the significance refers to the test statistic corresponding to Equation (4). We find strong evidence of a home bias, because the average coefficient of the home indicator variable is significantly greater than zero for all but one bookmaker on a 5% significance level. Furthermore, we observe a favorite-longshot bias, since the average β_1 is significantly greater than one for all bookmakers on a 5% significance level. As expected from these results, the joint hypothesis of overall market efficiency can also be rejected for all bookmakers at a 5% significance level.

Our findings are not in line with the results of Forrest and Simmons (2001), who observe a trend toward efficiency from 1996–1997 to 2000–2001. They attribute this to the rise of offshore bookmakers, which offer odds mainly over the internet. On the contrary, we find that major biases reported in the betting literature are present and highly significant also in this market. However, the question remains whether these inefficiencies are evenly distributed over the various leagues in our sample. Preliminary analysis of subsamples by leagues suggests that odds for those leagues that are likely to receive more attention from international betting clientele (e.g., England) are less biased

than leagues that are more of national rather than global importance (e.g., Austria). Furthermore, competitive balance might be of importance as in some leagues strength is distributed more evenly over the teams. Unfortunately, the subsamples corresponding to national leagues are still too small to perform statistical tests on the disaggregated data. Therefore, this question is postponed to future research when more data becomes available.

4.2. Economic Tests

A different approach to test for market efficiency is to examine various betting strategies. First, we are interested in differences in the performance of various strategies in general, and second we search for technical rules yielding positive returns. If the latter occur in the dataset, this would be a strong indication of inefficiency.

In a first step, we test, for each bookmaker separately, whether some predefined strategy would have yielded abnormal returns. Bettors with absolutely no relevant information will select randomly, which can be captured by a strategy involving all possible bets. This would have led to a loss between 0.145 and 0.220, depending on the bookmaker (see column two in Table 3). Simply playing home wins, ties, or away wins, every single game yielded very different performances. While home teams offered an above average rate of return, odds on away teams bore high losses (see columns three to five). This result is not surprising, since the previous section showed that a home bias is prevalent in odds-implied probabilities. Furthermore, we classify favorites and underdogs as teams with an average odds-implied probability of greater/less than 0.5/0.2 over

TABLE 3 Betting Strategies I

Bookmaker	All	H	T	A	F	U
Bookmaker A	-0.182	-0.063	-0.177	-0.305	-0.027	-0.434
Bookmaker B	-0.145	-0.044	-0.150	-0.241	-0.023	-0.351
Bookmaker C	-0.181	-0.085	-0.196	-0.263	-0.066	-0.370
Bookmaker D	-0.183	-0.069	-0.189	-0.289	-0.033	-0.431
Bookmaker E	-0.168	-0.082	-0.159	-0.264	-0.076	-0.316
Bookmaker F	-0.152	-0.050	-0.164	-0.243	-0.043	-0.355
Bookmaker G	-0.187	-0.066	-0.206	-0.289	-0.019	-0.458
Bookmaker H	-0.198	-0.088	-0.199	-0.309	-0.052	-0.445
Bookmaker I	-0.220	-0.085	-0.244	-0.331	-0.045	-0.473
Bookmaker J	-0.176	-0.092	-0.181	-0.253	-0.057	-0.376
Bookmaker K	-0.196	-0.081	-0.204	-0.303	-0.061	-0.419
Bookmaker L	-0.185	-0.076	-0.196	-0.283	-0.047	-0.403

NOTE: The numbers refer to the rate of return, which could have been achieved by playing all outcomes every game (All), only home wins (H), only ties (T), only away wins (A), only favorites (F), and only underdogs (U).

TABLE 4 Strategies Over All Bookmakers

Strategy	Games	Rate of return
Home win	1,790	−0.0004
Tie	1,790	−0.1146
Away win	1,790	−0.1740
All	5,370	−0.0963
Favorite	661	0.0080
Underdog	514	−0.2661
Favorite-home	634	0.0051
Favorite-away	27	0.0763
Underdog-home	15	−0.0667
Underdog-away	499	−0.2720

NOTE: The numbers refer to the rate of return, which could have been achieved by playing all outcomes every game (All), only home wins, only ties, only away wins, only favorites, and only underdogs.

the bookmakers, respectively. We find huge differences in the performance of these two betting strategies, which again confirms the favorite-longshot bias in our dataset (see the last two columns in Table 3). In any case, when restricting ourselves to a single bookmaker, we were not able to observe strategies yielding a positive rate of return.

In a second step, we simulate the strategy of a bettor maximizing expected return, who always searches for the highest odds on the respective game (see Table 4). Hence, this approach exploits interfirm differences in combination with the simple betting strategies presented above. Accordingly, we could have achieved a positive rate of return, betting on favorites (0.80%) and an even position betting on home teams (−0.04%). This once again highlights the home bias and favorite-longshot bias implicit in the bookmakers’ odds. The positive rate of return corresponding to betting on favorites originates from games where the home team is the favorite. Interestingly, favorites playing away seem to offer an even higher potential for successful strategies, but this constellation does not appear very often in our data. On the other hand, in most cases, the underdog coincides with the away team and therefore these teams offer the worst rates of return.

5. CONCLUSION

We have empirically investigated the efficiency of the online soccer betting market in Austria, one of the major betting markets in Europe. Our results indicate the existence of a statistically significant home bias as well as a favorite-longshot bias. Literature on traditional betting markets offers various explanations for the existence of these biases

(see Hausch and Ziemba, 1995; Sauer 1998; and Vaughan Williams, 1999). Still, it is surprising that they are also present in this highly transparent and competitive market. While one single bookmaker does not allow profitable betting strategies, this changes when we look at all of them simultaneously. Out-of-sample data will show whether these results are stable over time. Nevertheless, with respect to data on the season 2000–2001, we conclude that the Austrian online soccer betting market is not weak-form efficient.

References

- Cain, M., D. Law, and D. Peel. 2000. The Favourite-Longshot Bias and Market Efficiency in U.K. Football Betting, *Scottish Journal of Political Economy* 47, 25–36.
- Dixon, M., and P. Pope. 1996. *Inefficiency and Bias in the UK Association Football Betting Market*. University of Lancaster.
- Forrest, D., and R. Simmons. 2001. *Globalisation and Efficiency in the Fixed-Odds Soccer Betting Market*. University of Salford.
- Golec, J., and M. Tamarkin. 1991. The Degree of Inefficiency in the Football Betting Market, *Journal of Financial Economics* 30, 311–323.
- Greene, W. H. 2000. *Econometric Analysis*, 4th ed. Prentice Hall, Englewood Cliffs, NJ.
- Hausch, D. B., and W. T. Ziemba. 1995. Efficiency of Sports and Lottery Betting Markets, in R. Jarrow et al. (eds.), *Handbooks in OR & MS*, 9, 545–580. Elsevier Science B.V., Amsterdam.
- Lo, V. S. S. Y. 1994. Applications of Logit Models in Racetrack Data, in D. B. Hausch, V. S. Y. Lo, and W. T. Ziemba, *The Efficiency of Racetrack Betting Markets*. Academic Press, San Diego, CA, pp. 307–314.
- Pope, P., and D. Peel. 1989. Information, Prices and Efficiency in a Fixed-Odds Betting Market, *Economica* 56, 322–341.
- Sauer, R. 1998. The Economics of Wagering Markets, *Journal of Economic Literature* 36, 2021–2064.
- Vaughan Williams, L. 1999. Information Efficiency in Betting Markets: A Survey, *Bulletin of Economic Research* 51, 1–30.
- White, H. 1980. A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroscedasticity, *Econometrica* 48(4), 817–838.

DOWNLOAD FULL VERSION HERE: <https://bettingbooks.org/12>

PART VIII: Lotteries

Download More Sports Betting Books at: www.bettingbooks.org

DOWNLOAD FULL VERSION HERE: <https://bettingbooks.org/12>

This page intentionally left blank

Download More Sports Betting Books at: www.bettingbooks.org

CHAPTER 22

How to Design a Lottery

Ian Walker

Professor of Economics, University of Warwick, Coventry, UK.

1. Introduction	460
2. The Odds of Winning a (Pari-Mutuel) Lottery	464
3. The Expected Value Calculation	465
4. Higher Moments of the Prize Distribution	467
5. Econometric Methodology, Data, and Estimates	469
6. Game Design Simulations	474
7. Conclusion	476
<i>References</i>	477
Appendix: The Expected Value Formula	479

Acknowledgments

This chapter builds on earlier research that was funded by ESRC under research grant R000236821. This work was supported, in part, by a research studentship jointly funded by ESRC and the National Lottery Commissioners. However, the views expressed here are those of the author alone.

Abstract

This chapter provides an outline to the statistical, economic, and practical considerations relevant to designing lottery games. The focus is on lotto but some of the issues apply more generally.

We illustrate the analysis using an estimated statistical model of the determinants of lotto sales, that builds on recent research on horsetrack betting, to simulate the way in which changing the design of lotto games might affect their sales using data for the UK. The research complements recent lottery research that models the proximate determinants of sales, such as jackpot size, by exploiting changes induced by rollovers as a source of natural experiments. The approach pursued here assumes that what matters for sales is the shape of the prize distribution and this shape is summarized by the first three moments of the prize distribution (mean, variance, and skewness). The methodology builds on the characteristics approach to demand analysis that was pioneered by Lancaster. The approach has the advantage that it can be used to simulate the effects of hypothetical changes in the shape of the prize distribution that come about because of game design variation. We estimate the model, taking account of the endogeneity of the moments, and use the estimates to simulate how sales would evolve if game design were changed.

1. INTRODUCTION

The traditional economist's toolbox is not well equipped to understand why risk-averse individuals would participate in gambles that are actuarially unfair bets. Thus, understanding gambling has been something of a problem for economists. However, gambling is a pervasive feature of most economies, and a number of simple departures from traditional economic thinking have been suggested in the literature to attempt to reconcile gambling with the idea that individuals are risk-averse. One recent phenomenon in betting markets has been the worldwide growth of sales in lottery products over the last 40 years. The scale and scope of lotteries has expanded considerably over recent decades and there has been considerable controversy over their increasing use by governments as sources of revenue. In most economies, lottery operators are charged with raising revenue that can then be used either to support general government expenditure or is earmarked for particular purposes. Indeed, operators are generally expected, within bounds, to maximize the revenue that they raise.

We will take it for granted that such gambles are not (or, rather, are not just) financial assets that have no intrinsic value in themselves. Rather, we view a lottery ticket as a *product* within which is embedded a number of characteristics, some of which consumers like and some they don't, but that, at least for some individuals, the attractive characteristics outweigh the unattractive ones. Some of these characteristics are fixed, and while these might contribute to the average level of sales, because they are fixed, they cannot explain variation in sales over time. For example, the number of lottery sales outlets is usually relatively static over time and so cannot explain sales variation over

time. However, we would like to know how sensitive sales are to the design parameters of games—in particular, those parameters that are constant because they are embedded in the design of the game. For example, we are interested in the take-out rate (the proportion of revenue staked that is *not* returned as prizes through the various prize pools). Usually, a certain share, s_n , of the total prize money is reserved for the jackpot prize [claimed by the winner(s) who have “matched” *all* of the n numbered balls that are drawn from the N that are available] and the rest is allocated into separate prize *pools*, in predetermined shares s_{n-1} , s_{n-2} , and so on, to those who match $n - 1$, $n - 2$, and so on, balls. N and n determine the probabilities, p_n , p_{n-1} , and so on, of winning the n -ball, $n - 1$ -ball, and so on, prize pools.

Occasionally, games will change in design and, in principle, one might then investigate the effect of such a change on sales revenue. However, such changes in game design usually occur for reasons that are correlated with sales—for example, sales may be flagging and what was once a sensible design may now be deemed to be less suitable. So, to arrest the sales decline it may be sensible to change the design. Thus, it is not generally appropriate to use such events to show how some change in general game design, that was not correlated with previous sales, would affect sales. For example, policymakers may be interested in knowing how variations in the take-out rate will affect sales revenue of a given game. To cast light on such questions, we need to see the effects on sales of changes that are *not* chosen by the operator. Of course, there are *no* such changes in practice. However, a peculiar feature of many lottery games is that they are pari-mutuel bets where the number of winners of each prize pool in any particular draw is a random variable. This leads to a probability that there will be no winner in any draw (a rollover). Lottery games that have this feature are usually referred to as lotto games and are designed so that when a rollover occurs, the jackpot prize pool is transferred to the jackpot pool of the next draw. Rollovers change the prize distribution of the game. This chapter exploits the effects of rollover-induced changes in the shape of the distribution of prizes to make inferences about how design-induced changes in the distribution would affect sales. That is, even though we do not have any changes in game design in our data that could be informative, we do have rollovers that, in principle, are informative about design effects.

In this chapter, we explore how the sales of lottery tickets are affected by the *shape* of the prize distribution that is ultimately determined by the way in which the lottery game has been designed. By shape, we mean the distribution of the prize fund across the winners of different prize pools. The size of the overall fund is determined by the take-out rate (which is used to pay for the operator’s costs or is given to the government or its agent to be spent), and the level of sales. Most lotto games feature a jackpot prize pool that is reserved for winners of the hardest to win jackpot prize (usually matching all n of the balls drawn), another for those that fulfill the easier task of matching $n - 1$ balls, and so on. Each prize pool will consist of a given proportion of the overall prize fund. Thus, in lotto games, there is random variation in the number of winners of each prize pool from draw to draw. And, in particular, the number of jackpot winners varies so that there will sometimes be *no* jackpot winners in a given draw. The rules of lotto then usually dictate that the unclaimed jackpot pool gets transferred to the following

draw and added to that draw's jackpot pool. Casual empiricism shows that rollovers are associated with higher sales in the subsequent draw because they increase its jackpot prize pool.

Thus, the distribution of possible winnings is determined by: game design parameters, such as the take-out rate and the given proportions of the overall funding going to each prize pool; random fluctuation in the number of winners in each prize pool; and the size of any rollover amount from the previous draw. The design parameters are known to players and are fixed; the rollover amount varies from draw to draw but is known one draw in advance; the number of winners of each prize pool depends on the number of players and on the numbers that are drawn in that draw (if popular numbers¹ are drawn there will be many winners), which are not known in advance.

In practice, there are many complications that have to be faced, but the essence of our analysis is based on exploiting the variation in the shape of the prize distribution to explain sales variation over time. In particular, we examine how variation in the prize distribution affects sales to test the hypothesis that players are motivated, in part, by the *skewness* of the prize distribution. Lotto prize distributions are highly *left skewed* because almost all players lose their stakes, with a small proportion winning small prizes, and a very small proportion winning very large prizes. But when a rollover occurs, the size of the largest prize rises and so the degree of left skewness falls because the largest prize suddenly got larger.

In addition to examining the proposition that skewness motivates gambling,² our analysis has a practical purpose: we use this estimated model to try to make counterfactual inferences about how changes in the design of the lottery game that generates our data might affect the level of sales. While the dataset we use does not contain *any* changes in the design parameters of the game, we exploit the fact that there is a deterministic relationship between the mean, variance, and skewness of the prize distribution (the prize distribution's first three *moments*) and there is a deterministic relationship between the design parameters and the level of the rollover from the previous draw. Since we know how the moments depend on both of the game design parameters, and on the size of the jackpot pool, and since we can estimate how sales depend on these moments, we can combine these to infer how sales depend on the design parameters, even though there has been no variation in those parameters across the sample.

This chapter builds on earlier research in Walker and Young (2001). The major empirical difficulty with that earlier work is that it failed to resolve satisfactorily the fact that, although sales will depend on the moments of the prize distribution, the moments of the prize distribution themselves depend on sales. Thus the relationship between sales and the moments of the prize distribution is a simultaneous one—so the major empirical difficulty we face is to estimate the extent to which variations

¹Ziembra (1986) and MacLean et al. (1992) provide extensive evidence on such *conscious selection*—that players' choice of numbers is not random so that different numbers, while each having the same probability of being drawn, will generate different numbers of winners, and so can be exploited to improve one's expected winnings. See also Thaler and Ziembra (1988), Hartley et al. (2000), and Ziembra (this volume).

²In horse race betting, the idea that punters favor longshots unduly is a related phenomenon.

in the moments *cause* variations in sales. The usual way of resolving simultaneity in the relationship between two variables is to exploit factors that affect one relationship but not the other. A possible candidate factor is the size of the rollover—this affects sales in the next draw *via the effect on the moments of the prize distribution* and Walker and Young (2001) use rollovers as an instrumental variable. However, the current jackpot size includes any rollover from the previous draw, which will be a deterministic proportion of the sales in the previous draw. Unfortunately, sales in one draw may be correlated with sales in the next, even in the absence of rollovers. For example, this correlation might arise because a temporary factor that boosts sales in one draw may bring a few new individuals into the game who then gain (or regain) the habit of playing and so subsequent sales are affected. This *serial correlation* in sales implies that we cannot use sales in a previous draw. Unfortunately, this rules out exploiting the rollover size, because this is a given fraction of the sales in the previous draw.

Moreover, the moments are not known until the draw is closed and sales can be tallied. Thus, it seems plausible that individuals form *expectations* of the moments based on sales in the current draw and that these expectations are driven, in part, by any observed rollover from the previous draw. A large rollover will lead potential players to *expect* larger sales in this draw so that the expectation of the expected value will be larger, the expected variance will be larger, and the expected level of skewness will be less left skewed, than would be the case with no rollover.

Here, we exploit a peculiar feature of the game whereby the proportion of winners in each prize draw is a random variable that depends on the winning numbers that are actually drawn. Some numbers are more popular than others. In draws where unpopular numbers are drawn, there are few winners *of prizes below the jackpot level* (as well as few winners of the jackpot). One purpose of these smaller prizes in lotto design is that many small winners might encourage many players to think that winning big is more likely than it really is. Moreover, many players participate through consortia—groups of workplace colleagues or family members who agree (explicitly or not) to share their winnings. When such consortia win small prizes, they are thought to often reinvest their winnings, rather than take the trouble of dividing the winnings into very small amounts per consortium member. Thus, for a given level of sales in a particular draw, the number of small prize winners (which is a random variable, conditional on overall sales) does affect the level of sales in the subsequent draw. This will be true irrespective of any serial correlation in sales. Our modeling strategy is to use this property to isolate the effect of moments on sales. Once we allow for this endogeneity of the moments of the prize distribution, we find quite conclusive evidence that skewness is an important factor in driving lottery sales.

The rest of the chapter is organized as follows. Section 2 explains lotto games and how to calculate their prize distributions, which is fundamental to understanding lotto design. Section 3 builds on this to work out the expected value of a lottery ticket, and Section 4 extends this to include higher moments of the prize distribution. Section 5 explains the econometric methodology that provides unbiased estimates of the effects of exogenous variation in the moments. Section 6 presents some tentative simulations

of how sales would be affected by variation in the major parameters in the design of the game. Section 7 concludes with some directions for future research.

2. THE ODDS OF WINNING A (PARI-MUTUEL) LOTTERY

Lotteries have traditionally featured selection mechanisms that have numbered tickets drawn at random from an urn. While the details differ from game to game, modern incarnations of the mechanism usually feature N numbered balls bouncing in a transparent, often rotating, container, from which n are drawn (usually without replacement). Individuals buy tickets by marking n numbers on a printed matrix of N numbers on an entry form. The form is scanned electronically and a ticket is printed out and given to the customer as a record. Each combination of n numbers has the same chance of winning given by $p = N! / n!(N - n)!$. For example, in the commonly used $n = 6$, $N = 49$ design, $p = 1/13,983,816$. Thus, if sales are 40,000,000, then we would expect, on average, the number of winners to be approximately 3. The probability that a given ticket does *not* win is $1 - p$. If two tickets are sold, then the probability that *neither* are winning tickets is $(1 - p)^2$, and so on, for three tickets, four, and on and on. Thus the probability that there are *no* winners (that is, that there is a rollover) in draw t when sales are S_t is given by $(1 - p)^{S_t}$. Thus, suppose sales are around 40 (20) million and p is 1 in 14 million, then the probability of a rollover is approximately 6% (25%).

The figures correspond approximately to UK figures for the Saturday and Wednesday lotto games. In practice, we observe many more rollovers in lottery games than would be implied by the observed level of sales and the discrepancy are due to the extent to which there are similarities in the way players choose their numbers—a phenomenon known as conscious selection. This can be accommodated into the analysis by allowing the rollover probability to be $(1 - p)^{\alpha S_t}$ with $\alpha \leq 1$, such that $\alpha = 1$ implies no conscious selection (i.e., completely random choice) and $\alpha = 0$ implies that all players have chosen exactly the same combination of numbers.³

In practice, the tedium of lotto games causes sales to typically exhibit a long run secular decline. Thus, rollovers are doubly important if there is serial correlation in sales. Rollovers cause sales to rise in the next draw and the serial correlation then causes sales to rise in the draw after that (and after that . . .), albeit by to a lesser extent, in all subsequent draws. Thus rollovers, of the appropriate frequency and size, can offset the tendency for sales to decline because of tedium. However, more rollovers are not necessarily better. If rollovers have high frequency, then players will come to expect them and will engage in *intertemporal substitution*—reserving their lottery spend until a rollover (or several) has already occurred before playing. The optimal frequency of

³See Hartley et al. (2000) for an analysis of the *form* of conscious selection and its extent, where we estimate the extent to which the distribution of the number of winners of each prize pool departs from the theoretical distribution when there is no conscious selection.

rollovers will depend on the relative strengths of the trend decline, the intertemporal elasticity of substitution, and the degree of serial correlation.

3. THE EXPECTED VALUE CALCULATION

In raffles, the expected value of a ticket falls with sales since the prizes are fixed. At low levels of sales, tickets are valuable, and as sales rise their value falls to zero. The operator faces the risk that sales revenue will not exceed the cost of prizes and the players face the risk that they have to compete with many other players for the prizes. In lotto games, the design is pari-mutuel and players compete for *shares* of the prize pool(s) as opposed to fixed prizes. Thus, as sales rise, the prize pool expands, and the expected winnings (and losses) remain the same. However, the possibility of a rollover implies that the (top) prize pool might not be won at all in this draw, and this depresses the expected value of participating in the current draw. Thus, sales have an additional effect on the expected value of participating in any given draw—higher sales cause the rollover probability to fall which raises the expected value. Thus, the expected value of a ticket in a given draw depends on the size of the prize fund, which is the proportion of sales that is not taken out as tax and operator costs (i.e., $[1 - \tau]S_t$ where τ is the take-out rate) plus any rollover from the previous draw. But this is multiplied by the probability that there is at least one winner, that is, $1 - (1 - p)^{S_t}$.

Further details of the algebra of the calculation of the expected value are in the Appendix, but Figures 1 and 2 capture the important intuition. Figure 1 shows how the expected value or mean return, call this m_{1t} , of a lottery ticket that costs \$1 for common types of design in a regular (non-rollover) draw. In the figure the take-out rate τ is set at 0.55, which is a typical value. The shape of this figure has given rise to what has been called lotto's peculiar economies of scale,⁴ since it shows that the game gets cheaper to play (in the sense that the expected value gets higher) the higher are the sales. This is because the higher sales are, the smaller is the chance of a rollover occurring, because more of the possible combinations are sold. This makes the return higher in the current draw because rollovers take money from the current draw and add it to the next draw; and your ticket in this draw gives you a possible claim on prizes in this draw but not the next. So the higher the chances that a jackpot rolls over, the less a ticket for the current draw is worth. Indeed, from an individual player's perspective, it doesn't matter if the money is rolled over or given to another player in this draw, only that he or she does not win.

At very large levels of sales, all game designs have the same mean return, which simply equals the $1 - \tau$, because the chance of a rollover is very small when ticket sales are very large since most possible combinations will be sold. At any given level of sales, easier games offer better value in regular draws since the rollover chance is smaller.

⁴See Cook and Clotfelter (1993).

When a rollover of size J_{t-1} has taken place, the expected value function shifts upwards by an amount equal to J_{t-1}/S_t (because everyone has the same chance of getting the rollover amount) and this diminishes as S_t rises. Thus, the function shifts upward more at low sales than at high sales and the shifted function will, in general, have a single maximum. Figure 2 shows the shape of the m_{1t} function for regular and rollover cases. The effects of rollover on expected value, and thence on sales, will be small: when sales are high, and when J_{t-1} is small because s_n is small or because τ is high.

A double rollover, when the jackpot had not been won for two earlier draws, would shift the m_{1t} even higher upward. In principle, with multiple rollovers the expected value *could* exceed the cost of the ticket (see Ziemba et al., 1986)—although we would expect

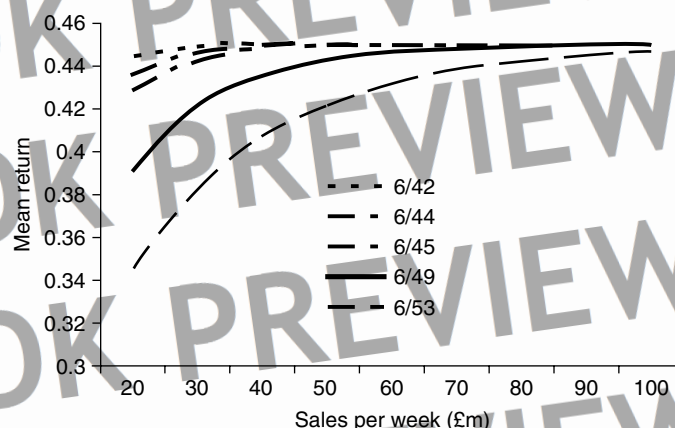


FIGURE 1 Lotto's peculiar economies of scale: no rollover case.

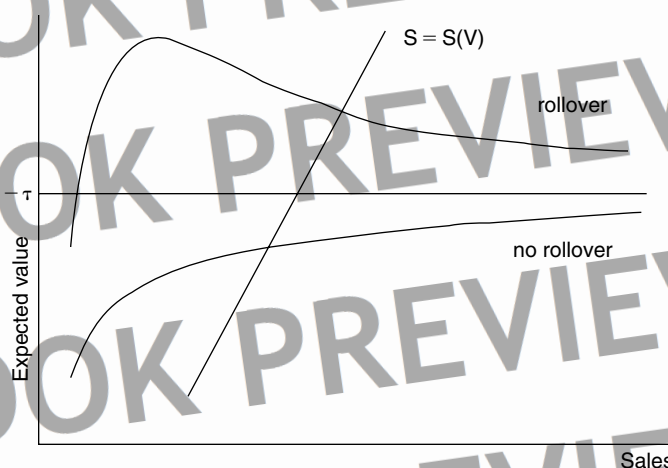


FIGURE 2 Expected value in rollover and regular draws.

sales to rise with the expected value and consequently we would be unlikely to observe tickets being worth more than their cost.

4. HIGHER MOMENTS OF THE PRIZE DISTRIBUTION

It seems likely that sales are affected by *higher* moments of the prize distribution as well as by the first moment (i.e., the expected value of the prize distribution)—in particular, sales are likely to be affected by the variance and skewness of the prize distribution. Indeed, if *only* the first moment mattered, then we would find it difficult to reconcile gambling with aversion to risk. Individuals who are risk-averse would not be expected to participate in lotto games that offered expected values that were less than the cost of participation. There is accumulating evidence that gambling is affected by the skewness in the prize distribution.⁵ That is, lotteries offer the possibility of a huge change in lifestyle. Lottery prize distributions are highly left skewed—the vast majority of players lose a small amount (the stake), a small number win small prizes, and a very small number win large prizes. However, the existing studies neglect the fact that the observed moments in prize distributions are themselves a function of the bets that are placed. For example, as sales rise in lotto, the expected value rises, and this makes participating more attractive. Moreover, lottery operators have chosen the design of the game to deliver a particular vector of moments, taking into account the likely size of the market. If the market is expected to be small, then we might expect the operator to choose a different design (one that is easier to win) than if the market were large. Similarly, as well as the expected value, the higher moments of the prize distribution will be affected by sales. Thus, the mechanical relationship between sales and the moments of the prize distribution may contaminate the response of sales to variation in the moments that are due to chance.

