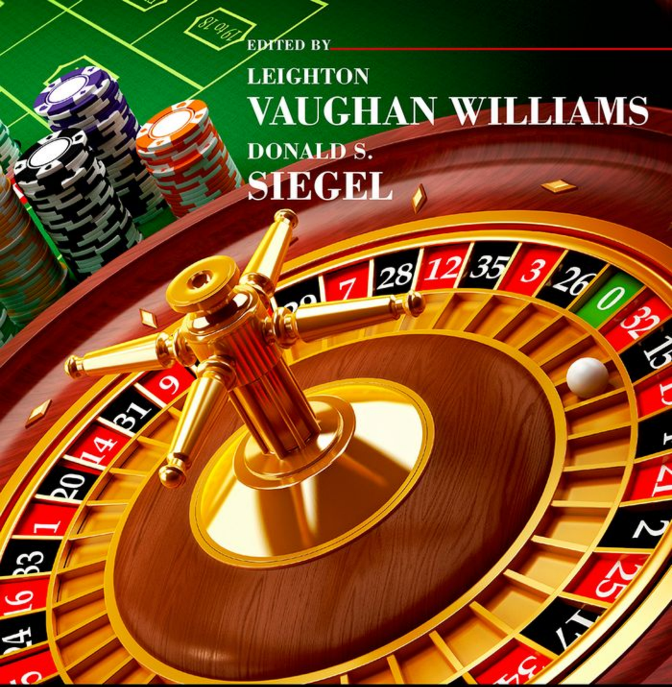


EDITED BY
LEIGHTON
VAUGHAN WILLIAMS
DONALD S.
SIEGEL

The Oxford Handbook *of* THE ECONOMICS OF GAMBLING

THE OXFORD HANDBOOK OF

THE ECONOMICS OF GAMBLING

THE OXFORD HANDBOOK OF

THE ECONOMICS OF GAMBLING

Edited by

LEIGHTON VAUGHAN WILLIAMS

and

DONALD S. SIEGEL

OXFORD
UNIVERSITY PRESS

OXFORD
UNIVERSITY PRESS

Oxford University Press is a department of the University of Oxford.
It furthers the University's objective of excellence in research, scholarship,
and education by publishing worldwide.

Oxford New York
Auckland Cape Town Dar es Salaam Hong Kong Karachi
Kuala Lumpur Madrid Melbourne Mexico City Nairobi
New Delhi Shanghai Taipei Toronto

With offices in
Argentina Austria Brazil Chile Czech Republic France Greece
Guatemala Hungary Italy Japan Poland Portugal Singapore
South Korea Switzerland Thailand Turkey Ukraine Vietnam

Published in the United States of America by
Oxford University Press
198 Madison Avenue, New York, NY 10016

Library