The distribution of winnings implied by the n/N design typically has a large spike at zero, since most players lose altogether, and successive peaks corresponding to matching more of the n winning numbers until a final peak occurs for matching the n -ball and winning (a share in) the jackpot prize pool. The distribution at these further local maxima associated with more difficult-to-win prizes arises because the amount won depends on the number of people who also win a share in each prize pool. Thus, instead of a spike, there is a small peak with a (local) maximum in the distribution for each prize type, which corresponds to the most probable number of winners for that type, but around this is a distribution that arises because there may be fewer winners each getting a larger share of the pool or more winners than expected each getting a smaller share. Successive peaks, corresponding to the mean winnings of bigger prizes, are lower (as the chance of winning is smaller) and narrower (because the variance in the number of prize winners is lower for the more difficult to win prizes). The overall distribution is thus left skewed (by the large majority of losers) but a

⁵See Golec and Tamarkin (1998) for U.S. racetrack betting, and Garrett and Sobel (1999) and Kearney (2005) for U.S. lotteries.

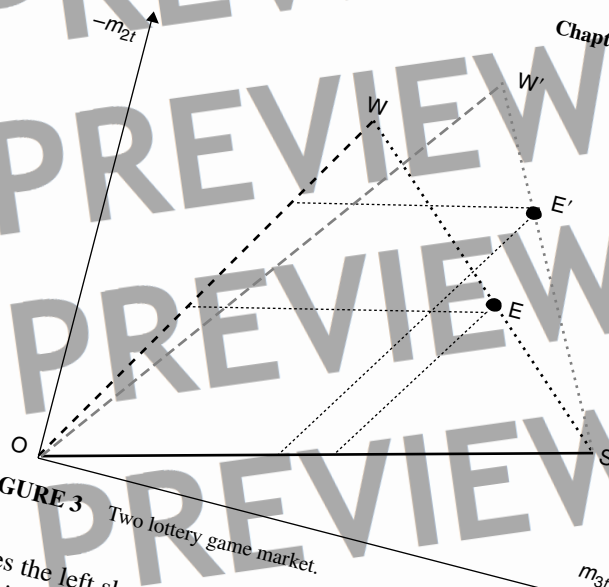


FIGURE 3 Two lottery game market.

rollover decreases the left skew since it increases the size of the jackpot pool. The formulae for the variance and skewness are rather complicated functions of the rollover size and the game design parameters and can be seen in the Appendix to Walker and Young (2001).⁶

The characteristics approach to demand due to Lancaster (1971) is useful for thinking about sales of closely related products.⁷ The essence of the approach is that, in contrast to conventional microeconomic theory, goods are not valued for their *goodness* but rather for the bundle of characteristics that they contain. So goods that are good substitutes for each other are likely to contain similar combinations of characteristics. In the context of the lottery market, the products sold effectively differ in their prize distributions and a convenient way of summarizing these distributions is via their first three moments. The first moment, the expected value m_{1t} , is effectively the *price* variable since the expected cost of participating is the face price (usually one unit of currency) minus the expected value. Therefore, each lottery ticket costs $1 - m_{1t}$ and the larger is the expected value of a ticket the better the bet is. Apart from this mean return to a ticket, a ticket is characterized by design parameters that imply specific variance (m_{2t}), which is a characteristic that is disliked, and specific skewness (m_{3t}), which is liked.

Suppose a market is characterized by two lottery games, labeled W and S, both with the same take-out rate for convenience. The position is illustrated in Figure 3 where each

⁶That paper makes clear that there is a strong correlation between each of the moments of the prize distribution. In particular, locally at least, they each respond in similar ways to a variation in sales. What is required to break this multicollinearity is data that contains sufficient variation in sales.

⁷The first application was by Gorman (1991) that modeled egg sales in the UK. More recently, the approach has been used for thinking about “brands” of goods (see Hausman, 1997; and Putrin, 2002).

product is portrayed as a ray in characteristics space. The lengths of the rays are given by multiples of $1 - m_{1t}$ to indicate their cost. Both games have specific combinations of $-m_{2t}$ (variance is disliked, so it is measured negatively in the figure) and m_{3t} (skewness) with the W game being less skewed, at given variance, than the S game. Each ray indicates a ticket type and movements along the ray away indicate higher levels of sales.

We could envisage the population having a distribution of preferences between $-m_{2t}$ and m_{3t} with many preferring the heavily skewed S game but, nonetheless, some preferring the W game. Aggregate sales of the two products can be described by a position along the line WS. Imagine that the position E is chosen where the market has a slightly higher proportion of S-type tickets than W-type tickets (as seen by their distances along the respective rays from the origin). Now consider what happens when the W game experiences a design change that results in a price fall so the line OW gets longer since more W-type tickets can be afforded. And suppose that the skewness in W's prize distribution, at any level of variance, gets larger, so OW gets less steep. Thus, W' has become a more attractive bet and the frontier for the aggregate market changes to W'S. Now that W has become both cheaper and more similar in the combination of characteristics that it offers to S, we would expect E to move to a position like E' where many more W-type tickets are bought and somewhat fewer S-type. Note that if only S existed, sales would be OS. If W is now introduced, then sales of S slump, but this is more than compensated for by the extra W sales. Thus, diversification of the game portfolio can increase aggregate sales because two products better cater to the distribution of preferences for the characteristics in the population.⁸

5. ECONOMETRIC METHODOLOGY, DATA, AND ESTIMATES

One way of summarizing the complications of how all the various aspects of game design impacts on sales is through the mean, variance, and skewness of the prize distribution.⁹ for example, one might allow $S_t = s(m_{1t}, m_{2t}, m_{3t}, X_t) + \varepsilon_t$ by choosing some parametric form for $S(\cdot)$.¹⁰ The term ε_t captures the effects of all unobservable factors (and may be correlated with the same factors in previous draws—serial correlation) on sales; while X_t is a vector of observable factors that, independently of

⁸Note that m_{2t} and m_{3t} are not, in general, independent of each other—it may be difficult to have more of one without having more of the other also. While it will not be possible to have any combination of the two, one may be able to expand the set of possible combinations through more a complex design—such as a bonus ball.

⁹Note that each moment depends in a known and deterministic way on n, N, τ, s (the vector of s 's), and on S_t and J_{t-1} . That is, $m_{kt} = m_{kt}(n, N, \tau, s, S_t, R_{t-1} J_{t-1})$, for $k = 1, 2, 3$, and so it is clear that, while sales depend on the moments, the values of the moments also depend on sales. Thus, when estimating the causal effects of the moments on sales, account needs to be taken of their endogeneity.

¹⁰For a simple example, Farrell et al. (2000) assumed that $S(\cdot)$ was a linear function of m_{1t} and did not depend on higher moments at all.

the parameters of the game, affect sales. Most other research has focused on the role of such X_t 's in determining sales.¹¹ One modeling approach would be to substitute the formulas that show how the moments are related to the game design variables into $S(\cdot)$ to obtain a *reduced form* model of sales whereby $S_t = F(n, N, \tau, \pi, R_t J_{t-1}) + u_t$ where $R_t = 1$ if there was no jackpot winner in $t - 1$. This $F(\cdot)$ function is a complicated, highly nonlinear function, and it is not practical to estimate the structural model coefficients from this reduced form. A special case of $F(\cdot)$ arises when all of the game design parameters are fixed in the time series being modeled, so that only the rollovers determine sales variation across time, although this will still be highly nonlinear.¹²

A fundamental issue is that while $R_t J_{t-1}$ affects S_t in $F(\cdot)$, it also depends on S_{t-1} , because lagged sales affect the rollover probability and also determine the size of the previous jackpot. Thus, the presence of serial correlation is sufficient to undermine the assumption that rollovers occur randomly, since it implies that S_t depends on S_{t-1} directly, as well as indirectly via any rollover.

However, conscious selection is a strong phenomenon in lotto markets. Operators encourage it since having a sense of ownership over the numbers that one bets on makes playing more persistent—individuals are more likely to play in every draw if they feel that they are playing their *own* numbers. That is, conscious selection induces serial correlation in sales. Moreover, it is thought that allowing players to choose their own numbers has a large impact on sales.

If players consciously select their numbers when they bet, then there is more likely to be large variation in the number of prize winners in each prize pool than if there were no conscious selection. In particular, in 6/49 the number of three-ball matches is likely to vary considerably from draw to draw. The number of winners of the higher prize pools will also vary for the same reason. With sales of 40 million in 6/49, we would expect the number of three-ball winners to be approximately 700,000 with a standard deviation of approximately 100,000. Three-ball winners usually win a modest prize¹³ and it is common for players to *reinvest* these modest levels of winnings in more lottery tickets in the next draw.¹⁴ Moreover, a substantial minority of sales is accounted for

¹¹Notice also that the specification treats rollover-driven *temporary* variation in the moments as equivalent to design-induced *permanent* variation in moments. Nonetheless, it is useful, for the moment, to think of $S(\cdot)$ as capturing the structural determinants of sales—that is, the variation in sales that is due to variation in the moments that arise for any reason.

¹²Moreover, $F(\cdot)$ is only the reduced form of $S(\cdot)$ if ε_t exhibits no serial correlation. If it does, then the rollover size would be determined by previous sales and by the level of sales *two* draws previously. The reduced form would then have to substitute $R_t J_{t-1}$, $R_{t-1} J_{t-2}$, and so on, out of the model recursively. Moreover, a further problem arises because total sales, and hence the moments, are not known when individuals are making purchase decisions, so players have to estimate the moments from the information that is known prior to the draw—and this information set will include the value of $R_{t-1} J_{t-1}$.

¹³In the UK game, the three-ball prize is not *pari-mutuel*. Rather it is a fixed prize of £10.

¹⁴Retailers are empowered to pay out the small prizes. This, in itself, will encourage players to spend their winnings on tickets for the next draw, when they collect their winnings. Indeed Guryan and Kearney (2005) identify a *lucky store* phenomenon that arises because of habit persistence in players' expenditures.

by syndicates¹⁵ who are thought to reinvest such small dividends to save the trouble of distributing them to syndicate members. Thus, in practice, the number of three-ball winners affects subsequent sales.¹⁶ Thus, we use an approach that uses the proportion of three-ball winners in the $t - 2$ draw to explain the variation in $R_t J_{t-1}$ and then use the predicted value of this instead of the actual value. The assumption here is that three-ball (and four, five, five+bonus, and six) winners in draw $t - 2$ affect sales in $t - 1$, and hence the probability and size of any rollover, but they are assumed not to *directly* affect sales in t . This seems like a reasonable assumption.

The data is the sales and rollover information, winning number, and number of winners of each prize pool that Camelot, the UK operator, is obliged to publish for every draw. The data is sales in the Saturday and Wednesday online lotto games in the UK from the first draw, though the introduction of the Wednesday game was introduced in early 1997, to draw 310 in mid-1999 just before the start of the Thunderball game was introduced.¹⁷ This spans the period when the portfolio of games consisted of the Saturday lotto game, then the Wednesday and Saturday 6/49 lotto games, then scratchcards, which were introduced in early 1998,¹⁸ as well as the two lotto draws.

Figure 4 shows the history of sales over this period and for some time subsequent to this. The estimation period covered a time when there was rapid growth in sales fueled, in part, by several large rollovers which are indicated by the sharp spikes in sales; followed by the introduction of the Wednesday lotto draw; and then the period, after the introduction of scratchcards, that saw a rapid decline in scratchcard sales, a slow decline in Saturday sales, and an even slower decline in Wednesday sales.¹⁹

We assume that the parametric form for Saturday and Wednesday sales, $S^s(\cdot)$ and $S^w(\cdot)$, are log-linear. That is, the log sales in each draw is assumed to be a linear function of the log of the moments.²⁰ We include only the moments relating to the *current* draw in each equation. We allowed for serial correlation and estimation was conducted using seemingly unrelated regression methods to allow for the correlation in the residuals across draws. That is, we allow the residual from the sales on a Wednesday (Saturday) to affect both the coming Saturday's sales as well as the following Wednesday's.²¹ That is, we estimate

¹⁵These are consortia of players who agree to contribute to the stakes each week and share the winnings. They are commonly organized at the workplace.

¹⁶A simple correlation in our data for our two draws suggests that an additional 100,000 three-ball winners in the large game draw at t will increase sales in $t + 1$ by a statistically significant 178,000 in the large game, while the same for the small draw raises subsequent sales by 128,000.

¹⁷We reserve the data beyond draw 310 as a *hold-out* sample that allows us to conduct forecasting tests.

¹⁸The results were very similar when the dataset was extended to later periods when further portfolio diversification took place.

¹⁹One of the peculiar features of the two games in the UK is that they were linked via their rollovers. That is, if there are no winners, then the jackpot from Wednesday (Saturday) is added to the jackpot on the following Saturday (Wednesday).

²⁰When computing the moments, we allowed for conscious selection by calibrating α for each game so that the level of sales was consistent with the correct number of rollovers across the estimation period. We use a log specification to avoid negative predictions in simulation.

²¹We ignore scratchcard sales in our analysis and some very small games—the *Extra* games which were extremely unpopular with players.

²²Estimation was conducted using GMM provided by the *xtabond* facility in STATA 10.

Our vector X includes data on daytime maximum temperature and rainfall (the averages over the week of sale prior to the draw, averaged over all UK weather stations); seasonal (month) control variables; a cubic trend; and the sales of scratchcards in the week of sale. Further control variables were included in experiments and checks for robustness, but even when they were statistically significant, they failed to change the coefficients on the moments. Indeed, the estimated coefficients hardly change at all even when *all* of the control variables (except the cubic trend) are excluded. An important innovation in this research is that our estimation method treats the dataset as an *unbalanced* panel (of two games)—it is unbalanced because the Wednesday game was not introduced until 15 months after the establishment of the Saturday draw. The specification is a logical one but explicitly excludes any dynamic effects apart from the residuals. This is quite restrictive in this context, but relaxing this specification gives rise to significant estimation difficulties.²² The results are presented in Table 1. We compare results with OLS (where the m 's are treated as exogenous but we still allow for serial correlation). The standard errors are robust to heteroskedasticity and a variety of tests of specification, parameter

$$S^s_t = \beta^s X^s_t + \gamma^s_1 m^s_{1t} + \gamma^s_2 m^s_{2t} + \gamma^s_3 m^s_{3t} + \varepsilon^s_t \quad \text{where } \varepsilon^s_t = \rho^{ss} \varepsilon^{s}_{t-1} + \rho^{ss} \varepsilon^{s}_{t-1}$$
$$S^w_t = \beta^w X^w_t + \gamma^w_1 m^w_{1t} + \gamma^w_2 m^w_{2t} + \gamma^w_3 m^w_{3t} + \varepsilon^w_t \quad \text{where } \varepsilon^w_t = \rho^{ww} \varepsilon^{w}_{t-1} + \rho^{ws} \varepsilon^{s}_{t-1}$$

FIGURE 4 UK sales (million pounds per week).

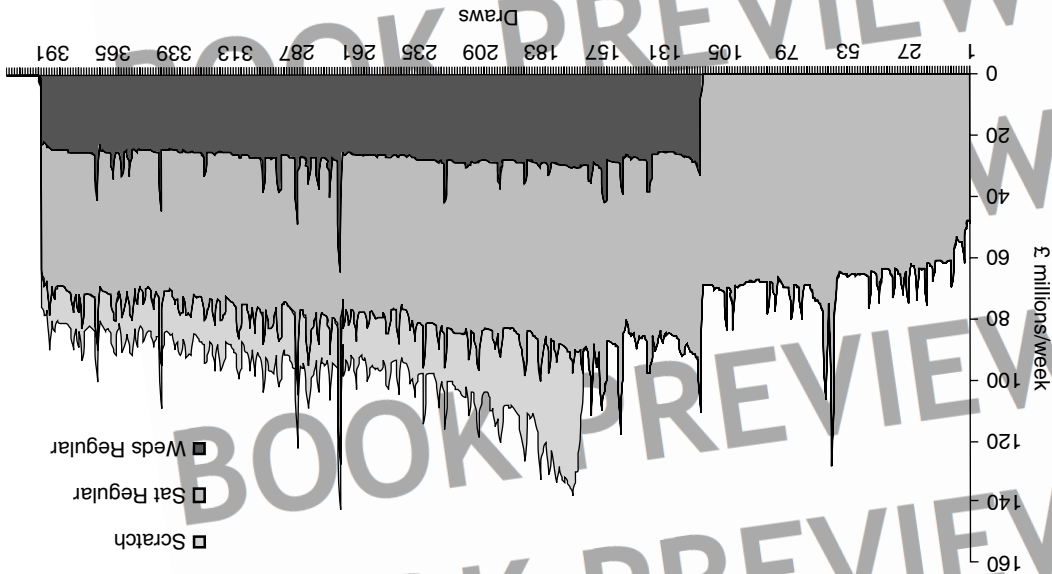


TABLE 1 Estimated Parameters of Log Sales Equations

	OLS using actual jackpot to determine m 's		GMM using $t - 2$ winners of each prize pool	
	Wednesday	Saturday	Wednesday	Saturday
m_1	0.344 (0.090)	0.211 (0.076)	0.462 (0.111)	0.442 (0.187)
m_2	-0.042 (0.009)	-0.072 (0.007)	-0.062 (0.016)	-0.098 (0.036)
m_3	0.062 (0.024)	0.111 (0.023)	0.204 (0.080)	0.160 (0.042)
ρ^w	0.057 (0.019)	0.003 (0.022)	0.050 (0.017)	0.010 (0.021)
ρ^s	-0.001 (0.009)	0.123 (0.020)	-0.002 (0.008)	0.273 (0.081)
R^2	0.932	0.800	0.938	0.821

NOTE: All equations contain control variables listed in the text. First stage results for the estimates show that the moments are significantly positively affected by the number of three-ball and four-ball winners, and significantly negatively affected by the number of jackpot winners. The results above were not sensitive to the exclusion of the number of jackpot winners, which was heavily dominated by one draw where, because popular numbers were drawn, there were more than 100 winners.

stability, and forecasting power, were employed.²³ The OLS results show that sales are a statistically significant increasing function of the mean of the prize distribution—so better bets are more attractive ones; sales are a statistically significant but decreasing function of the variance in the prize distribution—so riskier bets are less attractive; and sales are a statistically significant and increasing function of the skewness of the prize

²³A number of specification tests fail in the OLS forecast, suggesting that the endogeneity of the moments is an important factor. Tests of their endogeneity strongly suggest that this is a problem—especially in the Wednesday game. This is to be expected, since when sales are small, the moments are considerably more sensitive to sales variation and jackpot size than when sales are low. In the GMM model we find that the AR test is passed, indicating absence of higher order autocorrelation than the first order allowed here, although this is sensitive to the inclusion of the cross-correlation between the two games. The ARCH test for heteroscedasticity also passes, which is surprising given how the variance of sales seems larger in rollover weeks compared to regular weeks. The normality test passes, although this failed when we tried to include earlier data in the analysis—possibly because of the large outliers associated with early double rollovers. The parameter constancy tests for both models (which were carried out on the remaining observations up to draw 395) narrowly passed. The RESET test, however fails (albeit marginally) for Wednesday suggesting that there is some further specification problem. It seems likely that this is associated with the use of a fixed degree of conscious selection across draws. This implies that rollover frequency, conditional on sales, should be constant across draws. In fact, there is some suggestion that rollover frequency is falling. Moreover, it seems likely that the degree of conscious selection is lower in rollover draws, because players often wish to purchase more than their usual number of tickets and may be more likely to use the random number generator built in to the retailers' tills for this purpose. These nuances will be explored in subsequent research.

distribution—so players appear to exhibit a preference for skewness. The GMM estimates are also similar, although the Wednesday results, in this case, show somewhat higher sensitivity to all three moments which is consistent with endogeneity being a bigger problem in the Wednesday game because at lower level of sales, in 6/49, the moments are more sensitive to the level of sales, and to the jackpot size than is the case at higher level of sales.

Although the qualitative results support our original vision—that sales depend positively on the mean of the prize distribution, negatively on its variance, and positively on its skewness—the quantitative importance of the results are less straightforward to comprehend. To give a feel for what the results mean in quantitative terms, imagine a (large) rollover that adds £5,000,000 to the Saturday jackpot, where Saturday sales are approximately 40,000,000. This bonus would increase the expected value by approximately 25%, given the UK lotto design parameters, and would cause a modest rise in variance, and a large increase in skewness (of about 10%). Thus, according to the GMM estimates in the final column of Table 1, this would raise sales by about 13% though the effect on the expected value and a further 2.5% because of its effect on skewness, offset by an effect of more variance of slightly less than 1%. There would be a jump in sales of about 13%. But note that the serial correlation would raise sales the following Saturday by about 4% over what would have happened (with a small effect on Wednesday, too). Cumulating the effects over time, and across both games, the effect of a £5,000,000 rollover would be to raise sales by approximately £8,000,000.²⁴

6. GAME DESIGN SIMULATIONS

Making inferences from the observed estimated relationships between sales and the moments requires that we solve the estimated equations for sales as a function of the design parameters. This is complicated because these equations are highly nonlinear and do not admit an analytical solution.²⁵ Throughout, we use the estimates from the GMM estimation procedure—note that these estimates tended to imply that sales were more sensitive to the moments than was the case for OLS. The simulation strategy was to draw an initial value for ε_t for each game, solve the model for initial sales for given values of the design parameters, then use this solution to generate a rollover probability

²⁴The operator might be tempted to *lace* the jackpot in order to improve sales. However, even with the large estimated sensitivity to the jackpot implied by these results, this is unlikely to be profitable for the operator. This is for two reasons. First, the operator retains only 55% of the additional sales (i.e., a little over £4,000,000). Secondly, the assumption has been that this £5,000,000 arises from a randomly occurring rollover—if the operator were to try to make such bonuses occur on a regular basis, players would come to expect them, and the model would cease to be a valid description of sales behavior. In fact, in many cases, operators do have some discretion over adding bonuses to the jackpot pool (in the UK this is referred to as a *Superdraw*). In practice, they are used sparingly, and largely just to offset the effects of a temporary drought of rollovers.

²⁵The findroot command in *Mathematica*, which solves nonpolynomial expressions using the Jenkins-Traub algorithm was used and applied to simplified expressions for the moments. In practice, this proved very time-consuming, which limited the range of simulations that could be conducted to those with $S_n = 1$.

and a level of the rollover conditional on one occurring in the next draw. Whether or not a rollover occurred was chosen at random from the distribution with the mean given by the predicted rollover probability. This then generated a subsequent sales prediction and the rollover probability and conditional sales were solved, and so on. This exercise was repeated for 1,000 draws and the average sales computed. This process was looped a further 1,000 times, drawing a new initial ε_t on each occasion. The figures report the average sales from these 1,000 simulated 1,000-draw histories. Our simulations are entirely illustrative. To simplify the calculations, the π vector is fixed so that $\pi_n = 1$ and all π 's for lower prize pools were set to zero. The lack of lower level prizes implies that we are analyzing hypothetical games with a very high variance and high skewness. We concentrate here on the effects of τ and of n on revenue (N is fixed at 49). Figures 5 and 6 show the effects of varying τ and of n where we have added together the sales figures from Wednesday and Saturday games to compute the predicted sales and revenue (R) for an average week over the 250 draw period. In Figure 5 $\tau = 0.5$ while in Figure 6 $N = 49$. Figure 5 suggests that a harder to win game, such as 6/53 would raise higher revenue—almost 20% higher than 6/49. It also suggests that making the game very hard to win, beyond 53, would cause sales and revenue to fall steeply. Figure 6 shows the effects of the take-out rate—in this case revenue is fairly insensitive to the take-out over a wide range although the revenue maximizing rate approximately 0.4.²⁶

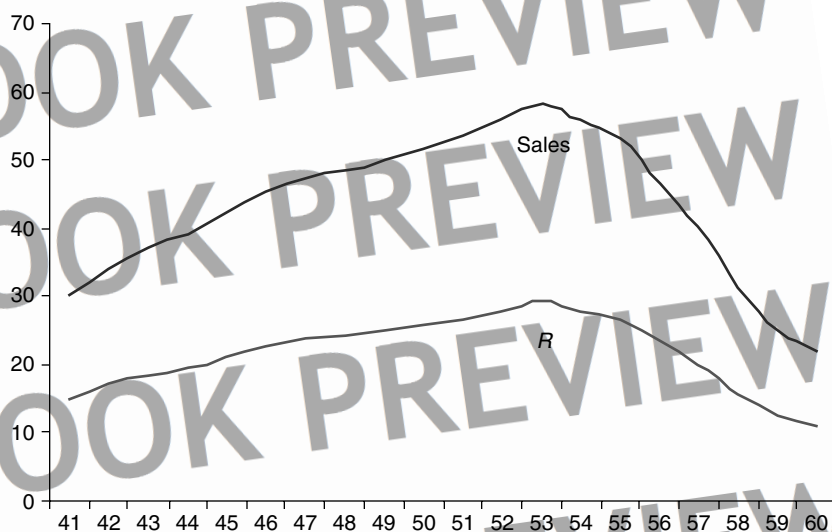


FIGURE 5 Effects of varying n .

²⁶Unfortunately, it has not been possible to compute standard errors around these forecasts, so it is difficult to say how precise these forecasts are likely to be.

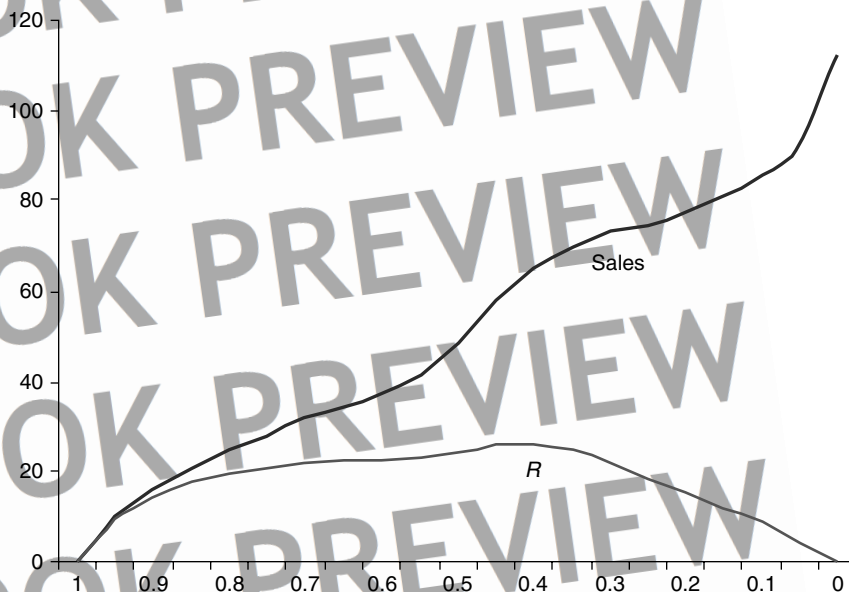


FIGURE 6 Effects of varying τ .

7. CONCLUSION

Our analysis has considered some of the important questions relevant to running a lottery. Our methodology for analyzing the implications of game design is analytically rigorous and yet it reflects the informal received wisdom that dominate industry debate. Thus, it probably captures many of the important features of realities of the game but provides a degree of abstraction from reality to allow counterfactual changes to be analyzed in a formal and quantitative way.

However, our results need to be qualified. The simulations assume that sales respond to variations in mean, variance, and skewness from design changes in the same way as they respond to these variables when rollovers occur. However, it is plausible that people may respond differently to these two types of changes. First, changes induced by occasional rollovers allow for the possibility of substitution between draws, but this possibility does not exist for changes coming through the game design rather than rollovers. This suggests that ticket sales are higher when changes come from rollovers than from game design. Rollovers are rather like sales promotions—they induce people to change their behavior quite differently to a temporary difference in the offer compared to how they would change their behavior for a permanent change. This failure to come to grips with intertemporal substitution is an important shortcoming on the present research. It would be difficult to overcome this in the UK data because, so far, there have been no clean and simple game changes that would allow us to challenge the implicit assumption that sales respond to temporary changes in moments in the same way as to permanent

ones. However, some preliminary work using Irish time series data, which contains several simple game design changes, suggests that, despite the large incentives to engage in this, intertemporal substitution does not seem to be statistically important.²⁷

There are several avenues for further research. The econometric analysis needs to be extended to incorporate newer games if we are going to be able to exploit a longer run of data. There are some practical difficulties in doing this because their introduction is potentially endogenous and because the newer games have typically involved rather low levels of sales. Second, hypothecation of the lottery revenue may play in determining sales. This is an interesting suggestion and could be incorporated into a time series analysis such as the present one providing there is sufficient variation in how the hypothecation is done. In the UK, there have only been limited changes. U.S. data offers better prospects for this kind of exploration because there have been changes in the hypothecation across states and across time. Third, we have not incorporated conscious selection into the moments of the prize distribution—Hartley et al. (2000) suggest that the first moment was largely unaffected by the considerable conscious selection that they found in the early sales data. However, we do not know if this generalizes to higher moments. Finally, this essentially time series work could usefully be complemented by a microeconomic analysis of the effects of income using cross section data in order to compute the regressivity of the take-out. The share of income spent on lotteries declines steeply across the income distribution and very high income shares seem to be confined to quite low income households. Such a microeconomic analysis could also incorporate the effects of rollovers if panel data were available. Ultimately, such an analysis could enable researchers to simulate the effects of game design on the size distribution of individual levels of play, as well as on aggregate sales, and hence offer the prospect of *designing out* so-called problem gambling.

References

- Cook, P. J., and C. T. Clotfelter. 1993. The Peculiar Scale Economies of Lotto, *American Economic Review* 83, 634–643.
- Farrell, L., R. Hartley, G. Lanot, and I. Walker. 2000. The Demand for Lotto and the Role of Rollovers and Conscious Selection, *Journal of Business and Economic Statistics* 18, 228–241.
- Garrett, T. A., and R. S. Sobel. 1999. Gamblers Favor Skewness, Not Risk: Further Evidence from United States Lottery Games, *Economics Letters* 63, 85–90.
- Golec, J., and M. Tamarkin. 1998. Bettors Love Skewness, Not Risk, at the Horsetrack, *Journal of Political Economy* 106, 205–225.
- Gorman, W. M. 1980. A Possible Procedure for Analysing Quality Differentials in the Egg Market, *The Review of Economic Studies* 47, 843–856.
- Guryan, J., and M. S. Kearney. 2005. Lucky Stores, Gambling, and Addiction: Empirical Evidence from State Lottery Sales. NBER Working Paper 11287. NBER, Cambridge, MA.
- Hausman, J. 1997. Cellular Telephones, New Products, and the CPI, *Journal of Business and Economic Statistics* 17, 188–194.
- Kearney, M. S. 2005. State Lotteries and Consumer Behavior, *Journal of Public Economics* 89, 2269–2299.

²⁷This, albeit provisional, finding is consistent with there being a strong *present-bias* in individual preferences. That is, individual decisions may be *time inconsistent* because they discount the immediate future more than the more distant future.

- Lancaster, K. 1971. *Consumer Demand: A New Approach*. Columbia University Press, New York.
- MacLean, L., W. T. Ziemba, and G. Blazenko. 1992. Growth vs. Security in Dynamic Investment Analysis, *Management Science* 38, 1562–1585.
- Petrin, A. 2002. Estimating the Consumer Surplus and Welfare Gains from the Introduction of Minivans, *Journal of Political Economy* 110, 705–729.
- Thaler, R., and W. T. Ziemba. 1988. Parimutuel Betting Markets: Racetracks and Lotteries, *Journal of Economic Perspectives* 2, 161–164.
- Walker, I., and J. Young. 2001. An Economist's Guide to Lottery Design, *Economic Journal* 111, F700–722.
- Ziemba, W. T. 2008. Efficiency of Racing, Sports, and Lottery Betting Markets, in D. B. Hausch and W. T. Ziemba (eds.), *Handbook of Sports and Lottery Markets*. North-Holland, Amsterdam, pp. 183–222.
- Ziemba, W. T., S. L. Brumelle, A. Gautier, and S. L. Schwartz. 1986. *Dr. Z's 6/49 Lotto Guidebook*. Dr. Z Investments, San Luis Obispo, CA, USA and Vancouver, BC, Canada.

APPENDIX: The Expected Value Formula

The expected value of a ticket is the first moment of the prize distribution—which is the product of the probability of it being won and the size of the prize if it is won. That is, it is given by $m_{1t} = [1 - (1 - p)^{S_t}] [(1 - \tau) S_t + R_t \times J_{t-1}]$, where J_{t-1} is the jackpot from the previous draw (which equals $(1 - \alpha) \times S_{t-1}$ in the case where the only prize is the jackpot) and R_t is an indicator for whether the previous draw was a rollover or not and is assumed to be known at the time that draw t tickets are purchased.²⁸ The first term in square brackets is the probability of no rollover (i.e., the probability that the jackpot is won) while the second term in square brackets is the expected size of prize.

The level of m_{1t} is unaffected by the nature of the prize distribution²⁹—in the absence of a rollover it depends only on the take-out rate and sales revenue, irrespective of how that fund of revenue is distributed across prize pools. In practice, it is only ever the jackpot prize pool that has no winner and is rolled over—the easier to win prize pools are usually sufficiently easy to win that there are always many winners. Thus, when a rollover from the previous draw occurs, it is only the jackpot n -ball prize pool that is rolled over. Thus, in practice, $J_{t-1} = S_n(1 - \alpha) \times S_{t-1}$ where S_n is the proportion of the overall prize pool going to the n -ball matches (i.e., shared by those who have chosen all n of the winning numbers).

²⁸In most lotto games, it is possible to commit to buying a ticket before it is known whether the previous draw is a rollover or not. For example, one can place an order to buy a ticket for every future draw. In practice, only a small proportion of tickets are bought in this way.

²⁹In practice, lotto games are designed so that the prize fund is spread, in predetermined shares, between those that match all n balls (the jackpot fund), those that match $n - 1$ balls, $n - 2$ balls, and so on. Let these shares be defined by $\pi = (\pi_n, \pi_{n-1}, \pi_{n-2}, \dots)$ which are usually set such that the expected prize in the n -ball pool exceeds that in the $n - 1$ pool, and so on. Some designs are more complex and feature additional *bonus* balls.

DOWNLOAD FULL VERSION HERE: <https://bettingbooks.org/12>

This page intentionally left blank

Download More Sports Betting Books at: www.bettingbooks.org

CHAPTER 23

The Statistics of Lotteries

John Haigh

Mathematics Department, University of Sussex, Falmer, Brighton, UK.

1. Introduction	482
2. Prize Structure and Winning Chances	485
3. Tests of Randomness	491
4. Gambler Choices	496
<i>References</i>	502

Abstract

Lotteries have been a source of state revenue intermittently through the ages. Modern lotteries, where desktop computers in widespread retail outlets can process sales swiftly and efficiently, are seen as a means of extending the tax base without offending voters. The science of statistics is relevant to lotteries in at least four ways.

The *prize structure* of a lottery might be set to maximize revenue. This involves a balance between the proportion of stake money returned to gamblers, its distribution among several prize levels, the expenses of running the lottery, and the tax take.

Lotteries are usually based around “select X numbers from this list of Y numbers,” with the values of X and Y chosen by reference to the expected sales. Simple *probability*, resting on counting the numbers of winning outcomes at a given level, enables the winning chances to be calculated.

The organizers intend that numbers are selected at random, and may televise draws live, with independent scrutineers, to give assurance on this point. However, many gamblers continue to seek patterns in the data. Statistics alone can never prove that the draws are free from fraud or bias, but *tests of randomness* can indicate whether the results are consistent with the organizers’ aims.

Gamblers may seek to use the data on the numbers of prize-winners in pari-mutuel lotteries to *make inferences* on the choices other gamblers are making, as the sizes of prizes depend on how many share them. Thus, even though using less popular choices should make no difference to the chances of winning, a gambler using this information can expect any prizes won to be larger than average.

1. INTRODUCTION

In Chapter 26 of the Book of Numbers, Moses is said to have used a lottery to award land west of the river Jordan. Morton (1990) notes references to lotteries in Roman times, and in the Han Dynasty in China almost 2,000 years ago; van Eyck’s widow used a raffle to dispose of his remaining paintings in 1446. According to Farebrother (1999), the first English state lottery was promoted by Elizabeth I in 1567, but the draw date was postponed to early 1569 to enable more of the 400,000 tickets to be sold. Sporadically, further lotteries to raise money for specific purposes were permitted under Royal Warrant for the next 125 years, and British state lotteries became regular events after 1694.

To give an assurance that all tickets bought had a chance to win a prize, some lotteries used two drums: copies of the tickets sold were selected from Drum A, and the corresponding prize then drawn from Drum B. This was a very slow process. The draw in 1569 took four months to complete. Later lotteries had winners determined instantaneously by devices based on the roulette principle, or by randomly opening a book which had many blank pages, but some with prize sums written in.

For about 30 years from 1694, prizes were paid out as annuities, then later in government stock. Worries about corruption and the morality of the State sponsoring lotteries led to a recommendation in 1808 that they be ended; the last UK lottery in this series took place in 1826. Small lotteries were not legal in Britain again until the 1920s. The Royal Commissions on Gambling in 1951 and 1978 considered their merits, and the present UK National Lottery was set up under an Act of 1993.

Colleges at Harvard, Yale, and Princeton were partly funded by lotteries from the time of American Independence to the Civil War, but a reaction against lotteries during the nineteenth century led to all states, except Louisiana, banning their operation. In 1912, the Totalizator became the only legal means of betting on Canadian horse races. The first modern lottery in the U.S. began in New Hampshire in 1964, and lotteries reentered Canada from 1970. Currently, some form of lottery runs in every Canadian province, and in a large majority of the U.S. states. No inhabited continent is lottery-free.

Even in very recent times, respectable organizations have set up lotteries that had embarrassing flaws in their design. Bellhouse (1982a,b) describes Canadian lotteries in 1978 and 1979 in which some of the tickets sold had significantly higher chances of winning than others. In a separate game in 1981, the organizers of the Scramble Prize would generate a six-digit number as the winning number, and ticket-holders could win prizes if their tickets matched this number, *or any permutation of the digits*; the organizers apparently thought that the ticket (111111) had the same chance as (123456)!

We shall not look at scratchcard-type games (first played in 1974 in Massachusetts). Nor do we do other than mention in passing supplementary games (spiel features) such as the Kicker: players for one game pay an extra sum that brings a random number printed on the ticket into play. In essence, there are four main lottery formats.

The first of these is the so-called Genoese lottery, with symbol m/M . Gamblers select m distinct numbers, with no regard to their order, from a list of M numbers, and win prizes if they match $m, m-1, \dots, m-r$ from a collection of m numbers drawn by the organizers (with small variations). Bellhouse (1991) draws attention to inconsistencies in the written accounts, and to a lack of contemporary evidence to support the specific contention that from about 1620 Genoa used such a 5/90 lottery to select five new senators annually from a list of 90 eligible citizens. However, it is clear that some such method was used in Genoa and other Italian city states, and that gambling on the names or numbers drawn in these lotteries took place. In 1709, Nicolas Bernoulli suggested that the value of a prize should be inversely proportional to the chances of winning, and calculated what values would give a fair game. In 1749, Frederick II of Prussia asked Euler to make some calculations for the chances of winning in a 5/90 lottery.

The second format is a small elaboration, in which as well as selecting m numbers from M , players make a separate selection of one number from N ; we write $m/M(N)$ as shorthand. The North American multi-state games Powerball and Mega Millions are examples, as is Thunderball in the UK. The top prize comes from matching all the numbers drawn in both selections.

Third, Keno. Here the central organizers select m numbers from a list of M , but gamblers make a selection of n of these M numbers, with $n < m$. In some formats, only one value of n is available, and prizes may be given for matching $n, n-1, \dots, n-r$ (and sometimes zero) of the m drawn; in other formats, the gambler has a choice of different values of n .

Finally, Numbers Games. The lottery organizers select m digits (typically $m = 3$ or 4) from the list $\{0, 1, \dots, 9\}$, in a particular order. A gambler hopes to match all of these, or a specified subset.

In all four formats, the prizes offered may be a definite amount (fixed odds) or through a pari-mutuel payout, in which all prize-winning tickets at a particular level share in the prize money allocated to that level. Frequently, a mixture of these methods is used: the lower prizes get a fixed amount, the residue goes to the higher prizes in definite proportions for a pari-mutuel payout. But even when the main publicity for a game asserts that the sizes of the prizes are fixed, the small print normally contains a clause to the effect that, if the number of prize-winners of a stated fixed amount exceeds some unexpectedly large level, then the prize money will be less than advertised, and paid out in pari-mutuel format.

Moore (1997) gives a brief account of the purposes for which lottery proceeds are used. Some lotteries have aimed at raising funds for particular projects—the Sydney Opera House and the 1976 Summer Olympics in Montreal, for example. Provincial lotteries in Canada tend to raise funds both for general purposes, and also specifically for sport, recreation, and culture. Many of the State lotteries in the U.S. name education as an intended prime beneficiary, and other popular causes include support for the environment, State parks, and senior citizens. Indiana subsidizes motorists, Nevada pays part of net proceeds into a compulsive gamblers assistance fund. Lottery profits are shared between government and voluntary organizations in Ireland and in some of the German *Länder*, lotteries in Switzerland and New Zealand support a variety of good causes. The UK government takes 40% of gross lottery sales, some as general funds, and the rest to the Department of Culture, Media, and Sport. Finney (1997) regards the lottery as “an ingenious device for channelling the surplus funds of citizens into good causes with most of which I sympathise.”

The gross proceeds of lotteries are allocated to (a) prize money, (b) central funds (i.e., available to national or local government to spend), (c) retailers of tickets, and (d) administrative costs (including advertising). But the tax treatment of lottery winnings in different countries means that comparisons are not simple. In the UK, the winner of even a £10,000,000 prize is paid the whole sum, immediately, in cash, with no deductions. In the U.S., similar winnings would be taxed as income, so the actual take by central funds is significantly higher than the quoted figure. When the same organizer runs different lotteries, there may be cross-subsidy from one format to another: the overall payback in the UK is set at 50%, but the typical payback on the main pari-mutuel game is normally 45%, with scratchcard paybacks at a higher rate. With these caveats, a fairly typical distribution of the gross ticket sales returns about 50% in prize money, provides 35–40% to central funds, allows retailers some 5–8%, while administrative costs and profits to the organizers account for 5–10%.

2. PRIZE STRUCTURE AND WINNING CHANCES

Lotteries intend that every ticket bought have the same winning chance. In the next section, we describe how this might be tested statistically, but here we assume that this ideal is met. Working out the winning chances for a single ticket in a lottery is then an elementary exercise in counting.

In the Numbers Game format, a player selects up to three (or four) digits from the list $\{0, 1, 2, \dots, 9\}$ *in order*, with possible repetitions, hoping to match those drawn by the lottery organizers. This is frequently played as a fixed odds game: plainly, in the three-number version, there are $10^3 = 1,000$ possible outcomes. The bets available in the Florida version, and the mean returns, are shown in Table 1a and Table 1b. (There is a simple explanation of the apparently peculiar payout of \$1,198 for winning the four-way box in Table 1a: the return to a 50 cent stake would then be \$599, the maximum amount that the player can collect direct from the retail outlet in which the machine is located!) The maximum payback is at the rate of 4999:1 (with a winning chance of one in 10,000), so life-changing winnings are not possible with modest stakes. But when the payout is pari-mutuel, as Chernoff (1981) pointed out, a different analysis is required:

TABLE 1a The Bets Available, the Chances, and the Winning Payouts to a Stake of \$1, in the Florida Play 4 Numbers Game

Name of Bet	Winning chance	Return to \$1	Payback %
Straight	1/10,000	5,000	50
Four-way box aaab	1/2,500	1,198	47.9
Six-way box aabb	3/5,000	800	48
12-way box aabc	3/2,500	400	48
24-way box abed	3/1,250	200	48

NOTE: To "box" a set of four digits is to ask that all different permutations of those digits should win the prize: digits such as 2223 thus give four bets, 2233 gives six, 2234 has 12 bets, and 2345 has 24 bets.

TABLE 1b The Bets Available, the Chances, and the Winning Payouts to a Stake of \$1 in the Florida Cash 3 Numbers Game

Name of Bet	Winning chance	Return to \$1	Payback %
Straight	1/1,000	500	50
Three-way box aab	3/1,000	160	48
Six-way box abc	6/1,000	80	48
Front Pair	1/100	50	50
Back Pair	1/100	50	50

NOTE: (There are other bets listed, but they are combinations of the above.) See Table 1a for the explanation of box.

the winning chances are the same, but observations of the correlation between the size of the prize and the winning numbers can help a player select number combinations that are likely to lead to higher than average winnings. We develop this point later.

(Chernoff also noted a peculiarity of the payout system that meant that players betting on just three of the four numbers drawn had an average payout that was larger than that for players betting on all four digits!)

In the Genoese, Powerball, or Keno lottery formats, the basic tool is the hypergeometric distribution. There is a field of M numbers which, after the lottery draw, can be split into m winning numbers and $M - m$ losing numbers. When a player selects n numbers, the chance they contain exactly k of the winning numbers is

$$\frac{C_k^m \times C_{n-k}^{M-m}}{C_n^M},$$

where

$$C_a^b = \frac{b!}{a!(b-a)!}.$$

When a lottery has this format, and the prizes are of fixed size, irrespective of the number of winners, the analysis is very simple. For example, in the 49's game as operated by the bookmakers in the UK, the Genoese format of selecting six numbers from 49 is used, but gamblers do not themselves select six numbers: they have the choice of selecting 1, 2, 3, 4, or 5 numbers, and win only when ALL the numbers they choose are among the six winning numbers. Table 2 shows the payout odds from Ladbrokes in September 2007. Notice how the mean return decreases as the odds against winning increase—exactly as in the well-known favorite-longshot bias found in betting on horse races.

Ziemba et al. (1986) describe various North American lotteries of Genoese type, noting the values of m and M then in use, the price of tickets, the payback percentage, the proportion of prize money allocated to the jackpot prize, frequency of draws, the size of the relevant population, and typical sales. The North American Association of State and Provincial Lotteries maintains a Website (<http://www.naspl.org>) that outlines what games are currently available in the U.S. and Canada. Even

TABLE 2 The Bets Available, the Chances of Winning, Payout Odds from Ladbrokes in September 2007, and the Rates of Return for the 49's Game, as Played in the UK

Number chosen	Winning chance	Payout odds	Payback %
1	6/49 ~ 0.122	6:1	85.7
2	5/392 ~ 0.013	53:1	68.9
3	5/4,606 ~ 0.00109	600:1	65.2
4	15/211,876 ~ 0.00007	8,000:1	56.6
5	1/317,814 ~ 0.000003	150,000:1	47.2

restricting attention to North America, and without any claim that the list is exhaustive, the following combinations were noted as being on offer: $m = 4$, $M = 100$; $m = 5$, $M =$ any of 26, 30, 31, 32, 33, 35, 36, 37, 38, 39, 42, 52; $m = 6$, $M =$ any of 30, 35, 36, 39, 41, 42, 44, 46, 48, 49, 51, 53, 54, 69; and $m = 7$, $M = 47$.

The 4/100 games are frequently of a special type: the gambler selects four numbers from the list $\{00, 01, 02, \dots, 99\}$, and then a computer-generated selection of six, 14, or even 20 further sets of four numbers are allocated to the gambler. These extra combinations may be grouped into two, three, or four subsets, and not more than one combination in a subset can win a prize. Prizes usually arise if two, three, or all four of the winning numbers drawn occur on a single ticket—merging of tickets is not permitted.

In the 5/ M lotteries, prizes are usually given for matching five, four, three, or occasionally just two of the winning numbers; with 6/ M , normally matching six, five, four, or three wins a prize. A common variation is that one or two supplementary or bonus numbers are drawn by the lottery organizers; gamblers who match $m - r$ winning numbers ($r = 1$, sometimes also $r \geq 2$) and one of the bonus numbers then qualifies for a prize tier. Occasionally, lotteries reward higher spenders by offers such as six entries for the price of five.

Where any form of pari-mutuel prize structure is in place, lotteries have rules about the consequences of there being no prize-winners at a particular level. In the Michigan 5/33 Roll-down, which ran for four years, if there were no jackpot winners, the unwon prize money “rolled down” to be shared by the winners in the next lower tier, but it is far more common for the unwon sums to roll over to the next lottery in sequence. The UK lottery allows up to three rollovers; if the jackpot were unwon for a fourth successive time, the prize money would roll down to the next tier, but many lotteries have no such restrictions. (The UK rules also state that if there are jackpot winners, but no winners in the second tier, the second tier prize money rolls up and is added to the jackpot prize! “Unto every one that hath shall be given,” St. Matthew 25:29.) Thus the payback ratio in rollover lotteries can be far higher than normal, theoretically in excess of 100%, and gamblers or syndicates may be tempted to try to buy large numbers of tickets, in the hope of a certain profit. But the evidence at the end of this section suggests that the increased demand at rollover times varies fairly consistently with the size of the rollover: increasing the average return above, say, 70% from its normal 45%, seems very difficult to achieve.

The use of a second drawing in the Powerball-type $m/M(N)$ enables the lottery organizers to tweak the winning chances even more finely. In September 2007, Powerball itself was 5/55(42), Mega Millions was 5/56(46), California’s Super Lotto Plus was 5/47(27), the UK’s Thunderball was 5/34(14), and 5/45(45) was played in Australia.

Keno is frequently played continuously, with fresh draws every 4–5 min. for over 20 hr each day. In the most common format, the lottery will select 20 numbers from 80, and players select either 1, 2, or up to 10 (occasionally 11 or 12) of the 80 numbers, winning prizes at fixed odds according to how many selections match the winning numbers. (Kansas adds the glitch of a progressive jackpot, paid pari-mutuel, to those who make selections of 6, 7, or 8 numbers.) Players of modest ambition can attempt to

match just one number (chance of success 1 in 4, payout usually \$2 or \$2.50 to a \$1 stake); optimists may attempt to match 10 numbers, having a chance of about one in 9,000,000 of winning the top prize of \$100,000, and better chances of lower prizes (e.g., getting a return of \$1 or \$2 for matching five numbers, at odds of about 1/20).

Tables 3a–f illustrate the variability in the prize structure by comparing four Keno games, identical in play except for the existence of the Kansas progressive jackpot noted earlier. Within a single U.S. state, the mean paybacks across the games offered are reasonably constant, but overall the payback across states varies from 48% to 70%. Massachusetts is consistently the most generous!

Given the overall payback proportion, lottery operators fix the number of prize levels, and how much should be allocated to each. Lotteries are in competition with each other, and with other forms of gambling, such as casinos, horse racing, and other sports events. Any particular lottery, in a given location, faces a different mixture of rivals. In New York, even aside from scratchcard and multi-state games, a resident could play Keno every 4 min. from 5:30 AM to midnight; three-digit or four-digit Numbers Games drawn twice a day; daily 10/80 and 5/39 games; and New York’s own 6/59 Lotto on Wednesdays and Saturdays.

TABLE 3a The Paybacks (in Dollars) to a \$1 Stake in Keno in Four Different States When Gamblers Select 10 Numbers

Match	Odds (1 in)	Georgia	Kansas	Massachusetts	Washington
10	8,911,711	100,000	100,000	100,000	100,000
9	163,381	5,000	2,000	10,000	5,000
8	7,384.5	500	250	500	500
7	620.7	50	50	80	50
6	87.1	10	10	20	5
5	19.4	2	1	2	2
0	21.8	5	5	2	3

NOTE: The mean returns are, respectively, 63.7%, 53.4%, 69.3%, and 48.8%.

TABLE 3b The Paybacks (in Dollars) to a \$1 Stake in Keno in Four Different States When Gamblers Select Nine Numbers

Match	Odds (1 in)	Georgia	Kansas	Massachusetts	Washington
9	1,380,688	30,000	25,000	40,000	25,000
8	30,682	3,000	2,000	4,000	2,500
7	1,690.1	150	100	200	100
6	174.8	25	20	25	10
5	30.7	5	5	5	5
4	8.8	1	1	1	1

NOTE: The mean returns are, respectively, 62.8%, 53.3%, 69.7%, and 49.2%.

TABLE 3c The Paybacks (in Dollars) to a \$1 Stake in Keno in Four Different States When Gamblers Select Eight Numbers

Match	Odds (1 in)	Georgia	Kansas	Massachusetts	Washington
8	230,115	10,000	10,000	15,000	10,000
7	6,232.3	500	250	1,000	500
6	422.5	75	50	50	50
5	54.6	10	10	10	5
4	12.3	2	2	2	2

NOTE: The mean returns are, respectively, 64.7%, 54.8%, 69.0%, and 49.6%.

TABLE 3d The Paybacks (in Dollars) to a \$1 Stake in Keno in Four Different States When Gamblers Select Seven or Six Numbers

Select	Match	Odds (1 in)	Georgia	Kansas	Massachusetts	Washington
7	7	40,979.3	4,000	2,000	5,000	2,500
	6	1,366	125	100	100	100
	5	115.8	15	10	20	10
	4	19.2	3	3	3	2
	3	5.7	1	1	1	1
6	6	7,752.8	1,200	1,000	1,600	1,000
	5	323	50	50	50	40
	4	35	7	5	7	4
	3	7.7	1	1	1	1

NOTE: The mean returns are, respectively, 65.0%, 54.0%, 70.0%, and 50.0% for seven numbers, and 63.9%, 55.7%, 69.1%, and 49.7% for six numbers.

TABLE 3e The Paybacks (in Dollars) to a \$1 Stake in Keno in Four Different States When Gamblers Select Five or Four Numbers

Select	Match	Odds (1 in)	Georgia	Kansas	Massachusetts	Washington
5	5	1,550.6	400	400	450	200
	4	82.7	17	10	20	17
	3	11.9	2	2	2	2
4	4	326.4	70	50	100	24
	3	23.1	5	4	4	5
	2	4.7	1	1	1	1

NOTE: The mean returns are, respectively, 63.2%, 54.7%, 70.0%, and 50.3% for five numbers, and 64.4%, 53.9%, 69.2%, and 50.3% for four numbers.

TABLE 3f The Paybacks (in Dollars) to a \$1 Stake in Keno in Four Different States When Gamblers Select Three, Two, or One Numbers

Select	Match	Odds (1 in)	Georgia	Kansas	Massachusetts	Washington
3	3	72	25	20	25	16
	2	7.2	2	2	2.5	2
2	2	16.6	10	9	5	8
	1	2.63	—	—	1	—
1	1	4	2	2	2.5	2

NOTE: The mean returns are, respectively, 62.5%, 55.6%, 69.4%, and 50% for three numbers, and 60.2%, 54.2%, 68.1%, and 48.2% for two numbers. For one number, the return is 50%, except in Massachusetts where it is 62.5%.

Factors that influence whether a potential lottery player buys a ticket, and if so how many tickets, can be expected to include:

1. The maximum prize possible.
2. The typical value of the maximum prize.
3. The chances of winning a very large prize.¹
4. The chances of winning a moderately large prize.²
5. The chances of winning some prize.
6. The average rate of return (the payback).
7. To what uses the State or Government take are put.

The work of Walker and his colleagues (see, e.g., Farrell et al., 1996) shows how the existence of “free money” rolled over from a previous unwon jackpot can greatly increase sales. This is reinforced by the experience of enormous jackpots in the Powerball and Mega Millions lotteries in the U.S., when gamblers have driven hundreds of miles to a different state, in order to buy the chance to participate in the draw. However, usually only the size of the *jackpot* is affected, with the chances of winning and the sizes of all the other prizes completely unchanged. It is not easy to disentangle the effects on a gambler’s actions of an increase in the mean return, and an increase in the potential jackpot return. Morton (1990) noted a curious phenomenon: when the jackpot is boosted by a given amount from a rollover, the increase in sales is much larger than when a similar boost to the jackpot is given by an artificial addition of funds!

The extent to which sales increase at times of rollovers is illustrated by the UK experience. In the period 1999–2000, typical sales for a normal midweek draw were around

¹Enough for a person of modest lifestyle to retire permanently at a significantly higher living standard.

²Enough to be able to indulge an extravagant dream.

25–28 million, implying a prize fund of £11–12.5 million, including a jackpot fund of about £4,000,000; normal weekend sales were about £50–52 million, with a prize fund of about £23,000,000, including about £7,500,000 in the jackpot. The expected jackpot prize for a winning ticket was £2,000,000.

When the *weekend* jackpot rolled over during that time, subsequent midweek sales increased by about 40% to 36–40 million. Thus the expected total prize fund increased to over £20,000,000, including some £13,000,000 in the jackpot, and the expected jackpot winnings for a single ticket increased to some £4,500,000. On the two midweek draws with a double rollover, sales were 49,000,000 and 64,000,000. At the average of these two figures, the total prize pool would be about £38,000,000, and the jackpot pool of about £20,000,000 would be expected to be shared among four winners. The mean payback in midweek rollovers increased from 45% to around 65%.

When the smaller midweek jackpot rolled over to the weekend, sales edged up only about 15% to some 58–60 million. The jackpot pool increased to about £12,500,000, thereby increasing the expected winnings/ticket to around £3,000,000, and the mean payback increased only to some 52%.

In the UK in early 1996, before the midweek draw was introduced, there were two double rollovers: sales at the first double rollover reached nearly £128 million, and £106 million at the second, at a time when normal sales were about £70 million. But even in these two double rollovers, the mean payback was about 65%; for single rollovers at that time, the payback was about 57%.

By 2007, overall sales were substantially less; typical midweek sales were below £20,000,000, typical weekend sales not much over £30,000,000, but the same phenomenon at rollovers occurred: the increased jackpot pool attracted sufficient extra sales to keep the overall mean payback under 70%. Attempting to buy all possible combinations to guarantee a jackpot share still made no sense.

3. TESTS OF RANDOMNESS

Whatever the lottery format, randomness has two components: the first is that, in any draw, all possible outcomes should have the same probability, and the second is that the outcomes of draws should be independent. It is built into the structure of most m/M lotteries that the number of possible outcomes is far larger than the number of outcomes that will be attained during the life span of the lottery game, so statistical testing of equal frequencies of all outcomes is a practical impossibility. We have to be far less ambitious. Similarly, testing for full independence of draws is not realistic. So we confine our attention to the independence of pairs of draws, or triples.

The most obvious requirement in a m/M lottery is to test for equality of individual numbers. Suppose that in D draws, the number k is selected $X(k)$ times. Under the null hypothesis of equal frequencies, the mean number of times each number is selected is $E = mD/M$; each draw selects m numbers without replacement, so, as

Stern and Cover (1989) pointed out, the usual χ^2 goodness-of-fit statistic needs a small modification to take account of that. It reduces to

$$W = \frac{M(M-1)(\sum X(k)^2 - m^2 D^2/M)}{(M-m)Dm}$$

Asymptotically, for large D , W has a χ^2 distribution on $M-1$ degrees of freedom, and the hypothesis of equal frequency is cast into doubt at large values of W .

This test is not particularly sensitive. For example, suppose one (unknown) number has probability $m/M + \varepsilon$ of selection in any draw, while the other $M-1$ numbers all have probability $m/M - \varepsilon/(M-1)$. Then the mean value of W can be calculated as

$$M-1 + \frac{\varepsilon^2 M^2 (D-1)}{m(M-m)}.$$

For a 6/49 lottery, the conventional 5% significance level asks that W should exceed 65: for $\varepsilon = 0.01$ —large enough to be a significant flaw—the mean value of W is about $48 + 0.00093D$, hence this mean value will not exceed the critical value until over 18,000 draws have been made! Detecting unequal frequencies among the numbers drawn is hard, unless there are substantial inequalities. (In a 6/49 lottery, suppose one number would actually occur at *double* its expected frequency: even then, it would take 120 draws for the mean value of W to exceed its 5% critical value.)

Joe (1993) extended this test to look for equal frequencies of pairs, and of triples, of numbers within draws. Here, account has to be taken not only of the nonreplacement of numbers within a draw, but also of the overlap between pairs and triples in the same draw [e.g., given that the pair (8,13) have both been selected, this increases the chance that (8,10) were both selected]. Let X_α denote how often the pair α are selected in D draws. Joe gave test statistics that again have distributions that are asymptotically χ^2 : to test for equality of frequencies of pairs within the same draw, use

$$W = \left[a \sum_{\alpha} (X_{\alpha} - E)^2 - b \sum_{m(\alpha,\beta)=1} (X_{\alpha} - E)(X_{\beta} - E) \right] / D$$

on $M(M-1)/(2-1)$ degrees of freedom. Here $E = m(m-1)D/(M(M-1))$, the second sum is over all pairs (α, β) that have exactly one member in common, and the constants (a, b) are given by

$$a = [(m-1)M - 5m + 7]C, \quad b = (m-2)C$$

where

$$C = \frac{M(M-1)(M-2)}{m(m-1)^2(M-m)(M-m-1)}.$$

However, the amount of data required to detect even modest association among pairs of numbers using this statistic is considerable. We omit Joe's corresponding formula for triples.

Genest et al. (2002) described a different way of using the usual goodness-of-fit statistic to test for equal frequencies of subsets of size c , for $1 \leq c \leq m$. Rather than modifying that statistic to give a quantity that would be asymptotically distributed as χ^2 , as above, they showed that its asymptotic distribution is the sum of c appropriately weighted independent χ^2 variables. For example, when $c = 2$, let X_α and E be as above, and define

$$W = \sum_{\alpha} \frac{(X_{\alpha} - E)^2}{E}.$$

Then the asymptotic distribution of W is that of the sum $a\chi_{M-1}^2 + b\chi_{M(M-3)/2}^2$, where $a = (m-1)(M-m)/(M-2)$ and $b = (M-m)(M-m-1)/[(M-2)(M-3)]$. For testing equal frequencies for pairs of numbers in a 6/49 lottery, this approach compares the actual values of W with those from a variable having the same distribution as the sum

$$\frac{215}{47}U + \frac{903}{1,081}V,$$

where U has the χ_{48}^2 distribution, and, independently, V is distributed as χ_{1127}^2 . For some alternatives, Joe's statistic is more powerful, for other alternatives, this version has the edge.

Joe (1993) also developed a test for independence of pairs of successive draws. Here let X_α denote how often the pair α has appeared in consecutive draws, for $\alpha = (1, 1), (1, 2), \dots, (M-1, M), (M, M)$. The test statistic is

$$W = \left[c \sum_{\alpha} (X_{\alpha} - E)^2 - d \sum_{m(\alpha, \beta)=1} (X_{\alpha} - E)(X_{\beta} - E) \right] / D$$

on $M^2 - 1$ degrees of freedom, with $E = m^2(D-1)/M^2$. The second sum is over those ordered pairs (α, β) in which either the first or the second (but not both) members of α and β are the same, and

$$c = (mM - 3m + 2)B, \quad d = (m-1)B$$

where

$$B = \frac{M^2(M-1)}{m^3(M-m)^2}.$$

Even though the amount of data needed to use these tests on the output of the actual draws may exceed the lifetime of the game, the tests could be applied both to dummy data that the lottery companies generate in testing their machines, and to the output of the Quick Pick generators for every game.

Johnson and Klotz (1993) took account of the order in which the numbers were drawn. They supposed that ball i had probability p_i of being selected first, and that the probabilities on second and subsequent draws were obtained by simply rescaling the remaining values. Their data on more than 200 draws gave a mild suggestion, via the log likelihood statistic, that the hypothesis of equal frequency was questionable ($p = 0.084$). They linked this with the observation that the balls are fed into the mixing machine in the same order each draw.

There are plainly many statistical tests that can be used to seek to identify particular ways in which the data may be nonrandom. Since the same data are used in each of these tests, they are far from independent of each other; and when (as often happens) the data are subjected to a large number of tests, the public may need to be reminded that genuinely random data will *fail* a 5% significance test 5% of the time! The tests used for the UK 6/49 Lottery may be representative: they are listed in Haigh (1997), and some of the test statistics, based on the seven numbers drawn (including the bonus number) are:

1. The sum S of the numbers drawn, so that $28 \leq S \leq 322$, and its null distribution is symmetrical about 175.
2. The number of *even* numbers selected, which should follow a hypergeometric distribution.
3. As well as the overall frequencies with which individual numbers are drawn, the frequencies with which those numbers are drawn in the n th position, $1 \leq n \leq 7$.
4. The maximum number of *contiguous values* among those drawn.
5. Define a success to be a draw in which at least one of the numbers selected also arose in the first draw; look at the gap between successes. Alternatively, instead of using the first draw as the permanent baseline, use the last draw that yielded a success.
6. Split the numbers into seven blocks of seven; then ask how many draws are needed until at least one (or two, or three) representatives of each block have been drawn.

For keno-type games, tests of randomness can follow this same pattern, while for powerball-type games, the separate sets of numbers drawn should be subject to similar tests—and for independence from each other.

Matters are easier with the numbers game format. In each position just one number is to be drawn, so the extra scaling factors needed in conventional χ^2 tests for equal frequencies are not needed. Moreover, with typically 10 possible digits in each position, and frequent draws, enough data for meaningful tests are quickly accumulated.

For a general numbers game format, with K positions, at each of which one of N digits is to be selected at random, suppose digit i has been selected $X(i)$ times

in D draws at a given position. Then the test statistic for equal frequencies at that position is

$$W = \frac{\sum (X(i) - D/N)^2}{D/N}$$

which, asymptotically, has a χ^2_{N-1} distribution under the null hypothesis, which would be thrown into doubt by high values of W .

For independence of the digits drawn at a pair of positions, let $Y(i, j)$ be the number of times digit i has been in the first position and digit j has been in the second. Then, under the null hypothesis that the digits are equally frequent, and there is independence between the two positions, the test statistic

$$W = \frac{\sum (Y(i, j) - D/N^2)^2}{D/N^2}$$

has a $\chi^2_{N^2-1}$ distribution.

Both these tests will, eventually, detect any departure from equal frequencies or from pairwise independence. If there is good reason to suspect a specific problem—for example, that particular digit i is appearing too seldom—then a test can be tailored to detect that. It ought not to be necessary to issue the standard reminder that it would be completely invalid to use the *same data* that suggested such a problem to make a statistical test for its occurrence: however, this point is not well appreciated by members of the public who may notice an apparent anomaly, and then find it hard to accept that data collection to test for the existence of that anomaly should begin afresh.

Lottery players can have a role in ensuring the integrity of a lottery by publicizing apparent anomalies, but the vast majority of such anomalies turn out to be false alarms. One reason is the great incentive by lottery corporations, who must retain the trust of the public, to keep their house in order. Another is that probabilistic intuition about what should, or should not, be expected is frequently misplaced. Here are two examples of calculations for a 6/49 lottery.

First, since any number has six chances in 49 to be drawn, the mean number of draws between consecutive appearances is $49/6 = 8.17$. The chance for any particular number not to appear in 50 consecutive draws is $(43/49)^{50} \approx 0.0014$; but the chance that *some* number makes no appearance in 50 given draws is about 7%, not at all small.

Second, bunching of numbers. Say that a bunch has occurred if at least four of the numbers drawn occur within six consecutive numbers; thus the winning combination (13 15 16 19 26 28) has no bunch, while (13 15 16 17 26 28) has two bunches (both the intervals (12,17) and (13,18), each with six numbers, contain four of the numbers drawn). Because each of the numbers 1 to 44 is a potential start point for a bunch to occur, it turns out that the mean number of bunches in D honest draws is about $D/23$, which is much more frequent than the layman anticipates. Just as surprising to many is

the fact that over 49% of the possible combinations in a 6/49 lottery contain a pair of consecutive numbers.

An incident reported by Chernoff (1981) illustrates this lack of intuition, and serves as a warning to consult the actual data, and not rely on press statements issued by the lottery corporations. A lottery representative asserted that, in the first 500 plays of the Massachusetts Numbers Game, there had been no exact repetitions of the winning numbers. As there are only 10,000 different four-digit numbers, the chance of *no* repetitions in 500 properly conducted draws (a variant on the birthday problem) is about 3×10^{-6} , so if the statement had been correct, there was a *prima facie* case of tampering with the outcomes! It turned out that the lottery representative had just assumed no duplications (on his own intuition), but that there had been several—close to the number expected.

4. GAMBLER CHOICES

The amount of information made public about what choices gamblers make varies considerably between lotteries. At one extreme, Riedwyl (1990) had access to the complete data for the draw in week six, 1990, of the Swiss 6/45 lottery. At the other extreme, in the UK 6/49 lottery, only snippets of information, such as the number of gamblers who tend to buy the combination {1 2 3 4 5 6}, and that {7 14 21 28 35 42} is the single most popular combination, have been generally released, although Simon (1997) was given access to more detailed data for one draw in 1996. For the Canadian 6/49 lottery, the relative frequencies with which gamblers select individual numbers are regularly published.

Many lotteries offer a Quick Pick = Easy Pick = Lucky Dip option, whereby a gambler allows a computer to use a built-in random number generator to make the selection. In the New Zealand 6/40 lottery, over 60% of tickets are bought using this facility, in the UK the figure is about 20%. In the Powerball game in the U.S., as many as 70% of tickets are sold using Easy Pick. Experiments show that, left to their own devices, humans asked to make random choices spread their selections much too *evenly* over the permitted range, so it is very likely that the actual number of tickets sold that genuinely are random selections differs little from the number who use Quick Pick.

The main points from Riedwyl's (1990) Swiss data are:

1. Some combinations are bought far more often than others. (A consequence, of course, is that other combinations are bought less frequently, and gamblers who can identify members of this latter category can expect any winnings to be above average.)
2. These popular combinations can be classified as
 - a. Previous winning combinations in the same lottery.
 - b. Winning combinations from other lotteries, in the same or neighboring countries.

- c. Combinations that make patterns on the lottery ticket—horizontal, vertical, or diagonal straight lines, zig-zag patterns, blocks, evenly spaced around the perimeter, or other systematic combinations.
- d. Small modifications of the above.

For example, although the 16,862,596 tickets bought in Riedwyl's dataset had an average of 2.07 for each of the 8,145,060 different combinations, over 5,000 of these combinations were bought at least 50 times! Two tickets whose numbers formed a diagonal straight line were bought over 24,000 times each, and the winning combination from the previous draw was bought over 12,000 times. The two combinations that differed from that last winning combination, either through adding, or through subtracting, unity from each number had 2,342 and 1,623 buyers, respectively. The winning combinations from the previous week's lotteries in France, Germany, and another Swiss Lotto all had over 1,400 buyers. Every winning combination in the same Lotto, for more than a year previously, was bought hundreds, sometimes thousands, of times. In all, 7% of total sales were spread among just 0.07% of the available combinations. In June 1999, draw 366 of the UK Lottery saw an astonishing 46 ticket-buyers share the jackpot, against the 3.5 or so expected from the sales figures; the winning combination {2, 17, 18, 23, 30, 40} was nothing remarkable—except that it had also arisen as the winning combination in the UK bookmakers' 49's game a few days earlier! As another illustration of unusually frequent selection, Morton (1990) reported that 14,697 punters had chosen the diagonal (8 15 22 29 36 43) in the New York State Lotto on June 7, 1986.

The Canadian information on the popularity of individual numbers is interesting, but the huge lack of independence between a gambler's choices makes it of very limited value in inferring the popularity of *combinations*. For example, it seems quite likely that particular combinations consisting of only the least popular numbers will be far more popular than average: as with the types of combinations described for the Swiss data above, buying these least popular numbers is a systematic way of constructing a combination, and so liable to be used by uncomfortably many gamblers. For what it is worth, the Canadian data for a 6/49 lottery (November 1995) show that the most popular individual numbers, in descending order of popularity, are 7, 11, 3, 9, 5, 27, 31, 8, and 17, while the least popular numbers, rarest first, are 40, 39, 48, 41, 20, 45, 38, 46 and 30. In rough summary, low odd numbers are most popular, high numbers, and round numbers, rather less so.

On the Australian 6/45 lottery, with the numbers displayed in rows of eight, Damien Broderick (personal communication) obtained similar data for a series of draws. The most popular choices were 19, 13, 20, 7, 11, and 27 (here 20 is a comparative surprise, perhaps explained by it being in the center of the ticket), and the least popular were 41, 42, 32, 44, 39, and 33. The frequencies were remarkably similar from draw to draw, with two caveats: the frequencies for the less popular midweek draws showed some significant differences from those in the weekend draws; and when sales were boosted by a rollover, the frequencies also altered. Overall, the constancy of choice outweighed these (small but noticeable) variations.

Information on the numbers selected by 17,001 people in one state for the Powerball draw on May 3, 1996 is given on the chance Website, maintained by Laurie Snell and others at Dartmouth College (<http://www.dartmouth.edu/~chance>). Five numbers were to be chosen from the list $\{1, 2, \dots, 45\}$: the most popular were 7, 9, 5, 3, 11, 12, 8, 4, and 10, while 37, 38, 43, 45, 39, 44, 41, 36, and 42 were least popular. Data on the full combinations of five numbers are also reported: aside from two people who each bought over 20 tickets with the same five numbers, but made different choices for the Powerball, the most popular combination was the 13-times choice of $\{3, 13, 23, 33, 43\}$, a combination making a diagonal line down the center of the ticket! It is plain that it is not only the numbers themselves, but also *where they fall* on the ticket, that influences punters.

A number of authors have attempted to use the data on the number of prize winners at the different levels, correlated with the winning combinations, to infer what choices gamblers are making. One of the earliest was Chernoff (1981), who observed that in the early days of the four-digit Massachusetts Numbers Game, zeros and nines were plainly unpopular, while other small digits were popular. At that time, the mean return was 60% of stakes. Chernoff considered the circumstances under which betting on unpopular numbers would give reasonable certainty of being ahead. Drawing on the Gambler's Ruin problem, he noted that betting on the 10% least popular numbers every day for a year, the advantage would need to be another 30% or so, on top of the 40% lost to the state's take, to be virtually certain of being ahead. He also noted that the phenomenon of regression to the mean would tend to depress expected winnings, if they are calculated on how much those numbers would have won in past draws.

Ziemba et al. (1986) used the data on the first 207 draws of the Canadian 6/49 lottery to infer gambler choice. Their basic model was that the mean return for a random ticket had the form $K \prod \vartheta_i$, the product being over the actual numbers drawn, and they estimated the values of the 49 parameters $\{\vartheta_i\}$. The results suggested that many combinations including the numbers 32, 29, 10, 30, and 40 would lead to a profit, on average, but that the combinations based on 5, 3, 13, 33, 28, and 7 would return only some 20% of the stake. (The standard deviations associated with these estimates are very large.)

The additive model proposed by Zaman and Marsaglia (1990) for the m/M lottery suggested that the probability of the combination t , $P(t)$, be given by

$$P(t) = \frac{1}{M^*} + K \sum_{i \in t} \left(\frac{\pi_i}{m} - \frac{1}{M} \right)$$

where $K = m/C_{m-1}^{M-2}$ and $M^* = C_m^M$. The values of $\{\pi_i\}$ would relate to the relative popularity of the individual numbers in the combination t . Attractive though this model may be at first sight, it suffers the deficiency that it demands a *negative* frequency for those combinations of m numbers that all have low popularity. This model demands a more even choice of numbers than is observed.

Zaman and Marsaglia (1990) also offered the multiplicative model

$$P(t) = \prod_{i \in t} \vartheta_i \quad (1)$$

that had been suggested by Stern and Cover (1989). On this model, Stern and Cover estimated the most popular choices in a 6/49 lottery were 3, 7, 9, 11, 25, and 27, while 20, 30, 39, 40, 41, and 48 were least popular.

Joe (1987) used ideas from *majorization* to estimate the frequencies of combinations; his work suggested that the most popular combination was $\{3\ 5\ 7\ 24\ 25\ 27\}$, about 14.5 times as popular as average. Later, Joe (1990) suggested another class of models leading to

$$P(t) = \left\{ \sum_{i \in t} \vartheta_i \right\}_+^{1/\alpha}, \quad \text{for some } \alpha > 0$$

which, in the limit as $\alpha \rightarrow 0$, reduces to the multiplicative model. Joe tested his model by comparing its predictions with the actual numbers of winners of prizes in the various categories; although these predictions were often several standard deviations out, they did not show systematic bias.

Finkelstein (1995) sought to estimate the frequencies of gambler choice of the individual numbers in a California 6/51 lottery. Let $W(d)$ be the winning combination in draw d , and let combination t be chosen $X(d, t)$ times. Then

$$\pi_j = \sum_{t: j \in t} E(X(d, t)) / \sum_t E(X(d, t))$$

is the frequency with which j is chosen by gamblers, supposed constant across draws. Hence, using the number of jackpot winners among D draws, we have

$$\pi_j^* = \frac{\sum_d X[d, W(d)] I[j \in W(d)]}{\sum_d X[d, W(d)]} \bigg/ \frac{\sum_d I[j \in W(d)]}{mD/M}$$

as an estimate of the individual popularity of j . For any finite D , these values do not necessarily sum to m , as they should, but Finkelstein proved the almost sure convergence of the estimate to its correct value (assuming, of course, the popularity of numbers is fixed). The rate of convergence depends on how many jackpot winners there are; in practice, convergence is very slow.

Finkelstein also used similar methods, based on the numbers of winners of the lesser prizes. Write

$$\pi_j^*(r) = \left\{ \frac{\sum_d \sum_{t \in M(W(d), r)} X(d, t) I[j \in W(d)]}{\sum_d \sum_{t \in M(W(d), r)} X(d, t)} \right\} \bigg/ \left\{ \frac{\sum_d I[j \in W(d)]}{mD/M} \right\}$$

where $M(t, r)$ denotes those combinations that have exactly r elements in common with t . He showed that

$$\pi_j^*(r) \rightarrow \frac{Mr - m^2}{m(M - m)} \pi_j + \frac{m - r}{M - m}$$

almost surely. For his data, based on 176 draws, all these estimates turned out to be compatible with a uniform choice of the M numbers.

He also offered an estimator based on the multiplicative model [Equation (1)]. In draw d , suppose $N(d)$ tickets are sold, there are $Y(d)$ winners of Match 3 prizes, and $p(d)$ is the probability, based on the model and the winning combination, that any ticket wins a Match 3 prize. The parameter estimates are chosen so as to minimize

$$\sum_d [Y(d) - N(d)p(d)]^2 / \{N(d)p(d)[1 - p(d)]\}.$$

Finkelstein's estimate of the order of popularity was

9 3 7 8 11 6.....51 43 49 48 46 50.

This sequence is plainly compatible with a birthdays model, in which the months 1 to 12 and the days 1 to 31 are much more popular than higher numbers.

Henze and Riedwyl (1998) built on Riedwyl's (1990) data to suggest a feasible way of selecting a combination that has a good chance of leading to a better than average prize (if it wins) in an m/M lottery. Their method is essentially one of seeking to eliminate large classes of combinations that can reasonably be taken as more popular than average, and then to choose at random from those that remain. To this end, for each combination compute

- (a) The sum of all the numbers.
- (b) The *edge number*, that is, how many of the numbers are on the edge of the ticket on which the gambler marks the selection.
- (c) The *cluster number*, that is, into how many clusters the numbers fall, if two that are adjacent (horizontally, vertically or diagonally) belong to the same cluster.
- (d) The *arithmetic complexity*, that is, count how many different numbers there are among the set of (positive) differences of pairs of numbers, and subtract $(m - 1)$.

The main ideas are that, for popular combinations, (a) tends to be low, (b) tends to be low, (c) is often unity or m , while (d) is also often low (e.g., geometric patterns are evenly spaced, and the sets of differences of pairs of numbers have many duplicates). Thus, for example, with a 6/49 lottery, they recommend selecting a combination at random, but then rejecting it unless:

- (i) The sum of the numbers is at least 177 (the mean and standard deviation of this sum, in a fair lottery, are 150 and 32.8, so this eliminates about 80% of all combinations, and ensures a bias toward higher numbers).
- (ii) The edge number is 3, 4, or 5.

- (iii) The cluster number is 2, 3, 4, or 5 (so eliminating combinations that are evenly and widely spread, which have a cluster number of 6; and the deliberate bunching of contiguous numbers, with a cluster number of unity).
- (iv) The arithmetic complexity is at least 8 (which eliminates many geometric patterns).

There are 1,521,650 combinations that remain, just over 10% of the entire list. (Any strategy based on eliminating swathes of combinations is doomed to become self-defeating if sufficiently many gamblers use it; it is thus important that the strategy retains a large number of combinations.)

Henze and Riedwyl tested the effectiveness of their suggestion by taking the data from two years' play in the UK and Florida 6/49 lotteries, and generating 10 sets of 5,000 tickets at random, according to this strategy. They then fictitiously added these 5,000 extra tickets to every draw for two years, and compared their average payoffs for the Match 4 and Match 5 prizes with the actual prize levels. In all cases, there was a significant increase in the average payoff (for Florida, about 20% on Match 4 and 35% on Match 5, even higher in the UK data).

Henze and Riedwyl also fitted a regression model for the dependent variable, $y = \ln(U)$, where U is the Match 5 payout, against dependent variables based on the criteria (1) to (4) above. Their model gave the best fit as

$$y = 3.4113 + 0.0238(d) + 0.7255S^* + 0.0771(b) - 0.0403(c)$$

where (d) is arithmetic complexity, S^* is the logarithm of the sum of all the numbers, (b) is the edge number, and (c) is the cluster number. The signs of the coefficients are consistent with their earlier recommendation. They noted that 25 of the 200 winning combinations in the Florida Lottery over 1993–1996 satisfied the conditions

1. The arithmetic complexity was at least five,
2. The cluster number was 2, 3, 4, or 5,
3. The predicted value of y was at least 7.5,

while 24 of them satisfied conditions (i) to (iv) above. The mean Match 5 prize for these 25 combinations was slightly higher even than the mean Match 5 prize from the 24 combinations that fit the earlier recommendations. Until lottery players' habits change substantially, it is clear that careful choice of combinations can lead to outcomes that are significantly better than average.

Suppose a gambler on a lottery with huge prizes is successful in identifying unpopular combinations, and is willing to make bets at times the prize funds are boosted by a rollover. MacLean et al. (1992; see also MacLean and Ziemba, 1999) considered two plausible scenarios. Case A has a medium-sized rollover, and the winning combination is fairly unpopular; Case B has a large rollover, and the winning combination is very unpopular. They use the Kelly criterion (Kelly, 1956) to find the optimal fraction of wealth to use to buy tickets, and reach gloomy conclusions.

Because the bulk of the expected winnings comes from an event with very low probability, the Kelly strategy advises a very low bet: in Case A, you need at least \$1,000,000

in capital to justify buying even one ticket! And if your aim is to have at least a 50% chance of turning your initial \$1,000,000 into \$10,000,000 before you lose half the initial capital, it will take on average some 22,000,000 years! With Case B, matters are *slightly* more optimistic: you can afford to purchase more tickets per draw, and if you are satisfied with a 95% chance of reaching \$10,000,000 before falling to \$25,000, the average time to wait is down to 2,500,000 years.

No lotto game has yet lasted 100 years without change.

References

- Bellhouse, D. R. 1982a. Fair Is Fair: New Rules for Canadian Lotteries. *Canadian Public Policy* 8, 311–320.
- Bellhouse, D. R. 1982b. The Need for a Federal Lotteries Review Board. *Canadian Journal of Statistics* 10, 213–217.
- Bellhouse, D. R. 1991. The Genoese Lottery. *Statistical Science* 6, 141–148.
- Chernoff, H. 1981. How to Beat the Massachusetts Number Game. *Math Intelligencer* 3, 166–172.
- Farebrother, R. W. 1999. A Brief History of the British National Lottery 1527–1826. *Chance* 12(1), 27–31.
- Farrell, L., R. Hartley, G. Lanot, and I. Walker. 1996. It Could Be You: Rollovers and the Demand for Lottery Tickets. Working Paper, Economics Department, University of Keele.
- Finkelstein, M. 1995. Estimating the Frequency Distribution of the Numbers Bet on in the Californian Lottery. *Appl. Math. Comput.* 69, 195–207.
- Finney, D. J. 1997. Contribution to Discussion. *Journal of the Royal Statistical Society* 160(2), 226.
- Genest, C., R. A. Lockhart, and M. A. Stephens. 2002. Chi-square and the Lottery. *The Statistician* 51(2), 243–257.
- Haigh, J. 1997. The Statistics of the National Lottery. *J. R. Statist. Soc.* 160(2), 187–206.
- Henze, N., and H. Riedwyl. *How to Win More: Strategies for Increasing a Lottery Win*. A. K. Peters, Natick, MA.
- Joe, H. 1987. An Ordering of Dependence for Distribution of k-tuples, with Application to Lotto Games. *Canadian Journal of Statistics* 15, 227–238.
- Joe, H. 1990. A Winning Strategy for Lotto Games? *Canadian Journal of Statistics* 18, 233–244.
- Joe, H. 1993. Tests of Uniformity for Sets of Lotto Numbers. *Statist. Probab. Lett.* 16, 181–188.
- Johnson, R., and J. Klotz. 1993. Estimating Hot Numbers and Testing the Uniformity of the Lottery. *Journal of the American Statistical Association* 88, 662–668.
- MacLean, L. C., W. T. Ziemba, and G. Blazenko. 1992. Growth versus Security in Dynamic Investment Analysis. *Management Science* 38, 1562–1585.
- MacLean, L. C., and W. T. Ziemba. 1999. Growth versus Security Tradeoffs in Dynamic Investment Analysis. *Annals of OR* 85, 193–225.
- Moore, P. G. 1997. The Development of the UK National Lottery 1992–1996. *J. R. Statist. Soc.* 160(2), 169–185.
- Morton, R. H. 1990. Lotto, Then and Now. *The New Zealand Statistician* 25(1), 15–21.
- Riedwyl, H. 1990. *Zahlenlotto: wie Man Mehr Gewinnt*. Haupt, Bern, Germany.
- Simon, J. 1997. An Analysis of the Distribution of Combinations Chosen by National Lottery Players. Working Paper, European University Institute, Florence.
- Stern, H., and T. M. Cover. 1990. Maximum Entropy and the Lottery. *Journal of the American Statistical Association* 84, 980–985.
- Zaman, A., and G. Marsaglia. 1990. Random Selection of Subsets with Specified Element Probabilities. *Commun. Statist. Theory Meth.* 19, 4419–4434.
- Ziemba, W. T., S. L. Brumelle, A. Gautier, and S. L. Schwartz. 1986. *Dr. Z's Lotto Guidebook*. Dr. Z Investments, Inc., San Luis Obispo, CA, USA and Vancouver, BC, Canada.

CHAPTER 24

U.S. Lotto Markets

Victor Matheson

Department of Economics, College of the Holy Cross, Worcester, MA, USA.

Kent Grote

Department of Business and Economics, Lake Forest College, Lake Forest, IL, USA.

1. Introduction	504
2. Differences Between American and European Lotteries	509
3. Fungibility of Lottery Revenues	512
4. Efficiency of Lottery Markets—Part 1	516
5. Efficiency of Lottery Markets—Part 2	519
6. Efficiency of Lottery Markets—Part 3	521
7. Conclusions	522
<i>References</i>	523

Abstract

Lotteries as sources of public funding are of particular interest because they combine elements of both public finance and gambling in an often controversial mix. Proponents of lotteries point to the popularity of such games and justify their use because of the voluntary nature of participation rather than the reliance on compulsory taxation. Whether lotteries are efficient or not can have the usual concerns related to public finance and providing support for public spending, but there are also concerns about the efficiency of the market for the lottery products as well, especially if the voluntary participants are not behaving rationally.

These concerns can be addressed through an examination of the U.S. experience with lotteries as sources of government revenues. State lotteries in the U.S. are compared to those in Europe to provide context on the use of such funding and the diversity of options available to public officials. While the efficiency of lotteries in raising funds for public programs can be addressed in a number of ways, one method is to consider whether the funds that are raised are supplementing other sources of funding or substituting for them. If lottery profits are fungible or substituting for other sources that would have been used in the absence of such profits, then the issues of equity and efficiency of lotteries relative to other sources are certainly heightened. The literature suggests that some degree of fungibility does exist, bringing these very concerns into question.

Whether the lottery markets are efficient can be addressed, in part, by examining the rationality of its participants. This can be done by considering how consumers participate in the market, how they respond to changing prices (or effective prices in the case of lotteries), and whether the market ever provides its participants with a “fair bet,” a gamble in which there is a positive expected value from participating. While empirical studies provide somewhat mixed results, there are indications that consumers of lottery products are relatively rational and that lottery markets seldom provide “fair bets,” both indicators of efficient markets.

1. INTRODUCTION

Lotteries have been commonplace in America from the earliest days of colonialism. Many public works, including Boston’s famous Faneuil Hall, as well as projects at illustrious universities such as Harvard and Princeton, were partly funded by lotteries, which remained popular throughout the country until the American Civil War. A nationwide backlash against gambling led to the decline of state-sponsored lotteries, however, and by the 1890s only Louisiana still operated a lottery game. Interestingly, as was seen again over a century later, the Louisiana Lottery Corporation’s monopoly on legalized gambling led to demand far outside the state’s borders with only 7% of the company’s revenues being generated within Louisiana (Louisiana Lottery Corporation, 2007). Allegations of corruption led to the collapse of the Louisiana Lottery in 1894 and left the U.S. without any state-sponsored games for 70 years.

In 1964, New Hampshire became the first state to reinstate a lottery game and other states soon followed suit. The first Canadian provinces restarted lotteries in 1970. By 2007, 42 states and the District of Columbia, as well as every Canadian province, sponsored lotteries. In the mid-1970s, state and provincial lottery associations began to join together to offer lotto games beginning with the formation of the Western Canada Lottery Corporation in 1974, the Tri-State Lotto, joining Maine, New Hampshire, and Vermont, in 1985, the Multistate Lottery Corporation (now more commonly known as Powerball) in 1988, and the Big Game/Mega-Millions Association in 1996 (Grote and Matheson, 2006a). Table 1 provides a list of every state lottery in the U.S. along with its year of initiation, the year that it joined a multistate lottery, as well as the multistate association it joined, the annual sales and profits of each lottery association, and the per capita sales of lottery tickets in each state.

The expansion of legalized gambling through state lotteries has proven popular for at least two reasons that will be explored in depth in this chapter. First, as more states legalized lottery games or other types of gambling, bordering states felt increasingly pressured to legalize lotteries within their own states. If gambling opportunities were widely available across state lines, a prohibition on gambling within the state may not result in a lower incidence of gambling, but could instead simply lead to gambling dollars being spent in neighboring jurisdictions. The potential loss of local revenues to lotteries or casinos in other nearby states has been a prime argument for legalizing and expanding gambling in the U.S.

Second, lottery associations typically designate all or a portion of the funds collected to “good works.” In the UK, for example, 40% of the sales price of each ticket is retained by the government with a significant percentage of this amount designated for the Department of Culture, Media, and Sport. In the U.S., more often than not, lottery funds are also designated for special purposes with education being the most common recipient of lottery proceeds. Thus, lottery tickets, like church bingo or other charitable gambling, may be perceived as a more conscientious choice by gamblers than privately run casinos or racetracks. Critics of lotteries, however, argue that all government revenues are fungible, and that by designating lottery proceeds toward education, for example, government officials simply find it easier to reduce other funding sources for education.

States typically offer a wide variety of gambling products through their lottery associations that can be placed in a variety of categories. The most popular lottery products in the U.S. are instant win scratchcard games. These lottery tickets sell for between \$1 and \$20 and allow gamblers to instantly win small to medium sized prizes. These games have the advantage of providing instant gratification (or despair) to players, but instant games cannot award large prizes without placing significant risk on the lottery association. For example, suppose a scratchcard game offers a single \$1 million prize to the lucky winner and suppose the lottery association distributes 2,000,000 \$1 tickets. On the surface it appears that this game will return a 50% payout to players and 50% to the lottery association. If players find out immediately whether they have won the grand prize, however, the lottery association will only be able to sell tickets to this game until the prize is won, which on average will occur at the 1,000,000th ticket. Thus, a game

TABLE 1 Descriptive Statistics for State Lottery Associations

State	Start date	Multistate	2006 Revenues (million dollars)	2006 Profit (million dollars)	2006 Per capita sales (dollars)
Alabama	None		—	—	—
Alaska	None		—	—	—
Arizona	1981	Powerball (1994)	468.70	141.12	76.01
Arkansas	ballot 2008		—		—
California	1985	MegaMillions (2005)	3,585.00	1,240.57	98.33
Colorado	1983	Powerball (2001)	468.80	125.60	98.62
Connecticut	1972	Powerball (1995)	970.33	284.87	276.86
Delaware	1975	Powerball (1991)	727.99	248.80	852.97
Florida	1988	None	4,030.00	1,230.00	222.78
Georgia	1993	Powerball (1995) MegaMillions (1996)	3,177.59	822.40	339.34
Hawaii	None		—	—	—
Idaho	1989	Powerball (1990)	131.13	33.00	89.42
Illinois	1974	MegaMillions (1996)	1,964.83	637.67	153.12
Indiana	1989	Powerball (1990)	816.40	218.00	129.31
Iowa	1985	Powerball (1988)	339.52	80.88	113.85
Kansas	1987	Powerball (1988)	236.05	67.09	85.40
Kentucky	1989	Powerball (1991)	742.30	204.30	176.48

Louisiana	1991	Powerball (1995)	332.12	118.76	77.46
Maine	1974	Tri-State Lotto (1985)	229.69	51.70	173.80
		Powerball (1990–1992)			
		Powerball (2004)			
Maryland	1973	MegaMillions (1996)	1,560.91	500.97	277.95
Massachusetts	1972	MegaMillions (1996)	4,534.12	951.24	704.36
Michigan	1972	MegaMillions (1996)	2,212.37	688.02	219.14
Minnesota	1990	Powerball (1992)	450.00	121.30	87.09
Mississippi	None		—	—	—
Missouri	1986	Powerball (1988)	913.52	260.67	156.35
Montana	1987	Powerball (1988)	39.92	9.11	42.26
Nebraska	1993	Powerball (1994)	113.11	30.32	63.96
Nevada	None		—	—	—
New Hampshire	1964	Tri-State Lotto (1985)	262.74	80.32	199.82
		Powerball (1996)			
New Jersey	1970	MegaMillions (1999)	2,406.57	849.25	275.84
New Mexico	1996	Powerball (1996)	154.71	36.86	79.15
New York	1967	MegaMillions (2002)	6,803.00	2,203.00	352.37
North Carolina	2006	Powerball (2006)	229.53	64.59	25.92
North Dakota	2004	Powerball (2004)	22.33	6.92	35.12
Ohio	1974	MegaMillions (2002)	2,221.00	646.30	193.50

(continued)

TABLE 1 (*continued*)

State	Start date	Multistate	2006 Revenues (million dollars)	2006 Profit (million dollars)	2006 Per capita sales (dollars)
Oklahoma	2005	Powerball (2006)	204.84	68.95	57.23
Oregon	1985	Powerball (1988)	1,104.00	483.00	298.32
Pennsylvania	1972	Powerball (2002)	3,070.00	975.85	246.77
Rhode Island	1974	Powerball (1988)	1,731.47	323.90	1,621.82
South Carolina	2002	Powerball (2002)	1,144.60	319.40	264.88
South Dakota	1987	Powerball (1990)	686.16	118.99	877.53
Tennessee	2004	Powerball (2004)	996.27	277.66	164.98
Texas	1992	MegaMillions (2003)	3,774.69	1,036.11	160.57
Utah	None	—	—	—	—
Vermont	1978	Tri-State Lotto (1985) Powerball (2003)	104.88	22.88	168.10
Virginia	1988	MegaMillions (1996)	1,365.00	454.90	178.60
Washington	1982	MegaMillions (2002)	477.89	116.95	74.72
Washington, DC	1982	Powerball (1988)	266.20	73.40	457.76
West Virginia	1986	Powerball (1988)	1,522.00	610.00	836.97
Wisconsin	1988	Powerball (1989)	508.90	150.60	91.59
Wyoming	None	—	—	—	—

Sources: National Association of State and Provincial Lotteries; Grote and Matheson (2007b).

that initially appears to have a 50% payoff to the lottery association will actually have zero net expected return to the seller. For this reason, instant win games generally award many modest prizes rather than a small number of larger prizes.

The other type of games are on-line or drawing games such as lotto, numbers, or keno. These games involve players selecting numbers from a set of possibilities. Players are issued a ticket with their choices, and these numbers are checked against numbers selected at a designated drawing. Players who match more of the numbers win increasingly larger prizes. Lotto games in particular have the interesting feature that when no player wins the grand prize by matching all of the numbers in a particular drawing, the money allocated to the jackpot pool is typically rolled over into the jackpot pool for the next drawing, raising the potential jackpot for the subsequent drawing. Because the jackpot prize fund is allowed to roll over in this manner, the jackpot prize can become quite large if no one hits the jackpot in a number of successive periods. Indeed, advertised jackpots exceeding \$50,000,000 are quite common in both the U.S. and Europe, and occasionally lotto jackpots have been known to exceed \$250 million.

In some states, on-line instant win games and video lottery are available. On-line instant win games are a hybrid of scratchcards and on-line games that provide the instant satisfaction of scratchcards with the ability to win larger prizes. Video lottery is simply a state-sponsored gaming machine more akin to slot machines or other casino gaming than traditional lottery games. The availability of video lottery explains at least some of the variation in state-by-state per capita lottery sales shown in Table 1.

2. DIFFERENCES BETWEEN AMERICAN AND EUROPEAN LOTTERIES

While in many aspects European and American lotteries tend to be quite similar, there are noticeable differences between the two continents. First, the share of tickets sales accruing to the government is typically larger in Europe than in the U.S. The UK National Lottery keeps 40% of ticket proceeds as government revenue and returns 50% as prize money with the remainder going to pay for retailer commissions and administrative costs. In the U.S., only Oregon and West Virginia exceed a 40% government take with the average association receiving only 28% of ticket sales. Two states, Rhode Island and South Dakota, retain less than 20% of revenues as profits. As administrative expenses and commissions are similar in the U.S. and the UK the portion of ticket sales designated to prize money is correspondingly higher in the U.S. It must be noted, however, that lottery winnings are subject to income taxes in the U.S. while they are exempt in Britain and Canada, at least, significantly reducing net returns in the U.S. and raising the government's share of the total ticket price.

Next, lotto jackpot prizes in Europe are paid in cash while lotto jackpots in the U.S. are paid in annuities usually over 20–30 years. The advertised prize in the U.S. is the undiscounted sum of the annuity payments. Lottery winners can choose to take their lottery winnings in a lump sum instead of the annuity payments, but the lump sum is typically 50–60% of the size of the advertised jackpot, depending on the length of

the annuity and the prevailing interest rates. Thus, while the large American multistate lotteries, Powerball and Mega-Millions, like to advertise that as of 2007 between the two games they have awarded the 15 largest jackpots in the history of gambling, in fact, at least three advertised jackpots in the EuroMillions lottery would rank among the five largest jackpots in history in terms of cash value rather than advertised value. (See Grote and Matheson, 2003, for an analysis of the effects of annuity payments on gambler behavior.) Combining the effects of annuities and the taxability of prize winnings, the net present after-tax value of the advertised jackpots of American lotteries tend to be roughly one-third the size of their advertised values.

The most popular European lotteries also tend to be much more egalitarian in their distribution of prizes than the most frequently played games in the U.S. In Europe, lower tier prizes are awarded larger shares of the prize pool and game matrices are set so that rollover jackpots are relatively less common. For example, the UK National Lottery sells tickets for £1, and players choose six numbers from a field of 49. Players who match three of the six numbers correctly win £10, the smallest prize that can be won. At least 11 state lottery games in the U.S. have offered an identical play matrix. In these games, the prize for a \$1 ticket for matching three of six numbers averaged roughly \$3.50 and ranged from \$0 to \$6, generally less than half that offered by the UK Lottery for its smallest prize. Similarly, while the UK Lottery and Euromillions each allocate 16–17% of every euro or pound wagered to the jackpot prize pool, a random survey of roughly 40 American lotto games finds the corresponding percentages allocated to the grand prize ranges from 19% to 43% of each dollar wagered with the average lottery providing slightly more than 30% of the funds collected to the jackpot, nearly double the percentage of the two European lotteries.

Finally, the jackpot prize pool tends not to roll over as much in European lotteries as compared to those in the U.S. Lottery associations face a trade-off in determining the optimal odds for a lotto game. By offering games with longer odds but bigger grand prizes, they could potentially attract more buyers. Numerous authors including Garrett and Sobel (1999, 2004) and Forrest et al. (2002) have suggested that lotto players are attracted by the high jackpots and not the expected return, and lotto is popular due to the “skewness” of the bet rather than its expected return. Lottery associations realize, however, that if the odds are too high, jackpots will be won very infrequently, and, therefore, the games will not benefit from frequent media exposure surrounding jackpot winners. Indeed, Britain’s Lotto Extra game was discontinued in 2006 after several long stretches without a winner (Forrest and Alagic, 2007). Lottery officials are, therefore, forced to choose between offering games with high jackpots and ones with frequent winners.

To this end, in the mid-1970s, state and provincial lottery associations began to join together to offer lotto games beginning with the formation of the Western Canada Lottery Corporation in 1974, the Tri-State Lotto, joining Maine, New Hampshire, and Vermont, in 1985, the Multistate Lottery Corporation (now more commonly known as Powerball) in 1988, and the Big Game/Mega-Millions Association in 1996. By merging games, states could offer larger jackpots, but the increased number of players would ensure that the grand prize was won on a regular basis.

Until the early 2000s, states with smaller populations generally offered lotto by being a member of one of the two major multistate games (Powerball and Big Game/Mega-Millions) while more populous states could offer high prizes through independent lotto games. For example, as of January 2000, eight states (California, Texas, New York, Florida, Pennsylvania, Ohio, Washington, and Colorado) operated lotto games but did not belong to a multistate game. Of these eight states, six ranked among the seven largest states by population. By the early 2000s, however, perhaps due to the record \$250 million advertised jackpots offered during several Powerball and Mega-Millions drawings, even these hold-out states began to join in the multistate associations so that by July 2005, only Florida remained independent from any multistate lotto game. Similarly, in 2004 the national lottery associations of the UK, France, Spain, and six other countries joined together to offer EuroMillions, which offers among the highest jackpots in Europe. See Table 1 for a list of state lotteries and the multistate lottery to which they belong.

Because of the larger number of ticket buyers, both the Mega-Millions and Powerball multistate games can offer substantially higher advertised jackpots than most state games. While the odds of winning these games are also lower than those of the state lotto games, there is not as much sacrifice in terms of the frequency of jackpot winners as there is in single state games. The relationship between population of potential ticket buyers and the structure of the game can be more precisely explained through an odds to population ratio.

Clotfelter and Cook (1993) note that the most frequent odds to population ratio for lotto games in the U.S. in the early 1990s was roughly one. That is, a lottery association serving a population base of 13,000,000 could offer a game with odds of roughly 1/13,000,000 and maintain a reasonable frequency of jackpot winners. The UK National Lottery, on the other hand, serves roughly 60,000,000 people with a game that offers odds of 1/14,000,000 for a 0.25 odds to population ratio. The EuroMillions game offers odds of 1/76,000,000 to a population base of just over 200,000,000 or a 0.38 ratio. The National Lottery ratio is less than half that of the lowest ratio reported by Clotfelter and Cook for state lotteries in 1990, and the EuroMillions lottery's ratio is less than one-third that of either of the two large multistate games in the U.S., Powerball (1/146,000,000 odds and 126,000,000 population for a ratio of 1.15) and MegaMillions (1/175,000,000 odds and 137,000,000 population for a ratio of 1.27).

Table 2 provides comparisons of two American lotto games in Florida and Texas, the two large U.S. multistate games, Mega-Millions and Powerball, as well as two European games, the UK National Lotto and EuroMillions. The time frames for each game are periods over which the prize structure in each game remained unchanged. Column 3 in Table 2 lists the average jackpot pool for each drawing of the games converted to net present value in the case of the American games and to dollar values using average annual exchange rates in the case of the European games. Column 4 lists the average number of times per year that the jackpot is won by at least one ticket in each game. Column 5 lists the average number of rollovers before the jackpot is finally won in each game. The final column lists the average number of winners when the jackpot is actually awarded.

TABLE 2 Jackpot Statistics

Game	Period	Average jackpot pool (million dollars)	Average number of times jackpot won per year	Average number of rollovers	Average number of winners
Florida	1/1/03–12/30/06	6.05*	28.75	2.65	1.27
Texas	5/7/03–4/22/06	14.72*	6.08	15.28	1.06
MegaMillions	5/17/02–12/29/06	32.56*	12.67	7.93	1.05
Powerball	10/9/02–8/27/05	27.50*	12.28	7.65	1.20
UK Lottery	11/19/04–12/29/07	10.77**	85.34	0.21	3.35
EuroMillions	2/13/04–12/28/07	31.28**	32.79	2.17	1.72

NOTE: *Value of cash option. **Pound and euro values converted to dollars.

As can be seen, the American games offer larger jackpots that are less frequently won than their European counterparts. Note that EuroMillions has weekly rather than biweekly drawings, as did the UK National Lottery for roughly its first two years of existence, so the figures in Column 4 actually understate the relative frequency at which American games are won in comparison to European games. The data also show that the UK lotto only infrequently rolls over, and EuroMillions rolls over at a rate less than one-third that of its big American counterparts. Even when a jackpot is won, it is much more likely to be shared among multiple winners in Europe than in the U.S.

3. FUNGIBILITY OF LOTTERY REVENUES

As stated previously, one possible reason for the popularity of state (and multistate, national, and multinational) lottery games is that the revenues from such games can be used to enhance funding of particular state programs. This earmarking of funds for a designated purpose appears to be important to both the successful passage of and the ongoing support for state lottery games. Of the 42 U.S. states (plus the District of Columbia) that provide lottery games, only 17 allow for the revenues from those games to go directly into that state's general fund.¹ Ten of those states earmark at least a portion of lottery revenues for a designated purpose. The remaining 25 states earmark all revenues from lottery games for specific government programs, with education being the primary beneficiary. Table 3 provides a more detailed summary of the legislated use of lottery revenues by state.²

¹Some states designate that excess lottery revenues will be available for general funds if a threshold level of revenues for earmarked spending is met.

²Note that "revenues" are more accurately designated as "profits" in Table 3 since it is assumed that administrative costs and prize money are already removed by the time the money is used to fund state expenditures.

TABLE 3 Use of Lottery Profits by State

State	Education	General fund	Environment / conservation	Development	Gambling treatment	Tax relief	Other
Arizona		x	x	x			x
California	x						
Colorado		x	x				
Connecticut		x					
Delaware		x					
Florida	x						
Georgia	x						
Idaho	x						x
Illinois	x						
Indiana						x	x
Iowa		x			x		x
Kansas		x		x			x
Kentucky		x					
Louisiana		x			x		
Maine		x	x				
Maryland		x					x
Massachusetts							x
Michigan	x						
Minnesota		x	x				
Missouri	x						
Montana		x					
Nebraska	x	x	x		x		
New Hampshire	x						
New Jersey	x						x
New Mexico	x						
New York	x						
North Carolina	x						
North Dakota		x					
Ohio	x						
Oklahoma	x						
Oregon	x		x	x			x
Pennsylvania							x
Rhode Island		x					
South Carolina	x						
South Dakota		x				x	x
Tennessee	x						

(continued)

TABLE 3 (continued)

State	Education	General fund	Environment / conservation	Development	Gambling treatment	Tax relief	Other
Texas	x						
Vermont	x						
Virginia	x						
Washington	x						
Washington, DC		x					
West Virginia	x		x				x
Wisconsin						x	
Total:	23	17	7	3	3	3	12

Sources: Novarro (2005) and the Web sites of state lottery associations.

A question that has arisen in the literature on lotteries as a source of state finances is whether these earmarked funds actually enhance spending dollar-for-dollar for the designated programs or if state legislators substitute earmarked dollars for dollars that would have come from the state's general funds had earmarking not occurred. The latter concept of substitution of state funds is referred to as fungibility, and the fungibility of funds can either be partial or total depending on the degree of substitution that occurs.

Several published studies have tested for the fungibility of government revenues from lotteries in U.S. states using different variables and statistical techniques, but most tend to agree that fungibility, at least to some degree, is present when funds are earmarked for specific state and local programs.

Mikesell and Zorn (1986) construct a time-series for government expenditures on education in a state as a percent of overall state and local government spending. They find that this percentage increases in only one of the three states examined after the introduction of earmarked funding from a state lottery game. In the other two states, there was actually a decrease in the percentage of funding to education immediately following the introduction of earmarked lottery revenues. While the authors note that this is not the best test of fungibility, since other factors may also be influencing the change in relative spending on education, it is an indicator that there was not a greater relative emphasis on education spending after the introduction of earmarked funding for that specified purpose.

Borg and Mason (1988, 1990) provided two studies of fungibility. The 1988 contribution considers the state of Illinois and its expenditures on education both before and after the introduction of a state lottery with profits earmarked for education. Using regression analysis and a Chow test, there is shown to be a statistically significant change in the trend for expenditures on education, with education expenditures rising at a lower rate after the introduction of the lottery, in spite of the lottery revenues available for such spending.

The 1990 contribution by Borg and Mason includes analysis of state expenditures on education in five states with lotteries that earmark profits for education and in seven states without lotteries. While there are mixed results for nominal spending on education in the five lottery states, real spending on education declines in all five of those states. Taken alone, this may indicate that fungibility of real spending on education is occurring; however, the seven non-lottery states also experienced a decline in real education spending over the same time period. Similar to Mikesell and Zorn (1986), this is not necessarily direct evidence of fungibility, but it certainly brings into question the commitment of funding to education after the introduction of lottery games that pledge the commitment of funds for that purpose.

Borg et al. (1991) performed a cross-sectional analysis of states to detect the impact of lottery funding on per-student expenditures on education. A dummy variable is included in the regression analysis to indicate if a state provides for earmarked funding to education via a state lottery. Their findings indicate that states with such funding have a statistically significantly lower level of spending per student, providing an indirect indication of fungibility.

Spindler (1995) tests for the fungibility of lottery revenues in seven states that earmark such revenues for educational programs. Using the ratio of education expenditures to general expenditures for each state as the dependent variable, Spindler constructs time-series ARIMA models to provide statistical evidence of fungibility in varying degrees in all seven states. Even more conclusively, however, there is evidence that the ratio of education to general expenditures actually declines significantly in four of those states after the introduction of a lottery game.

Three studies consider the impact of earmarked funding on state education expenditures in the state of Florida. Stark et al. (1993) provide evidence that there is not enough of an increase in per-student funding for education in Florida to account for the added state revenues from its lottery. They estimate that over 55% of the funds devoted to education from lottery revenues were, in fact, substituting for funding that would have come from the state if the lottery were not present. Summers et al. (1995) provide some support for the fungibility of education spending in Florida by considering the impact of lottery revenues on total allocations to community colleges in the state. They find that the combined allocation to community colleges from both the lottery and general funds from the state account for a smaller share of total funding sources available to community colleges after the state lottery began. Similarly, Land and Alsikafi (1999) find that there is a statistically significant decline in the growth rate of per-student (FTE) expenditures in community colleges in Florida after the introduction of the lottery. This is due, in large part, to the significant decline in per-student allocations from the state to community colleges in the post-lottery years. Part of this decline is due to a substantial increase in community college enrollments in the post-lottery years. However, the authors note that rather than providing additional funding to maintain current levels of per-student revenues to community colleges, the legislature opted to substitute lottery revenues for the necessary general funds. Garrett (2001) also focuses his empirical study of fungibility on a single state, Ohio, that like the state of Florida also earmarks its profits from the state lottery to education. Similar to the study by Spindler,

Garrett also uses an ARIMA model for his regression analysis, although real education expenditures per student are used as the dependent variable. Garrett also attempts to measure the degree of fungibility in lottery funding that occurs. His study finds that the earmarking of lottery funds in Ohio does not lead to a significant increase in per-student expenditures on education by the state, concluding that the funds are, to a large degree, fungible.

Erekson et al. (2002) conduct both a cross-sectional and time-series analysis of all 50 states over a five-year period to provide for a more complete study of fungibility. The models regress the expenditures on education as a percentage of general revenues for each state on a variety of theoretically important economic variables as well as a dummy variable for states that introduce a lottery and a variable for lottery revenues per capita. The estimation on the coefficient for lottery revenues per capita is negative and significant, indicating that fungibility does occur when lottery revenues are used to finance state expenditures, regardless of whether they are earmarked or not. Additional results indicate that for every \$1 per capita in lottery revenues generated as funding for a state, there is a loss of approximately 1–1.5% of education funding available.

Novarro (2005), similar to Erekson et al., also theorizes the importance of including both cross-sectional and time-series analysis to address the fungibility issue. The dependent variable used in the analysis is similar to previous studies, however, in that she utilizes state expenditures on education per student. She also uses lottery profits per-student as one of the independent variables in the model; however, she separates the effects of lottery profits depending on whether the profits are earmarked for educational purposes or are used as general funds by the state. By separating out the two types of earmarking, Novarro is able to conclude that while earmarking funds does indeed result in fungibility, earmarking provides relatively more revenues to a designated program than if the lottery revenues are not earmarked. Her model estimates that earmarked lottery profits for education tend to increase spending on education by approximately 79 cents for every \$1 in lottery profits, while \$1 in nonearmarked lottery profits tend to increase education spending by only 43 cents on average.

Given the statistical evidence, both direct and indirect, on the presence of fungibility of earmarked lottery revenues in these studies, it should bring into question the practice of earmarking lottery revenues if it merely allows for substitution of state dollars for legislative programs rather than supplementing those dollars. Obviously, the degree of fungibility that occurs is highly important, as is the issue of whether earmarking lottery revenues is relatively better than allowing state legislatures more discretion regarding their use.

4. EFFICIENCY OF LOTTERY MARKETS—PART 1

Since the price of a lotto ticket and the odds of winning remain fixed regardless of the size of the jackpot, the expected return from the purchase of a lottery ticket continuously changes along with the size of the jackpot. This varying return from a repeated game with fixed odds makes lotto almost unique among games of chance. Craps, slots,

roulette, bingo, keno, instant win lottery tickets, and lotto games without a rollover component all have fixed odds but also constant expected returns. Horse racing provides varying rates of return but is not a repeated game with fixed odds. Perhaps the only other similar gamble is blackjack when played by an expert card-counter where the game exhibits fixed payoffs but varying odds of winning depending upon which cards have already been played. The non-constant nature of the expected return of lotto has made the game the subject of extensive academic research and provides for interesting opportunities to explore the efficiency of betting markets and the rationality of gamblers.

Of course, some may question whether one can ever consider rational any gambling activity with a negative expected return. While this is a valid concern, gambling clearly offers non-pecuniary benefits to players in the form of thrills or excitement. In the words of one Big Game ticket buyer during the record \$363,000,000, May 2000 drawing, "One dollar is a small price to pay to be able to dream about winning \$300,000,000."

Accepting the idea of gambling itself as rational behavior, one may address more detailed concepts of rationality and market efficiency. At least three notions of rationality can be explored using lotto games. First, rationality requires that individual bettors choose the gamble with the highest expected return per dollar played. Second, as expected return rises, more bettors should enter into the market and existing bettors should gamble more. Finally, lotto games should never provide a positive expected return.

It is generally conceded that state lotteries have among the worst average expected payoffs among games of chance. While sports betting returns 91%, slot machines return 89%, bingo returns 74%, and blackjack returns 97%, state lotteries generally return only 50–60% of gross revenues to players in the form of prizes. Several theories explain the popularity of lottery tickets in the face of such low expected returns.

First, lottery tickets are an extremely convenient form of gambling. While horse racing and dog racing are offered at roughly 150 and 45 tracks around the U.S., respectively, and casino gambling is legal in about 1,200 American casinos (roughly two-thirds of which are in just five states: Nevada, Montana, California, Washington, and Oklahoma), lottery tickets are sold at over 150,000 retailers across the country. Furthermore, unlike casinos and racetracks, which are specialized gambling institutions, most lottery tickets are sold in gas stations and convenience stores and can be purchased along with other items.

Second, as noted previously, lottery associations more often than not designate proceeds to specific "good works" such as education or sport and recreation. Similarly, bingo, which is often offered by churches or other non-profit organizations, also offers a relatively low return.

Finally, the skewness of the bet and the high potential winnings offer one of the few gambling opportunities that present the possibility of a truly "life-changing" event. Few gamblers are likely to dream about what their life would be like if they won \$100 in their weekly local football pool, but thoughts of instantly becoming a multi-millionaire are another thing entirely. Indeed, the handful of lotteries known to return even less than half of revenues to prizes have offered very high jackpots. High maximum prizes tend to reduce the importance of expected value in lotteries.

Once the decision to play lotto over other games of chance is made, the question becomes whether or not bettors play the game in a way that reflects rationality in terms of maximizing expected return subject to the conditions of the game. The evidence of rationality on the part of lotto players is mixed, but tends to reflect at least some degree of rational decision-making on the part of lotto players.

Since the jackpot and often the lower tier prizes are paid in a pari-mutuel fashion in lotto games, players can increase their expected returns by playing “rare” numbers. On the reasonable assumption that every number combination is equally likely to be chosen, by selecting rarely played numbers, bettors can decrease the number of fellow players with whom they have to share the prize pool if they win. Most lotto games either allow a computer to randomly select numbers or allow players to choose their own numbers. When players select their own numbers, certain combinations, such as multiples of 7, birthdays, or vertical or diagonal columns on the play slip, are more commonly played than others.

For example, an examination of the first 801 drawings in the Texas Lotto shows that the average payout for choosing five out of six numbers correctly was \$1,656 and \$105 for choosing four of six correctly. However, in the six drawings where the smallest number drawn was 29 or higher, the average payouts were \$2,040 and \$141 respectively while in the 13 drawings where the highest number drawn was 28 or lower, the average payouts were \$922 and \$67 on average. Playing rare numbers, in this case numbers that did not correspond with dates, resulted in roughly a 25% increase in return above the average and over a 100% increase over the “common” numbers. Similarly, the January 14, 1995 drawing of the UK Lotto resulted in 133 grand prize winners, approximately 25 times the expected number, due to the selection of a set of numbers corresponding to an interesting pattern on the lotto play slip. The resulting jackpot prize of £122,510 per winner was the lowest in the history of the National Lottery and roughly 5% of the size of the typical grand prize.

The extent to which the distribution of numbers played deviates from a uniform distribution, and hence the ability that players have to earn above normal returns, is examined in depth elsewhere in this volume (Haigh, 2008) as well as by others (Farrell et al., 2000; Papachristou and Karamanis, 1998). As an approximation, however, since roughly 70% of all lotto tickets sold in the U.S. use computer generated numbers, which can be reasonably assumed to follow a uniform distribution, any supernormal expected returns are limited to the deviation from uniformity by the 30% of tickets that are sold to players who select their own numbers. Furthermore, as lotto jackpots grow, the percentage of players selecting their own numbers falls, further reducing any ability of players to select advantageous numbers during periods of high jackpots. Still, this phenomenon is a clear violation of rationality and has been examined by Clotfelter and Cook (1989), MacLean et al. (1992), Thaler and Ziemba (1988), and MacLean and Ziemba (1999), among others. The observed deviation in existing lotteries has been shown to frequently be large enough to allow some lotteries to provide positive net expected returns to bettors playing the rarest combinations. While mean returns exceeding \$2.00 per dollar played have been reported, due to the long odds involved, the player would have to play hundreds of thousands of draws before the strategy would, on average, have a

good chance of winning (Ziemba et al., 1986, MacLean et al., 1992, and MacLean and Ziemba, 1999).

5. EFFICIENCY OF LOTTERY MARKETS—PART 2

Another possible definition of rationality is that ticket sales will always increase when the expected return rises and will always fall when expected returns fall. An examination of the correlation between advertised jackpots and ticket sales shows a clear increase in ticket sales in response to higher expected returns as would be expected in efficient markets.

Violations of rationality that occur when ticket sales rise despite a decrease in the expected return can occur during rollovers when the number of ticket buyers rises at a faster rate than the advertised jackpot and have been named “Lottomania” or “Lotto Fever” by Beenstock and Haitovsky (2001) and Grote and Matheson (2004).

Testing whether lotto fever exists in actual lottery ticket markets requires an estimate of the expected return from the purchase of a lottery ticket. Several researchers have presented estimates of this expected return, starting with Clotfelter and Cook (1989) and including DeBoer (1990), Shapira and Venezia (1992), Gulley and Scott (1993) and Matheson (2001). Matheson (2001) presents the most detailed equation for the expected return, ER_t , from the purchase of a single lottery ticket.

$$ER_t = \left[\sum_i w_i V_{it} + (AV_{jt}/dvr_t) (1 - e^{-B_t w_j})/B_t \right] (1 - \theta) + \left(\sum_i w_i + w_j \right) \theta \tau \quad (1)$$

where w_i is the probability of winning lower-tier prize i , V_{it} is the cash value of lower-tier prize i at time t , w_j is the probability of winning the jackpot prize, AV_{jt} is the advertised jackpot prize at time t , dvr_t is a divisor used to convert the advertised annuitized jackpot into a net present value, B_t is the number of other ticket buyers for the drawing in period t , θ is the tax rate, and τ is the price of a ticket.

Lottery ticket sales almost always increase from drawing to drawing if the jackpot is not won, so rationality requires the expected return from the purchase of a lottery ticket to also be strictly increasing from drawing to drawing in order to explain the increasing ticket sales. This requires $ER_t > ER_{(t-1)}$ for all drawings within a jackpot cycle. Setting $ER_t > ER_{(t-1)}$ and canceling like terms, assuming that the conversion factor from the advertised jackpot to the net present value of the jackpot remains unchanged between drawings, leaves Equation (2).

$$V_{jt}(1 - e^{-B_t w_j})/B_t > V_{j(t-1)}(1 - e^{-B_{t-1} w_j}/B_{t-1}) \quad (2)$$

This arrangement is convenient because it eliminates problematic issues such as the appropriate tax rates to use as well as avoiding the problem of determining the size of the

lower-tier prizes when these prizes are determined in a pari-mutuel fashion. Equation (2) can be further rearranged to leave equation (3).

$$V_{jt}/V_{j(t-1)} > B_t(1 - e^{-B_{t-1}w_j})/B_{t-1}(1 - e^{B_t w_j}) \quad (3)$$

If equation (3) does not hold as the jackpot rises, then the purchase of a lottery ticket becomes an increasingly worse investment as the jackpot rises. In practice, however, Grote and Matheson (2004, 2005) have shown that lotto fever is exceedingly rare, occurring in only 12 cases out of over 23,000 American lottery drawings examined. Such instances are concentrated in record-size jackpots in large games and have become less common over time.

Violations of rationality that occur when ticket sales do not rise despite an increase in expected return are known in the literature as “lottery apathy” or “jackpot fatigue” and have been investigated by DeBoer (1990), and Grote and Matheson (2005, 2007a). It is an observed fact that lottery sales for most individual games have fallen over time. This decline is explained in part by the rise of recently legalized forms of non-lottery gambling or the introduction of new lottery products. For example, the expansion of casino gaming or the adoption of lotteries by neighboring states may have significant effects on lotto sales within a state.

The effects of casino gaming on lottery sales in the U.S. have not been well explored because of the difficulty in obtaining gaming revenue data from Native American casinos, which operate in roughly half of the states. The effects of neighboring lottery games have been well explored, however, as have the effects of the introduction of new games on existing games within a state. Researchers including Stover (1990) and Garrett and Marsh (2002) have clearly identified significant cross-border effects for lottery gambling. The expansion of lotteries to nearly every state has led to a decline in lotto play for states that had state lotteries previously as a decrease in cross-border play occurs. Some cross-border gambling still exists, particularly between states that are members of different multistate games. Advertised jackpots exceeding \$250 million are attractive lures for neighboring states. Border counties have been shown to experience disproportionately large increases in ticket sales during large multistate jackpots when the neighboring state is not a member of their particular multistate game (Oster, 2004).

Forrest et al. (2004) find that within-country competition between lotto games appears quite limited in the UK, but Grote and Matheson (2006a, 2007b) suggest a significant degree of cannibalism between games in American states that offer multiple lotto games. The degree of substitutability appears to be particularly high in states where two or more lotto games have similar characteristics in terms of average jackpots. Forrest et al. discovered no such evidence in the UK Lottery, attributing this result to the fact that the UK lottery association, Camelot, “has successfully designed and marketed games that each appeal to bettors in different ways.” Forrest et al. also find little evidence that the different lotto games in the UK are complements of one another, while Grote and Matheson find that while the presence of a multistate lottery game decreases sales overall for an existing state lotto game, during periods of large multistate jackpots, ticket sales for other lotto games within states that are members of the multistate lottery

association increase modestly as well. Grote and Matheson attribute this increase to a reduction in transaction costs.

A final anomaly identified in lotto sales is the halo effect, that is, an increase in lottery ticket sales in the periods immediately following a large jackpot being won. Various researchers have attributed this bump in sales to irrational bettors influenced by increased media attention surrounding the recent large jackpot. Grote and Matheson (2007), however, suggest that the anomaly may be explained as simply as bettors cashing in tickets winning smaller prizes and reinvesting the proceeds in new tickets.

6. EFFICIENCY OF LOTTERY MARKETS—PART 3

A final definition of rationality in lottery markets, first proposed by Scott and Gulley (1995), is that lottery games should never, or at least quite rarely, provide their participants with a bet with a positive expected value. Several papers have identified specific instances of fair bets in lotto drawings including Krautmann and Ciecka (1993) and Matheson (2001). Grote and Matheson (2005, 2006b) present the most ambitious tests of this definition of rationality by examining nearly 23,000 drawings of American lottery games. Using the expected return found in Equation (1), they find 290 instances where the purchase of a single randomly selected lottery ticket would have provided an after-tax expected return exceeding the cost of the ticket. The returns here exclude any additional money that could be earned by playing rare combinations as described previously. Examples of fair bets tend to be concentrated in smaller state lotteries that advertise relatively low jackpots but with substantially better odds of winning than the biggest state and multistate games. The smaller games do not attract as many additional ticket buyers when their jackpots become relatively large, and therefore the higher returns they offer are not as diluted by the prospect of potentially having to share the jackpot among multiple winners.

With less than 1.3% of drawings providing a positive expected return, it can reasonably be concluded that lottery games are generally efficient. Even those drawings providing positive returns subject the player to substantial risk, and only provide a fair bet if the player is assumed to be risk neutral. Investment strategies based on buying single tickets during draws with the “best” jackpots would only provide positive median returns with investment horizons that, literally, exceed 100,000 years in length.

Haigh (2008), supposes a gambler utilizes a strategy that considers both large jackpots and the playing of unpopular number combinations as examined by MacLean et al. (1992) and MacLean and Ziemba (1999). Under scenario A, the lotto game has a medium-sized rollover, and the winning combination is fairly unpopular, while under scenario B the game has a large rollover, and the winning combination is very unpopular. Because the bulk of the expected winnings in either case comes from an event with very low probability, in scenario A if an investor aims to have at least a 50% chance of turning an initial \$1,000,000 into \$10,000,000 before losing half the initial capital, it will take on average some 22,000,000 years for this to occur playing an optimal Kelly betting strategy. Scenario B offers only a slightly better investment opportunities. If an

investor is satisfied with a 95% chance of reaching \$10,000,000 before falling to \$25, the average time to wait is down to a “mere” 2,500,000 years. But Ziemba, in a private communication, notes that you could win on the first draw and that several very high, not-shared jackpots were won using unpopular numbers from the 1986 guidebook, but not by him (unfortunately).

Matheson (2001) and Grote and Matheson (2005, 2006b) note, however, that while the purchase of individual lottery tickets rarely provide a fair bet, the purchase of every number combination is much more likely to result in a positive net expected return at a significantly reduced level of risk. First, the purchase of every combination, denoted as the “Trump Ticket” by Krautmann and Ciecka (1993), guarantees the purchaser at least a portion of the jackpot, reducing the risk simply to how many other tickets have the winning combination as opposed to whether or not the jackpot is won in the first place. Second, the purchase of a Trump Ticket results in a higher jackpot payoff due to the large number of tickets purchased and the allocation of the proceeds to the jackpot pool. Third, the purchase of the Trump Ticket has certain tax advantages as described by Matheson (2001). See Ziemba et al. (1986) for an example of when it was optimal to buy the pot, as he calls it, in six consecutive draws. Sales fell despite the rollover increasing because buyers thought the prize could not be won.

Grote and Matheson (2005, 2006b) find that nearly 12% of the almost 23,000 drawings they examine would have provided a positive net return for the purchase of a Trump Ticket with many drawings providing an expected return in excess of 50%. The fact that few attempts to corner a lottery drawing have been attempted is likely due to two factors. First, even the purchase of a Trump Ticket may involve significant risk. While the Trump Ticket guarantees a share of the jackpot, it does not preclude other tickets from winning. In most of the cases identified by Grote and Matheson, the return from the Trump Ticket is only positive if no other tickets share the jackpot prize.

Furthermore, the act of physically purchasing the every possible combination for a particular lottery drawing is a daunting task. In fact, in February 1992, an Australian consortium attempted to corner a \$25,000,000 advertised jackpot in the Virginia Lotto. Despite a massive effort that included enlisting the aid of a major lottery ticket retailer, the consortium was only able to purchase 2,400,000 of the 7,059,052 possible combinations before time ran out. Luckily, they had the winning ticket and so made a profit. Cornering one of the larger games, such as Powerball, MegaMillions, or EuroMillions, would be even more difficult. Such a strategy is likely to be possible only for the smallest state games. However, with smaller games, while the rate of return might be high, the small size of the jackpot would limit the total return from such an effort.

7. CONCLUSIONS

Lottery games have considerable appeal as sources of public revenues. The diversity of products available as well as the adaptability of lotto structures, allow government officials to choose games that appeal to their constituents as well as provide for appropriate levels of public funding. However, as sources of public funding, the literature suggests

that fungibility of lottery revenues does exist, providing for lesser gains to public programs than might be expected. In fact, if the funds are completely fungible, programs designated as beneficiaries of lottery profits may receive just as much revenue after this designation as before.

The evidence on fungibility as an argument against the efficiency of state-run lotteries is both consistent and stronger than the arguments that the market for the lottery products are not efficient. Particularly in the instance of state-run lotto games in the U.S., consumers tend to exhibit rational behavior and the markets themselves do not tend to exhibit positive net expected returns on a general basis. However, individual violations of market efficiency do appear to occur in the form of positive expected returns from certain number combinations, the presence of lottery fatigue, and the potential positive expected returns from a Trump ticket.

References

- Beenstock, M., and Y. Haitovsky. 2001. Lottomania and Other Anomalies in the Market for Lotto, *Journal of Economic Psychology* 22(6), 721–724.
- Borg, M. O., and P. M. Mason. 1988. The Budgetary Incidence of a Lottery to Support Education, *National Tax Journal* 41, 77–85.
- Borg, M. O., and P. M. Mason. 1990. Earmarked Lottery Revenues: Positive Windfalls or Concealed Redistribution Mechanisms? *Journal of Education Finance* 15, 289–301.
- Borg, M. O., P. M. Mason, and S. L. Shapiro. 1991. *The Economic Consequences of State Lotteries*. Praeger, Westport, CT.
- Clotfelter, C. T., and P. J. Cook. 1989. *Selling Hope: State Lotteries in America*. Harvard University Press, Cambridge, MA.
- Clotfelter, C. T., and P. J. Cook. 1993. The Peculiar Scale Economies of Lotto, *American Economic Review* 83(3), 634–643.
- DeBoer, L. 1990. Lotto Sales Stagnation: Product Maturity or Small Jackpots? *Growth and Change* 21(1), 73–77.
- Erekson, O. H., K. M. DeShano, G. Platt, and A. L. Ziegert. 2002. Fungibility of Lottery Revenues and Support for Public Education, *Journal of Education Finance* 28(2), 301–312.
- Farrell, L., R. Hartley, G. Lanot, and I. Walker. 2000. The Demand for Lotto: The Role of Conscious Selection, *Journal of Business and Economic Statistics* 18(2), 228–241.
- Forrest, D., and D. Alagic. 2007. The History of a Lottery Game that Was Seldom Won, *The Journal of Gambling Business and Economics* 1(1), 57–68.
- Forrest, D., O. D. Gulley, and R. Simmons. 2004. Substitution Between Games in the UK National Lottery, *Applied Economics* 36(7), 645–651.
- Forrest, D., R. Simmons, and N. Chesters. 2002. Buying a Dream: Alternative Models of Demand for Lotto, *Economic Inquiry* 40(3), 485–496.
- Garrett, T. A. 2001. Earmarked Lottery Revenues for Education: A New Test of Fungibility, *Journal of Education Finance* 26(3), 219–238.
- Garrett, T. A., and T. L. Marsh. 2002. The Revenue Impacts of Cross-Border Lottery Shopping in the Presence of Spatial Autocorrelation, *Regional Science and Urban Economics* 32(4), 501–519.
- Garrett T. A., and R. S. Sobel. 1999. Gamblers Favor Skewness, Not Risk: Further Evidence From U.S.' Lottery Games, *Economics Letters* 63(1), 85–90.
- Garrett, T. A., and R. S. Sobel. 2004. State Lottery Revenue: The Importance of Game Characteristics, *Public Finance Review* 32(3), 313–330.
- Grote, K. R., and V. A. Matheson. 2003. Jacking up the Jackpot: Are Lotto Consumers Fooled by Annuity Payments? *Public Finance Review* 31(5), 550–567.
- Grote, K. R., and V. A. Matheson. 2004. Lotto Fever: Do Lottery Players Act Rationally Around Large Jackpots? *Economics Letters* 83(2), 233–237.

- Grote, K. R., and V. A. Matheson. 2005. Rationality and Efficiency in Lotto Games, in *Information Efficiency in Financial and Betting Markets*. Cambridge University Press, Cambridge UK, pp. 313–329.
- Grote, K. R., and V. A. Matheson. 2006a. Dueling Jackpots: Are Competing Lotto Games Complements or Substitutes? *Atlantic Economic Journal* 34(1), 85–100.
- Grote, K. R., and V. A. Matheson. 2006b. In Search of a Fair Bet in the Lottery, *Eastern Economic Journal* 32(4), 673–684.
- Grote, K. R., and V. A. Matheson. 2007a. Examining the “Halo Effect” in Lotto Games, *Applied Economics Letters* 14(4–6), 307–310.
- Grote, K. R., and V. A. Matheson. 2007b. Gamblers’ Love for Variety and Substitution among Lotto Games, *Journal of Gambling Business and Economics* 1(2), 85–99.
- Gulley, O. D., and F. A. Scott, Jr. 1993. The Demand for Wagering on State-Operated Lotto Games, *National Tax Journal* 46(1), 13–22.
- Haigh, J. 2008. The Statistics of Lotteries. *Handbook of Sports and Lottery Markets* by D. B. Hausch and W. T. Ziemba (eds.). North-Holland, Amsterdam, 481–502.
- Krautmann, A. C., and J. E. Ciecka. 1993. When Are State Lotteries a Good Bet? *Eastern Economic Journal* 19(2), 157–164.
- Land, V. Q., and M. H. Alsikafi. 1999. A Lottery’s Impact on Instructional and Noninstructional Expenditures and Unrestricted Revenues for Education, *Journal of Education Finance* 25(2), 149–174.
- Louisiana Lottery Corporation. 2007. Louisiana Lottery Corporation Web site. Available at <http://www.lalottery.com>.
- MacLean, L. C., W. T. Ziemba, and G. Blazenko. 1992. Growth Versus Security in Dynamic Investment Analysis, *Management Science* 38, 1562–1585.
- MacLean, L. C., and W. T. Ziemba. 1999. Growth Versus Security Tradeoffs in Dynamic Investment Analysis, *Annals of Operations Research* 85, 193–225.
- Matheson, V. A. 2001. When are State Lotteries a Good Bet (Revisited)? *Eastern Economic Journal* 27(1), 55–70.
- Mikesell, J. L., and C. K. Zorn. 1986. State Lotteries as Fiscal Savior or Fiscal Fraud: A Look at the Evidence, *Public Administration Review* 46, 311–320.
- Novarro, N. K. 2005. Earmarked Lottery Profits: A Good Bet for Education Finance? *Journal of Education Finance* 31(1), 23–44.
- Oster, E. 2004. Are All Lotteries Regressive? Evidence from the Powerball, *National Tax Journal* 57(2), 179–87.
- Papachristou, G., and D. Karamanis. 1998. Investigating Efficiency in Betting Markets: Evidence from the Greek 6/49 Lotto, *Journal of Banking and Finance* 22(12), 1597–1615.
- Scott, F. A., Jr., and O. D. Gulley. 1995. Testing for Efficiency in Lotto Markets, *Economic Inquiry* 33(2), 175–188.
- Shapira, Z., and I. Venezia. 1992. Size and Frequency of Prizes as Determinants of the Demand for Lotteries, *Organizational Behavior and Human Decision Processes* 52, 307–318.
- Spindler, C. J. 1995. The Lottery and Education: Robbing Peter to Pay Paul? *Public Budgeting and Finance* 15(3), 54–62.
- Stark, S. R., C. Wood, and D. S. Honeyman. 1993. The Florida Education Lottery: Its Use as a Substitute for Existing Funds and Its Effects on the Equity of School Funding, *Journal of Education Finance* 18, 231–242.
- Stover, M. E. 1990. Contiguous State Lotteries: Substitutes or Complements? *Journal of Policy Analysis and Management* 9(4), 565–568.
- Summers, S. R., D. S. Honeyman, J. L. Wattenbarger, and M. D. Miller. 1995. An Examination of Supplantation and Redistribution Effects of Lottery Allocations to a Community College System, *Journal of Education Finance* 21, 236–253.
- Thaler, R. H., and W. T. Ziemba. 1988. Parimutuel Betting Markets: Racetracks and Lotteries, *Journal of Economic Perspectives* 2(2), 161–174.
- Ziemba, W. T., S. L. Brumelle, A. Gautier, and S. L. Schwartz. 1986. *Dr Z’s Lotto Guidebook*. Dr Z Investments, Inc., San Luis Obispo, CA, USA and Vancouver, BC, Canada.

Subject Index

- 4/100 Games, 487
 5/M Lotteries, 487
 1963 Betting, Gaming and Lotteries Act, 372
 1986 Financial Services Act, 372
 2005 Gambling Act, 372
- A**
 Abnormal returns, 277, 285, 295, 301
 Advantageous bets, 368–372
 Adverse selection, 96, 409
 American football, 227–230, 370, 444
 point spread and odds betting in, 223–237
 American Lotteries, 510–512
 Anomaly, market, examining, 103–130
 combinatoric markets for testing models, 119–129
 Conlisk's model, 110
 definition of models and implications for combinatoric bets, 114–119
 markets controlled by bookmaker, 113
 mental accounting, 110
 perception-based models, 107, 111–112, *See also individual entry*
 preference-based model, 107, *See also Favorite-longshot bias*
 utility function, 114–116
 Anticipated utility specification, 37
 Arbitrage, 96, 120, 138, 142, 145, 148–150, 254, 258–261, 265, 278, 374, 406, 410–412, 429
 between betting pools, 279–282
 between parallel markets, 278–285, *See also Parallel markets*
 in team Jai Alai, 258–261
 Arbitrageurs, 148–150, 157
 ARCH test for heteroscedasticity, 473n24
 Arrow-Debreu market, 93
 Association football, *See Soccer*
 Asymmetric information, 113–114, 409
 Asymmetric probability weights, 39–42
- B**
 Backer, 405, 408
 Back price, 405, 406
 Bain, Hausch, and Ziemba (BHZ) approach, Kentucky Derby, 308–331
 Kelly betting model, 297, 300, 320–321
 Kentucky Derby, 1981–2007, 322–327
 win probabilities estimation, 314–320
 base-case model, 316
 dual-or-asterisk-qualifier model, 316–319
 dual-qualifier model, 316
 Barriers to arbitrage, 282, 284, 292
 Baseball, 139, 156, 226–227
 point spread and odds betting in, 223–237
 Basketball, 184, 211–212, 230, 236
 betting market, efficiency of, 211–212
 point spread and odds betting in, 223–237
 Behavioral theories, favorite-longshot bias, 107
 Belmont Stakes, *See Gramm and Ziemba (GZ) approach*
 Betdaq, 407
 Betfair, 394, 405–407, 412–414
 Bet flexibility, 416
 Bets, timing of, 99–100, *See also Timing*
 Betsson, 407
 Bettors, *See also individual entries*;
 Casual bettor; Gamblers;
 Representative bettor; Serious or regular bettor
 representative bettors, 108
 three-group heterogeneous bettor model, 157
 Betting
 betting markets efficiency, 138–139, 183–217, 242–249, 435, *See also Efficiency*
 exchanges, *See Exchanges*, betting systems, 288
 volume, 289–290
 Betting in running, 360
 Betting market efficiency (both strong form and weak form), 138–139, 183–217, 242–249, 435
 Betting on the spread, 358
 Betting systems, 288, 292, 331
 Betting tax, 18, 374, 425, 449
 Betting volume, 289–290
 Bettors, racetrack, preferences
 bettors off-track, 29
 bettors on-track, 29
 Chiappori, Gandhi, Salanié, and Salanié (CGSS) work, 42–45
 distortions of probabilities, 36–38
 empirical evidence on, 27–47

- exotic bets, 45–46
- expected utility representation, 37–38
- expected utility, 33–36
- facts, 30–33
- heterogeneous preferences, 42–45
- non-expected utility model, 37
- reference points and asymmetric probability weights, 39–42
- Bias, 83, 157, 161, 245, 247, 251, 409–412, 415–416
- Binary betting, 407
- Black Scholes model formula, 169–170
- Bond markets, 392
- Bookie, 241, 247
- Bookies margin, 378
- Bookmaker, 18, 32–33, 42, 85–86, 95–96, 113, 210, 227, 282–284, 404–409, 412, 414, 426, 438, 448–451, 453, 455, 486
- Bookmaker market, 278, 281–282, 286, 293, 302, 404, 406–410, 412, 422, 443
- Box bets, 144, 460, 485
- Breusch-Pagan test, 410
- Britain, soccer betting in, 421–444
 - development, 423–426
 - diffattend* variable, 442
 - Dobson Goddard presentation, 435
 - economic testing, 439
 - exact score betting, 425
 - fan bettors effect in, 441
 - fundamental analysis as an aid, 432–437
 - market efficiency, 429–431
 - modeling, limitations, 430
 - multivariate analysis, 438
 - ordered probit forecasting model, 434
 - ordered probit model, 434
 - Poisson modeling, 433
 - Pope-Peel model, 430, 435
 - Pope-Peel/Dobson-Goddard model, 438
 - regression test, 438
 - sentiment, 440–443
 - technical analysis, 437–440
 - through internet, 425
 - tipsters, 431–432
 - transactions costs
 - fall in, 426–428
 - in fixed odds market, 426–428
- Brownian motion model, 234–235
- Buy trade, 360
- C**
 - Call option, 361
 - Capital Gains Tax, 359
 - Casino gaming
 - pari-mutuel horse race wagering and, 5–7
 - Casual bettors, 142–144, 149, 151–152, 155–157, 289, 409
 - Casual vs. serious bettor behavior, 144–147, 156–157
 - Central paradox of racing, 78–79
 - Chefs
 - brilliant, 311–312
 - categories, 311–312
 - classic, 311–312
 - intermediate, 311–312
 - professional, 311–312
 - solid, 311–312
 - Chiappori, Gandhi, Salanié, and Salanié (CGSS) work, on racetrack bettors, 42–45
 - Closing bets, 360
 - Combinatoric bets, 114–119
 - combinatoric markets for testing models, 119–129
 - conditional independence, 123–125
 - fit of models, 123
 - pricing using conditional independence, 134
 - Commission, 209–210, 241, 310, 359, 406–408, 413–415, 509–510
 - Commodity trading, 358–359
 - Conditional independence, 118, 134–135
 - Conditional independence
 - combinatoric bets pricing using, 134
 - relaxing, 125–129
 - testing in combinatoric bets, 123–125
 - Conditional logistic regression, 411
 - Conscious selection, 462, 464, 470, 477
 - Consistency conditions, 366
 - Constant relative risk-aversion utility (CRRA), 34–36
 - Contracts
 - index, 387
 - simple contracts, 387
 - spread, 387
 - types of, 387
 - winner-take-all, 387
 - Cross-track betting, 184, 208–209, 278–279, 292
 - Cumulative prospective theory (CPT), 39–41
 - D**
 - Decision-making (under uncertainty), 104, 130, 277–278, 358
 - Decision markets, 397–398, 400
 - Design, lotteries, 459–477
 - ARCH test for heteroscedasticity, 473n24
 - econometric methodology, 469–474
 - expected value calculation, 465–467
 - game design parameters, 462
 - game design simulations, 474–476
 - GMM model, 473n24, 474
 - higher moments of the prize distribution, 467–469
 - odds of winning a (pari-mutuel) lottery, 464
 - OLS forecast, 473
 - possible winnings, determination, 462
 - RESET test, 473n24
 - rollovers in, 461–463
 - Diffattend* variable, 442
 - Distance preferences, 292
 - Distribution-based methods, 294–298
 - Distribution-free methods, 298–300
 - Dobson Goddard presentation, 435
 - Dosage breeding theory
 - for horse racing predictions, 307–331
 - Bain, Hausch, and Ziemba (BHZ) approach, 308–331
 - Chef-de-Race Listings, 333
 - dosage index, 311–314, *See also* Chefs
 - EFH listings, 334
 - Gramm and Ziemba (GZ) approach, 308–331
 - pedigrees, 333
 - performance measures, 311–314
 - preakness stakes, 1946–2006, 327–329
 - public's wagering, 333
 - racetrack as markets sequence, 310–311
 - semi-strong-form efficiency, 309
 - strong-form efficiency, 309
 - weak-form efficiency, 309
 - Dr. Z system, 60, 280, 411
 - Dutch book, 407

- E**
 Early gambling markets, 387
 Econometric methods on racetrack data
 advantage, 29
 drawback, 29
 Economic derivatives, 394
 Economic testing, soccer betting in
 Britain, 439
 Efficiency, 84, 120, 138–140, 148, 152,
 157, 183–184, 191, 224,
 242–249, 275, 330, 358, 409,
 411, 429–430, 433, 451, 504,
 516–517, 519, 521, 523
 Efficiency, betting markets, 183–217,
 See also Semi-strong form
 information efficiency
 basketball betting market, 211–212
 football betting market, 209–211
 in horse race betting markets,
 341–351, *See also* Professional
 tipsters role
 lotteries, 212–217
 of odds betting markets, 224–227, *See*
 also under Odds betting
 of point spread betting markets,
 227–230
 racetrack betting markets, 190–209,
 See also individual entry
 semi-strong efficiency, 194
 strong form tests, 185
 transactions costs, 186
 U.S. extent of gambling in, 187–189
 weak-form efficiency, 190–191, 196
 Efficiency, soccer betting odds, 447–456
 classical rational expectations test, 451
 data description, 450–451
 soccer betting market, 449–450
 statistical tests, 451–454
 Equity, 164, 174, 185, 389–390, 393,
 407–408, 504
 Equity markets, 166, 170, 390
 Euromillions, 510–512, 522
 European Lotteries, 510–512
 European vs. American lotteries, 509
 Exacta bets, 123, 128
 analysis, 53, 58
 F-L exacta, 128–129
 Exchanges, betting, 270, 276, 403–417,
 422, 425, 443
 Betdaq, 407
 Betfair, 405
 Betsson, 407
 cost-based explanation, 409
 degree of bias in, 412–416
 favorite-longshot bias in, 409
 Flutter, 407
 information-based model, 409
 internet innovation, 404
 Intrade, 407
 operation, 404–408
 ‘in running’ markets operation, 406
 Tradesports, 407
 WBX (World Bet Exchange), 407
 weak-form information efficiency in,
 409–412
 Exotic bets, 19, 107, 125, 130, 142, 207
 Exotics/Exotic bets, 45–46, 106,
 207–208
 Expected Value Formula, 479
 Explosion process, 294
- F**
 Fair bets, 442, 460, 504, 521
 Favorite, 90, 110, 113, 241–242, 250
 Favorite-longshot bias, 30–37, 42,
 83–100, 104–106, 112–113, 121,
 137, 140–150, 161–162, 278,
 289, 372–373, 404, 409–410,
 416, 425, 428, 437–439, 442,
 449, 455–456, 486, *See also*
 Reverse favorite-longshot bias
 advantages, 106
 arbitrageur, 148–150
 behavioral theories, 107
 Black Scholes model, 169–170
 explanation, 31–33
 information vs. risk in, 150–156
 quinella bets, 153–154
 information-perception theories, 141
 notation, 87
 objective vs. subjective probabilities,
 153–155
 objective vs. subjective win
 probabilities, 146
 opposite favorite-longshot bias,
 139–140, 157
 perception-based models, 107
 preference-based model, 107
 prospect theory, 107
 at racetrack, causes, 140–150
 regular favorite-longshot bias, 139
 representative bettor, 141
 risk-loving attitudes generating, 32, 34
 risk-preference theories, 141
 in S&P 500 and FTSE 100 index
 futures options, 161–179
 implied volatility skew effect,
 163–165
 soccer betting in Britain, 437
 standard neo-classical theory, 107
 theoretical explanations, 85–86
 heterogeneous beliefs, 86, 93–94
 limited arbitrage by informed
 bettors, 86, 96–97
 market power by informed bettors,
 86, 89–90
 market power by uninformed
 bookmakers, 86, 95–96
 misestimation of probabilities, 86,
 87–88
 preference for risk, 86, 90–92
 simultaneous betting by partially
 informed insiders, 86, 97–98
 theories, classes, 107
 time of day effects on, 147–148
 timing of bets, 99–100
 unifying with other market anomalies,
 137–158
 weekday vs. weekend betting,
 141–143
 Financial markets, 138, 184, 276–277,
 289, 386
 Financial Services Authority, 359, 372
 Financial Times Stock Exchange
 (FTSE), *See* FTSE 100
 Firm’s interval, 370
 Firm’s region, 371
 Fixed-odds Betting, 374, 424
 Fixed odds market, transactions costs in,
 426–428
 Fixed value approximation, Henery
 model, 56
Flutter, 407
 Football betting market, 209–211
 efficiency test, 210
 Football Pools, 423–424
 Forecasting, 184, 292, 343, 348, 374,
 392, 395, 397, 434
 Forecasting models, 292, 343–345,
 431–432, 434–437
 Foresight Exchange, 394
 Friedman-savage theory, 108
 FTSE 100 index market futures, 161–179
 monthly options, 175–178
 quarterly options, 174–175
 Full-card (whole-card) simulcasting, 12

Fungibility, 504, 514–516

Fungibility of lottery revenues, 512–516

ARIMA models, 515–516

Spindler tests, 515

G

Gambler choices, popularity of, 497, 499

Gamblers/Gambling markets, 103–130,

See also Anomaly, market

as risk-lovers, 109

skew-love of, 109

Gambling Act 2005, 372

Games, taxonomy, 186

Genoese lottery (m/M), 483, 486

GMM model, 473n24, 474

Gramm and Ziemba (GZ) approach,

Belmont Stakes, 308–331

Belmont Stakes, 1946–2006, 329–330

Grand National, 405

Greyhound racing, 187

Gross Profits Tax, 373–374, 408

H

Harville model, 52–53

Harville's formulas, 197–202, 208

Hausch, Ziemba, and Rubinstein (HZR) approach, 196–206

Hedging, 164, 187, 263, 360, 374, 408, 412

Henery fit, 198

Henery model, 53–54

approximated, empirical analysis for, 56–57

fixed value approximation, 56

varying value approximation, 56

Heterogeneous beliefs, 86, 93–94, 192

Heterogeneous preferences, racetrack bettors, 42–45

Hollywood stock exchange, 394–395

Home-away bias, 437, 439, 440

Hong Kong horse races

money bet in, modeling, 17–21

economy, 23

LRREBNREB, 19

LTPCPI, 18

pool availability, 23

race information, 23

variables examined, 18–19

weather, 23

Horse race betting, 190, 224–226,

276–278, 280, 286, 293–294,

300–302, 404–405

Horse race betting, markets, 224–226, 277, 293, 301, 342

semi-strong form information

efficiency in, 275–302

Horse race wagering, pari-mutuel, *See*

Pari-mutuel horse race wagering

Horse races, 68, 79, 169, 224, 226, 307, 518

money bet on, 17–21, *See also* Hong Kong horse races

predictions, dosage breeding theory for, 307–331

Horses

asterisk qualifiers, 309

categories, 309

dual qualifiers, 309

I

Implied volatility skew effect, 163–165, 168

In-play betting, 422

In running markets, 406

Income Tax, 359

Index betting, 358–361, 365–366, 368

Index betting, sports and stock indices, 357–374

advantageous bets, 368–372

biases, 372–374

correspondence with fixed odds, 364

equivalent spread bet, 373

fixed-odds bets, 373

gamblers participation, requirements, 359

nonlinear estimation technique, 372

operation, 360–361

regulating authority, 359

regulation, 372–374

setting spreads, 362–365

Shin's model, 372

spreads in performance indices, 365–367

taxation, 372–374

transactions, 359

Index combination bet, 359

Inefficiency, 245, 249, 285, 435, 442

Information, 23, 150, 275, 277, 286, 294, 314, 346, 409

Information asymmetries, 112

Information based model, 151–152, 155, 409–410

Information efficiency, 404, 409

Information levels, 410–411

Informational effects, favorite-longshot bias, 112–114

Information-based model, betting exchanges, 409

Information-perception theories, favorite-longshot bias, 141

Inside information, 372

Insider trading, 33, 372, 408–411, 416

Insiders, 86–87, 95–99, 113, 185, 282, 285, 410, 412, 416–417

Instant scratch-off games, 213

Internet

soccer betting in Britain and, 425

innovation, 404

Intrade, 407

Intra-state intertrack wagering (ITW), 11–12

Intuition, misleading, 495

Iowa Electronic Market (IEM), 391–394

Iraq war, 395–396

J

Jackpot fatigue, 520

Jackpots, 490, 509–512, 518–522

Joint Committee on the Draft Gambling Bill (2004), 407

Joint pricing of exotic and win bets, 107–108

K

Kelly betting model, 297, 300, 320–321

Kelly criterion, 201, 208, 279, 326, 369, 372, 381, 502

Keno, 484, 486–490, 494, 509, 517

Keno, comparisons across States, 48

Kentucky Derby, *See* Bain, Hausch, and Ziemba (BHZ) approach

L

Laboratory markets, 391–393

Lay opportunities, 412

Lay price, 406

Layer, 405–408

Liabilities, 361, 366, 405–406

Limited arbitrage, 120, 148, 258, 278

Limited arbitrage by informed bettors, 86, 96–97

Limit order, 360

Line, 241–243, 250

- Live racing, pari-mutuel horse race
 - wagering and, 8–11
- Lo and Bacon-Shone's discount model, 200
- London International Financial Futures Exchange (LIFFE), 166–167
- Lotteries, 212–217, 481–487, 489–490, 496, 500
 - brief history, 481–483
 - design, *See* Design, lotteries
 - design flaws, 482
 - efficiency of, 212–217
 - inefficiencies with unpopular numbers, 214–217
 - instant scratch-off games, 213
 - lottery betting markets, efficiency of, 183–217, *See also* Efficiency
 - lotto, 213
 - numbers game, 213
 - prenumbered games, 213
 - purpose of, 484
 - simple version, 213
 - statistics of, 481–502, *See also* Statistics of lotteries
 - types, 213
 - use of proceeds, 483
- Lotto, 213
 - lotto 6/49, 216
 - lotto markets, 503–523, *See also* U.S. lotto markets
- Lotto fever, 519–520
- M**
 - m/M(N) format, 483
 - Market anomaly, *See* Anomaly, market
 - Market biases, 157, 409–412, 415–416
 - Market design, 391
 - Market efficiency, 84–85, 105, 138, 140, 224, 230, 242–249, 289, 292–293, 330, 408, 416, 429–431
 - Market liquidity, 407
 - Market-makers, 359, 361–362, 364, 374
 - Market manipulation, 399, 408
 - Market power, 100
 - by informed bettors, 86, 89–90
 - by uninformed bookmakers, 86, 95–96
 - Mega-Millions, 505, 510
 - Mental accounting, in market anomaly, 110
 - Mid-event wagering, 234–236
 - Misestimation, 86–88
 - Misestimation of probabilities, 87–88
 - Misperceptions of probabilities, 31, 117
 - Modeling distance preference in thoroughbred turf racing, 67–80, *See also* Thoroughbred turf racing
 - Money bet on horse races in Hong Kong, modeling, 17–21, *See also* Hong Kong horse races
 - Multi-entry competitions, ordering probabilities
 - approximating by a simple method, 51–60
 - Multinomial logit model, 286, 294–299
 - 20-variable, 295
 - 70-variable, 296
 - Multinomial probit model, 297
 - Multiple variable models, 293–301
 - semi-strong form information efficiency, 293–301
 - 20-variable multinomial logit model, 295
 - 70-variable multinomial logit model, 296
 - distribution-based methods, 294–298
 - distribution-free methods, 298–300
 - Kelly betting strategy, 297, 300
 - multinomial logit model, 294–295
 - stochastic utility model, 294
 - support vector machine classification (SVC), 300
 - support vector machine regression (SVR) model, 299
 - Multistate lotteries, 510
 - N**
 - National Football League (NFL) games, 209–211
 - over/under-bets in, efficiency, 242–249
 - NBA, *See* U.S. professional basketball association
 - NBA National Basketball Association, 139, 230, 242, 245
 - Neo-classical theory, favorite-longshot bias, 107
 - Newsfutures, 394
 - Newspaper predictions, 285–287
 - NFL, *See* National Football League
 - NFL National Football League, 47, 139, 185, 209, 227, 231, 242
 - Normal approximation, 234–236
 - Normal distribution, 52, 231–233, 297, 326
 - Numbers Games, 213, 484–488
 - O**
 - Objective probability, 85, 87–88, 92–93, 144, 151–155, 169, 412
 - Odds, 29, 31, 34, 42–43, 45, 85–87, 105–108, 167–168, 173, 190, 205, 255, 277, 281–283, 285, 293, 302, 343, 405–415
 - Odds and probability, 224
 - Odds betting, 223–237
 - efficiency of, 224–227
 - baseball, 226–227
 - horse race betting, 224–226
 - normal model and mid-event wagering, 234–236
 - point spread and, 230–234
 - normal approximation, applications, 233–234
 - normal distribution result, 231–233
 - P-point favorites, 231
 - Odds classifications, 414
 - Offtrack betting sites (OTB's), 12–13
 - OLS forecast, 473–474
 - Online betting exchange
 - Tradesports.com, 397
 - Online sports betting, 449
 - Opening bets, 359, 373
 - Opposite favorite-longshot bias, 139–140, 144–145, 157
 - Ordered probit forecasting model
 - soccer betting in Britain, 434
 - Ordered probit model
 - soccer betting in Britain, 434
 - Ordering probabilities of multi-entry competitions
 - simple method approximating, 51–60, *See also* Harville model; Henery model; Stern model
 - empirical analysis for approximated Henery model, 56–57
 - exacta bet analysis, 53
 - Harville model, 52
 - quinella bet analysis, 53
 - simple approximation, 56–59
 - Stern model, 54
 - trifecta bet analysis, 53
 - Outlier price, 411
 - Over-round, 364–367, 372–373, 379–380, 407–409, 426, 429, 441

- Overtime, 242, 245
- Over/under, 240–242, 250
- Over/under-bets
 - NFL and NBA games, efficiency, 239–251
 - point spreads and, 240–249
- P**
- Pace character in thoroughbred turf racing, 67–80, *See also* Thoroughbred turf racing
- Pan-European electronic market, 447–456, *See also under* Efficiency
- Parlay, 207, 280, 410
- Parallel markets (bookmaker and pari-mutuel), 282, 285
- Parallel markets, arbitrage between, 278–285
 - arbitrage between betting pools, 279–282
 - cross-track betting, 278–279
 - parallel betting mediums, 282–285
 - place betting market, anomalies, 281
 - win probabilities derived, 281
- Pari-mutuel, 6
- Pari-mutuel betting, 358, 364
- Pari-mutuel horse race wagering, 84–85
 - from casino gaming, 5–7
 - competition for, 3–14
 - econometric model of, 7
 - effect at a racetrack, 10
 - demand model, 9–10
 - from live racing, 8–11
 - from professional sports, 8
 - from state lotteries, 7
 - from simulcast wagering, 11–13, *See also* Simulcast wagering
 - VLT's impact on, 6
 - from within and outside the industry, 3–14
- Pari-mutuel market, 35, 86, 99, 279, 282, 285, 301, 428
- Pari-mutuel prize structure, 487
- Payback rates, 257, 261, 488–491
- Pedigree, 291–293, 308, 310–312, 314
- Perception-based models,
 - favorite-longshot bias, 107
 - betting odds, 116
 - informational effects, 112–114
 - objective probabilities theories, 111
 - Potters and Wit model, 114
 - pricing functions for exactas for, 119
 - prospect theory, 111
 - subjective probabilities theories, 111
 - weighting of true probabilities, 111–112
- Performance index, 358, 365, 367, 378
- Performance indices, spreads in, 365–367
- Person-to-person betting, 373, 404
- Place pools, 195, 281
- Point spread, 240–242, 250
- Point spread betting, 139, 227–230, 223–237
 - efficiency of, 227–230
 - American football, 227–230
 - basketball, 230
 - normal model and mid-event wagering, 234–236
 - and odds betting, 230–234
 - normal approximation, applications, 233–234
 - normal distribution result, 231–233
 - P-point favorites, 231
- Poisson Distribution, 369, 433
- Poisson modeling, soccer betting in Britain, 433
- Pope-Peel model
 - soccer betting in Britain, 430, 435, 439
 - Pope-Peel/Dobson-Goddard model, 438
- Post Position, 290–292, 294
- Powerball, 483, 486–487, 490, 494, 498, 505, 510–512
- Powerball-type games, 494
- Powerball-type m/M(N) format, 487
- Preakness stakes, 1946–2006, 327–329
- Prediction markets, 110, 241, 385–401, *See also* Contracts
 - in business, 393–397
 - current uses, 393–397
 - decision markets, 397–398
 - economic derivatives, 394
 - equity markets, 390
 - first prediction markets, 388–390
 - foresight exchange, 394
 - gambling markets and, 386
 - hollywood stock exchange, 394–395
 - Iowa Electronic Market (IEM), 391–394
 - markets in the lab, 391–393
 - newsfutures, 394
 - in policy, 393–397
 - potential pitfalls, 398–400
 - RE model in a laboratory setting, 391
 - tradesports, 394
- Preference for risk, 86, 90–92
- Preference-based model,
 - favorite-longshot bias, 107, *See also* Gamblers
 - expected utility models with linear probabilities, 108–110
 - pricing functions for exactas for, 119
 - risk-loving gamblers, 109
- Prenumbered games, 213
- Presidential elections, 392
- Private information, 86, 95, 99, 113–114, 184, 276, 289, 344, 349, 427
- Prize structure, 482, 485–490, 511
- Probabilities, 35–38, 46, 96, 111, 124, 163, 280, 300, 320, 322, 327, 378, 390, 393, 400, 407, 409, 412–415
- Professional bettors, 289, 428, 438
- Professional Predictions, 285–288
 - in semi-strong form information efficiency, 285–288
 - newspaper predictions, 285–287
 - predictions available by subscription, 287–288
- Professional sports, competing pari-mutuel horse race wagering, 8
- Professional tipping services, 288
- Professional tipsters role
 - accuracy, 348–350
 - anti-herding, 350–351
 - data, 345–347
 - exaggeration, 342–343, 349
 - excess originality, 350–351
 - forecast originality, measuring, 347
 - frequency tests, 348
 - in horse race betting markets
 - efficiency, 341–351
 - originality, 348–350
 - private signals, 343
 - public information, 346–347
 - tips and rewards rules, 345–346
- Profitable betting strategies, 282, 292, 296, 298, 434, 439, 456
- Prohibition, 404, 417, 505
- Prospect theory, 39, 107, 111
- Publicly available information, 277–278, 294, 301–302, 331, 410
- Put option, 361

Q

Quarb, 374
 Quasi-arbitrage quarbs, 374
 Quasi-arbitrages, 374, 410
 Quinella bets, 123, 128
 analysis, 53

R

Racetrack betting markets, 190–209,
 310–311
 Blough's model, 193
 cross-track betting, 208–209
 exotic markets, 207–208
 Hausch, Ziemba, and Rubinstein
 (HZR) approach, 196
 implementing the system and
 empirical results, 203–206
 multinomial logit model, 194
 optimal capital growth, 201–203
 place and show markets, 195–196
 place and show probabilities, 196–201
 weak-form efficiency, 190–191, 196
 win market, 190–195
 Racetrack bettors, *See* Bettors, racetrack
 Racing, betting markets, efficiency of,
 183–217
 Racing Levy (UK), 408
 Randomness, tests of, 490–496
 Rank dependent expected utility
 (RDEU), 37–38
 Rational expectations (RE) model in a
 laboratory setting, 391
 Reference points, 39–42
 Regressors, 432
 Regulation, 372–374, 404, 424, 426, 429
 Representative bettor, 108, 110, 141
 RESET test, 473n24
 Reverse favorite-longshot bias, 104, 113
 Ridge regression, 75–76, 297
 Risk arbitrage in team Jai Alai, 261–269
 prediction models, 264
 single arbitrage, 263–267
 volatility, 264
 Risk preference model, 150–151, 154,
 409
 Risk preferences, 86, 90–92, 157, 163
 Risk-preference theories,
 favorite-longshot bias, 141
 Risk-love, 104, 109, 118, 123, 126, 134
 Risk vs. information models of bettor
 behavior, 150–156

Rollovers, 460–466, 470–471, 476–477,
 487, 490–491, 511–512, 519
 Royal Randwick Racecourse, 71–77
S
 S&P 500 index market futures, 161–179
 implied volatilities, 164
 monthly options, 175–178
 quarterly options, 174–175
 Sell trade, 360
 Semi-strong form information efficiency,
 194, 275, 277, 293, 301
 in horse race betting markets, 275–302
 arbitrage between parallel markets,
 278–285, *See also* Parallel
 markets
 betting systems, 288
 betting volume, 289–290
 distance preference, 292
 multiple variable models, 293–301,
 See also individual entry
 pedigree, 291–292
 post position, 290–291
 professional predictions, 285–288,
 See also individual entry
 professional tipping services, 288
 single variable models, 292–293
 Sentiment, 422, 439–443
 Sentiment, soccer betting in Britain,
 440–443
 Serious or regular bettor, 144–148, 157
 Setting intervals, 362
 Sha Tin Racecourse, 69–71
 Shin's z, 409–410, 416
 Show pools, 278, 280, 302, 411
 Silky Sullivan phenomenon, 200
 Simple method
 approximating ordering probabilities
 of multi-entry competitions,
 51–60
 Simple version, 213
 Simulcast wagering
 competing pari-mutuel horse race
 wagering, 11–13
 full-card simulcasting, 12
 live races at offtrack betting sites
 (OTB's), 12–13
 Simultaneous betting by partially
 informed insiders, 86, 97–98
 Single variable models, 277–293, 302
 Single variable semi-strong form
 efficiency studies, 292–293

Skewness, 462–463, 467–469, 473–476,
 510, 517
 Soccer, 46, 156, 184–185, 209–211, 224,
 227–228, 233, 372, 422–428,
 432–440, 442–444
 Soccer betting in Britain, 421–444,
 See also Britain
 Soccer betting odds, efficiency of,
 447–456, *See also under*
 Efficiency
 Sports betting, 104, 139, 156, 162,
 185–186, 189, 224, 240–242,
 359, 403, 407, 424, 449–450, 517
 Sports betting markets, efficiency of,
 183–217, *See also* Efficiency
 Sports index betting, 357–374, *See also*
 Index betting
 Spread betting, *See* Index betting
 Spread interval, 359, 360, 362, 366, 371,
 374
 Staking levels, 405
 Stamp Duty, 359
 State contingent claims, 276
 State lottery wagering
 competing pari-mutuel horse race
 wagering, 7
 Statistics of lotteries, 481–502
 4/100 games, 487
 5/M lotteries, 487
 6/49 lottery, 494–496
 gambler choices, 496–502
 Genoese lottery (m/M), 483–492
 goodness of-fit statistic needs,
 492–493
 Keno, 484
 m/M(N) format, 483
 Numbers Games, 484
 Powerball-type m/M(N) format, 487
 prize structure and winning chances,
 485–491
 tests of randomness, 491–496
 Statistical modeling, 436–437
 Stern model, 54, 58
 empirical analysis for, 57–59
 Stochastic utility model, 294
 Stock index betting, 357–374, *See also*
 Index betting
 Stop loss, 361
 Strong form tests, 185
 Strong law of large numbers, 368
 Subjective probability, 85, 151–153, 155,
 191, 295, 416

Support vector machine classification (SVC), 300
Support vector machine regression (SVR) model, 299

T

Take out rate, 428, 450–451, 461–462, 465, 468, 475, 479
Taxation, 372–374, 407, 504
Team Jai Alai, 253–271
arbitrage in, 258–261
risk arbitrage in, 261–269
Technical analysis
soccer betting in Britain, 437–440
Technological revolution in sports betting, 403–417
Thoroughbred turf racing
modeling distance preference and pace character in, 67–80
distance preference/gradient model, 77–78
jockeys, 79
pace, class, and time, 78–79
qualitative questions, 77–79
Royal Randwick Racecourse, 71
Sha Tin Racecourse, 69–71
wind resistance, 79
Timing, 85, 99
Timing of bets, 99–100
early, 99
late, 99–100

Tipsters, 287–288, 292–293, 342–351, 422, 431–432, 434, *See also* Professional tipsters role
Totalisator (Tote), 282–283, 412
TradeSports, 393–394, 396–397, 400, 407
Transactions costs, 163, 165, 185–187, 212, 291, 308–310, 318, 320, 373, 409–410, 416, 422, 425–428, 433, 436–437, 443, 521
in the fixed odds market, 426–428
sports betting, 186
Trifecta bets, 123
analysis, 53, 58
Turnover Tax, 373

U

UK National Lottery, 483, 511–512
Underdog, 210, 241, 249–250, 455
Unified theory of betting, 140, 157
Unpopular combinations, 501
U.S. lotto markets, 503–523
American vs. European lotteries, 509–512
efficiency, 516–521
rationality in, 516–522
revenues, fungibility, 512–516
ARIMA models, 515–516
Spindler tests, 515
state lotteries, 505–508
U.S. professional basketball association (NBA)
over/under-bets in, efficiency, 242–249

U.S. extent of gambling in, 187–189
Utility of gambling, 33, 42, 109

V

Varying value approximation, Henery model, 56
Video lottery terminals (VLT's), 5
Vigorish, 241–242
Volatility bias, 359
Volatility smiles, 163–165

W

Wagering on pari-mutuel horse racing, *See* Pari-mutuel horse race wagering
Walrasian equilibrium, 93
WBX (World Bet Exchange), 407
Weak form inefficiencies, 196, 411
Weak-form information efficiency in betting exchanges, 409–412
Welfare gains, 416–417
Winning chances, 483, 485–491
Win pools, 119, 225, 278, 280
World Bet Exchange, 407

Z

Ziembra, Brumelle, Gautier, and Schwartz (ZBGS), 212–214

Author Index

A

Abbate, G., 189
 Abraham, F. J., 240
 Ahlers, D., 140
 Aitchison, J., 19
 Akst, D., 185n1, 189
 Alagic, D., 510
 Albright, C., 211n11
 Algeot, P., 201n7, 320
 Ali, M. M., 5, 6, 7, 8, 9, 9n9, 10, 10n10,
 11, 12, 13, 18, 31n13, 34, 52, 86, 93,
 94, 105, 109, 110, 115n21–22, 117,
 120, 121, 125, 138, 140, 145n7, 162,
 187, 192, 193, 207, 280, 295, 297,
 309, 317n8, 410
 Allais, M., 36, 38
 Alsikafi, M. H., 515
 Amit, R., 404
 Anderson, R. A., 271
 Ariel, R., 185
 Arrow, K. J., 28
 Asch, P., 29n10, 32, 110, 114, 120, 125,
 139, 193, 196, 207, 208, 280, 295,
 309, 317n8
 Asimakopoulos, I., 431, 434, 435, 436
 Athanasoulis, S., 386n1
 Aucamp, D., 201
 Avery, C., 442
 Ayton, P., 46

B

Bacon-Shone, J. H., 52, 53, 58, 59, 60,
 200, 206, 207, 208
 Badarinathi, R., 139
 Baesel, J., 271
 Bain, R., 205, 292, 293, 307, 308, 315,
 317, 319, 320, 322, 325

Bakshi, G., 166
 Barone-Adesi, G., 167
 Bassett, G. W., 111n15
 Beenstock, M., 519
 Beetsma, R., 47
 Bellhouse, D. R., 483
 Benter, W., 75, 185n1, 194, 198, 200,
 209n9, 292, 295, 296
 Berg, J., 391, 392, 397
 Bernhardt, D., 342
 Betton, S., 195, 291
 Beyer, A., 185n1, 195, 290
 Bird, R., 109, 284, 286, 295
 Black, F., 166, 167, 168, 169, 169n10,
 170
 Blackburn, P., 283, 284
 Blough, S. R., 94, 193
 Bohara, A. K., 9
 Bolger, F., 301
 Bollen, N. P. B., 164, 165, 174, 178, 179
 Bolton, R. N., 194, 290, 294, 295, 296,
 297, 298, 299, 300
 Borel, E., 85, 85n4, 89
 Borg, M. O., 514, 515
 Borghesi, R., 444
 Bradley, I., 39, 110
 Breiman, L., 201, 320
 Bremer, M. A., 140
 Brown, K. C., 140
 Brown, K. H., 240
 Brown, W. O., 105n4, 212
 Bruce, A. C., 277, 301, 302, 283, 284,
 289, 295, 296, 412
 Brumelle, S., 212
 Brunk, G. G., 108n10
 Buraschi, A., 164
 Busche, K., 18, 30n12, 52, 104n3, 139,
 162n1, 191, 289

C

Cadsby, C., 185
 Cain, M., 104n3, 283, 285, 372, 425,
 437, 439, 449
 Camerer, C. F., 28n1, 37, 107n7, 110,
 111, 121, 149, 211, 212, 289,
 398, 399
 Canfield, B. R., 195, 290, 291
 Carlton, D. W., 408
 Chadha, S., 90
 Chan, K. C., 140
 Chapman, R. G., 194, 290, 292, 294,
 295, 296, 297, 298, 299, 300
 Chen, K.-Y., 394, 395
 Chen, Q., 342
 Chernoff, H., 214, 485, 486, 496,
 498
 Chevalier, J., 442
 Chew, P., 334
 Chiappori, P.-A., 30, 42
 Chiu, C., 277
 Chopra, N., 185
 Christiansen, E. M., 4n2
 Church, A. M., 9
 Ciecka, J. E., 521, 522
 Clotfelter, C. T., 187, 212, 214, 215n17,
 216, 465n4, 511, 518, 519
 Coate, D., 8, 12
 Cochran, W., 439
 Coleman, L., 119n26, 410
 Coles, S., 433
 Conlisk, J., 109, 118n25, 187
 Constantinides, G. M., 170
 Cook, P. J., 187, 212, 214, 215n17, 216,
 465n4, 511, 518, 519
 Copeland, T. E., 95n15
 Cotton, M., 290
 Cournot, A. A., 99

Coval, J. D., 164, 164n2, 165, 169n9,
170, 174, 174n13, 178
Cover, T. M., 201n7, 215, 320, 492, 499
Crafts, N. F. R., 32, 104n2, 112n17, 288,
426, 433

D

Dana, J. D., 185, 211
Dansie, B. R., 52, 197
Danthine, J. P., 271
De Bondt, W. F. M., 140
DeBoer, L., 519, 520
Deschamps, B., 288, 342, 372, 425,
431, 439
Dixon, M. J., 363, 369, 429, 433,
437, 448
Dobson, S., 424, 427, 431, 434, 438
Dolbear, E., 208
Donkers, B., 28n2
Dowie, J., 104n2, 295
Dumas, B., 165, 174, 178

E

Easton, S., 444
Edelman, D. C., 71, 75, 281, 292, 297,
298, 299, 300, 301
Eisenberg, E., 94n14
Elton, E., 157
Epstein, R. A., 190, 255
Erekson, O. H., 516
Ethier, S. N., 369

F

Fabricand, B., 191
Fama, E. F., 106, 184, 309
Farebrother, R. W., 482
Farmer, A., 96n19, 112, 428
Farrell, L., 469n9, 490, 518
Fauman, B. C., 291
Feiger, G., 271
Ferrell, W. R., 301
Février, P., 47
Figlewski, S., 105n4, 139, 165, 174,
178, 179, 194, 276, 286
Finkelstein, M., 368, 499
Finney, D. J., 484
Fishburn, P. C., 109
Forrest, D., 423, 424, 424n1, 431,
434, 436, 439, 441, 441n14,
443, 443n17, 449, 452, 453,
510, 520
Forsythe, R., 391
Friedman, J. H., 254n1
Friedman, M., 40, 107, 108, 110,
141, 187

G

Gabriel, P. E., 107, 156, 282, 283, 284
Galai, D., 95n15
Gale, D., 94n14
Gander, J. M., 104n3, 105n4, 112n17,
139, 145, 210, 295, 426, 437n8
Gandhi, A. K., 30, 42, 43, 119, 120
Garrett, T. A., 189, 467n5, 510, 515, 520
Gautier, A., 212
Genest, C., 493
Gergaud, O., 288, 342, 372, 425, 431,
439
Gigerenzer, G., 301
Gil, R., 444
Gilovich, T., 211
Gilpatric, S., 348
Glosten, L., 95n15
Goddard, J., 424, 427, 431, 434, 435,
436, 438
Golec, J., 34, 34n18, 35, 36, 91n12,
105n4, 109, 119n26, 140, 210, 240,
437, 438, 452, 467n5
Goodfriend, M. J., 254n1
Gorman, W. M., 468n7
Graham, I., 434, 436
Gramm, M., 139, 205, 292, 307, 308
Grant, D., 271
Gray, P. K., 104n3
Gray, S. F., 104n3
Greene, W. H., 452, 430
Greimel, T., 392
Griffith, R. M., 30, 31, 33, 37, 38, 39,
84, 86, 87, 104, 104n1, 153, 162,
162n1, 191, 317n8
Grossman, S., 409
Grote, K. R., 505, 510, 519, 520,
521, 522
Gu, M. G., 297
Guiso, L., 28n7
Gul, F., 46
Gulley, O. D., 519, 521
Gürkaynak, R., 396, 397
Guryan, J., 470n15

H

Hahn, R., 397
Haigh, J., 362, 364, 368, 494, 518, 521
Haitovsky, Y., 519
Hakansson, N. H., 201
Hall, C. D., 18, 30n12, 52, 104n3, 139,
162n1, 191
Hamid, S. S., 109n13
Hanson, R., 149, 395, 397, 398
Hartley, R., 462n1, 464n3, 477
Harvey, C., 179

Harville, D. A., 52, 53, 58, 59, 85,
118n24, 124, 196, 198, 201, 208
Hausch, D. B., 28n5, 30n11, 53, 60, 105,
115, 118n24, 120, 139, 162, 162n1,
163, 174, 183, 187, 190, 191, 195,
196, 196n5, 202, 203, 204, 205, 208,
225, 270, 278, 279, 280, 282, 293,
295, 307, 308, 309, 309n1, 311,
317n8, 321, 321n9, 411, 456, 462n1
Hausman, J., 468n7
Hawawini, G., 184
Henery, R. J., 38, 52, 53, 54, 56, 58, 59,
112, 117, 197, 198, 200, 208, 297
Hensel, C., 185
Henze, N., 500, 501
Herron, M., 392
Hodges, S., 30
Hollander, Z., 254
Hong, Y., 277
Howe, J. S., 140
Hurley, W., 86, 90n11, 96, 112, 113n19,
117, 149, 162n1, 163, 409

I

Isaacs, R., 85n4, 86, 89, 89n9, 99,
193, 294

J

Jackson, D. A., 281, 359
Jackwerth, J. C., 164, 167, 174
Jacquillat, B., 271
Jiang, W., 342
Joe, H., 215n17, 492, 493, 499
Johnson, R., 214n16
Johnson, J. E. V., 277, 283, 284, 289,
291, 295, 296, 300, 301, 302, 372, 412
Johnson, J., 184
Johnson, R. S., 139, 494
Jones, P., 404, 407, 408
Jullien, B., 30, 32, 37, 38, 40, 41, 42,
43, 85, 111, 111n16, 115n21, 119,
120, 372

K

Kahneman, D., 39, 46, 107, 111, 163,
174, 193
Kallberg, J., 202n8
Kallio, M., 254
Kanto, A. J., 34, 208
Karamanis, D., 518
Kearney, M. S., 467n5, 470n15
Keim, D. B., 184, 309
Keller, J. B., 68, 201, 261, 278, 320,
368, 371
Kendall, M., 184
Kida, T., 301

Author Index

535

Kim, C., 404
Kirzner, I. M., 149
Kleinfeld, N., 187, 188
Klotz, J., 214n16, 494
Knetter, M., 185, 211
Knight, B., 392
Kochman, L., 139
Koehler, D., 46
Koessler, F., 114
Koning, R., 434
Krautmann, A. C., 521, 522
Kuypers, T., 426, 434, 438, 441

L

Lacey, N. J., 139
Lakonishok, J., 140, 185
Lancaster, K., 468
Land, V. Q., 515
Lane, D., 187, 209n10, 256, 257, 260, 261, 263, 265, 266, 267, 268, 269
Langer, E., 213
Latané, H., 201
Law, D., 276, 277, 410
Leeson, P. T., 150
Lehmann, B. N., 140
Leong, S. M., 209, 279
Lessmann, S., 300
Levin, N., 193
Levitt, S. D., 47, 113, 242, 442, 444
Lim, K. G., 209, 279
Linnemer, L., 47
Little, I., 440n12
Lo, V. S. Y., 18, 53, 58, 59, 60, 200, 206, 207, 208, 280, 297, 452
Loomes, G., 46
Losey, R. L., 105n4, 139, 194, 285, 286
Lucas, R. E., 391
Ludlow, L., 194

M

Machina, M., 36, 40
MacLean, L. C., 201n7, 215, 216, 261, 321, 372, 462n1, 501, 518, 521
Maher, M., 433
Manski, C. F., 110, 119, 398
Markowitz, H. M., 39, 108n9, 187
Marsaglia, G., 498
Marsden, J. R., 107, 156, 282, 283, 284
Marsh, T. L., 520
Mason, P. M., 514
Matheson, V. A., 505, 510, 519, 520, 521, 522
Mauborgne, R., 404

McClelland, A. G. R., 301
McCrae, M., 284, 286, 295
McCulloch, B., 280
McDonough, L., 86, 90n11, 96, 112, 113n19, 117, 149, 162n1, 163, 409
McGlothlin, W. H., 30, 35, 104n1, 105, 121, 153, 162, 162n1, 191, 192, 295, 317n8
McHale, I., 425, 441n14
McKinnon, R. I., 271
Merton, R. C., 169n10, 170
Metrick, A., 104n3
Metzger, M. A., 28n6, 35, 46, 192, 214n15, 301
Miers, D., 424n2
Mikesell, J. L., 514, 515
Milgrom, P., 95n15
Mitchell, D., 195
Moore, P. G., 484
Morgan, W. D., 8, 9, 9n8
Morton, R. H., 482, 490, 497
Munting, R., 423, 424

N

Nelson, F., 395
Newton, B. A., 288
Noble, B. J., 68
Novarro, N. K., 516

O

O'Brian, N. R., 281
Oliver, P., 407
Ortner, G., 395
Oster, E., 520
Ottaviani, M., 30, 86, 96n17, 97, 99, 99n21, 114, 162n1, 163, 342nn1-2, 348
Owens, D. H., 139

P

Pope, P., 429, 437
Pankoff, L. D., 210, 230
Papachristou, G., 47, 518
Parayre, R., 270
Paton, D., 113, 113n19, 117, 295, 372, 409, 410
Paul, R. J., 240
Pawlukiewicz, J. E., 139
Peel, D. A., 104n3, 276, 277, 410, 411, 429, 430, 438, 440, 448
Peirson, J., 283, 284
Pennock, D. M., 100, 395
Perloff, J. M., 408
Petrin, A., 468n7

Piron, R., 110, 143n4
Plackett, R., 197
Plott, C. R., 119n26, 394, 395, 398
Pope, P. F., 411, 429, 430, 438, 440, 448
Potesman, A. M., 165
Potters, J., 113, 114
Prelec, D., 38, 88, 88n8
Prendergast, C., 342n2

Q

Quandt, R. E., 32, 42, 90, 91, 92, 109, 110, 120, 125, 141, 163, 192, 193, 207, 208, 280
Quinn, J., 195
Quirin, W. L., 290

R

Raines, S. T., 119n26, 141, 142, 143, 143n4, 147, 149, 150, 153, 154, 155, 162n1, 163, 409
Ray, M., 18, 19
Rhode, P. W., 149, 388, 391, 399
Riedwyl, H., 496, 497, 500, 501
Rietz, T., 397
Ritter, J., 185, 205
Roberts, B., 392
Roberts, H., 184, 309
Roberts, P. M., 288
Robinson, M. E., 363, 369
Rolfo, J., 271
Roll, R., 184
Roman, S. A., 311, 313, 323, 324, 333
Rosenqvist, G., 208
Rosett, R. N., 105, 192
Rosner, B., 193, 294, 297
Ross, G., 8, 12
Rubinstein, M., 157, 164, 167, 174, 179, 196
Rue, H., 433
Ryan, M. E., 119n26

S

Sackrowitz, H., 370
Salanié, B., 30, 32, 37, 38, 40, 41, 42, 43, 85, 111, 111n16, 115n21, 119, 120, 372
Salvesen, O., 433
Santa-Clara, P., 389
Sauer, R. D., 33, 85, 105n4, 139, 162n1, 210, 212, 276, 279, 301, 372, 431, 434, 438, 456
Savage, L. J., 40, 107, 108, 110, 141, 187, 197
Scharfstein, D., 342

Schnytzer, A., 195
 Scholes, M., 166, 169, 169n10, 170
 Schotman, P., 47
 Schultz, D., 254
 Schwartz, S., 212
 Scott, F. A., 519, 521
 Scriabin, M., 210, 230
 Serbin, V., 165
 Servan-Schreiber, E., 394, 395
 Shapira, Z., 519
 Sharp, R., 7
 Sharpe, G., 423, 424, 426
 Shastri, K., 157
 Shaw, J., 270
 Shaw, R., 139, 150, 280
 Shiller, R., 386n1
 Shilony, Y., 195
 Shin, D., 440n13
 Shin, H. S., 32, 33, 86, 95, 96n18, 113, 163, 195, 276, 372, 409, 416
 Shmanske, S., 441n14
 Shumway, T. G., 164, 164n2, 165, 169n9, 170, 174, 174n13, 178
 Siddique, A., 179
 Simmons, R., 373, 424n1, 431, 434, 439, 441, 443, 443n17, 449, 452, 453
 Simmons, S., 7
 Simon, J., 47, 496
 Sinclair, S., 4n2
 Skiena, S. S., 254n1
 Slemrod, J., 392
 Smidt, S., 185
 Smith, L. R., 110, 143n4
 Smith, M. A., 286, 287, 288, 301, 373, 374, 407, 407n1, 409, 410, 411, 412, 415, 416
 Snedecor, G., 439
 Snowberg, E., 30, 45, 141, 164, 309n1, 317n8, 389, 390, 393, 400
 Snyder, W. W., 52, 105, 139, 162, 194, 276, 277, 278, 285, 286, 295, 412
 Sobel, R. S., 119n26, 141, 142, 143, 143n4, 147, 149, 150, 153, 154, 155, 162n1, 163, 409, 467n5, 510, 515
 Song, C., 432
 Sørensen, P. N., 30, 86, 96n17, 97, 99, 99n21, 114, 162n1, 163, 342nn1–2, 348
 Sosik, J., 437
 Spindler, C. J., 515
 Stackelberg, H., 99
 Staelin, R., 296
 Stark, S. R., 515
 Stein, J., 342

Stern, H., 52, 53, 54, 56, 59, 198, 200, 208, 211, 215, 231, 234, 235, 236, 370, 492, 499
 Stiglitz, J. E., 409
 Stole, L., 342n2
 Stott, H., 434, 436
 Stover, M. E., 520
 Strumpf, K. S., 149, 388, 391, 399, 442
 Sugden, R., 46
 Summers, S. R., 515
 Sung, M. C., 184, 296, 297, 300, 372
 Sweeney, R. J., 140
 Swidler, S., 139, 150, 280

T

Talbott, J. C., 105n4, 139, 194, 285, 286
 Tamarkin, M., 34, 34n18, 35, 36, 91n12, 105n4, 109, 119n26, 140, 210, 240, 452, 467n5
 Tavaré, S., 369
 Terrell, D., 96n19, 112, 428
 Tetlock, P., 397
 Thaler, R. H., 28n3, 33, 42, 47, 85, 105, 107n8, 108, 110, 115, 121, 138, 140, 157, 163, 187, 193, 295, 372, 462n1, 518
 Thalheimer, R., 5, 6, 7, 8, 9, 9n9, 10, 10n10, 11, 12, 13, 18, 187
 Thorp, E. O., 185, 201n7, 261, 321, 372
 Tomarkin, M., 437, 438
 Tompkins, R., 30
 Tryfos, P., 210, 230
 Tuckwell, R. H., 139, 281, 284, 286, 287, 293, 295
 Tversky, A., 39, 46, 107, 111, 163, 174, 193
 Twomey, P. M., 373

U

Ulyangco, K., 444

V

Valkanov, R., 389
 Van Harlow, W., 140
 Van Zijl, A., 280, 295
 Varola, F., 311
 Vasche, J. D., 7, 8, 9, 9n8
 Vaughan Williams, L., 28n5, 113, 113n19, 117, 187, 284, 295, 342n1, 372, 373, 374, 406, 409, 410, 456
 Venezia, I., 519

Vergin, R. C., 210, 230, 288, 437
 Vuillier, J.-J., 311

W

Wagner, G. A., 189
 Waldron, P., 281
 Walker, I., 462, 463, 468
 Walls, D. W., 289
 Weinbach, A. P., 240
 Weitzman, M., 28n6, 33, 34, 46, 86, 91, 108, 109, 115n21, 141, 192, 414, 415
 Welles, C., 187
 Wethyavivorn, K., 157
 Whaley, R. E., 164, 165, 167, 174, 178, 179
 White, E. M., 298, 299
 White, H., 456
 Whitley, R., 368
 Wit, J., 113, 114
 Wolfers, J., 30, 45, 141, 149, 164, 309n1, 317n8, 386n2, 387, 388n3, 390, 391, 392, 394, 395, 396, 397, 398, 400
 Woodland, B. M., 104n3, 111n5, 139, 437
 Woodland, L. M., 104n3, 111n5, 139, 437

Y

Young, J., 462, 463, 468

Z

Zaman, A., 498
 Ziegelmeyer, A., 114
 Ziemba, R. E. S., 261
 Ziemba, W. T., 28n3, 30, 33, 42, 47, 85, 105, 107n8, 108, 110, 115, 118n24, 120, 121, 138, 139, 162, 163, 174, 183, 184, 186, 187, 191, 193, 195, 196, 196n5, 201, 201n7, 202, 202n8, 203, 205, 206, 208, 209n10, 212, 213n14, 225, 254, 256, 257, 260, 261, 263, 265, 266, 267, 268, 269, 270, 278, 279, 280, 282, 291, 292, 293, 295, 307, 308, 309, 309n1, 311, 317n8, 321, 324, 372, 411, 456, 462n1, 466, 486, 498, 501, 518, 519, 521
 Zitzewitz, E., 149, 386n2, 387, 388n3, 390, 391, 392, 394, 395, 396, 397, 398
 Zorn, C. K., 514, 515
 Zott, C., 404
 Zuber, R. A., 105n4, 139, 210, 230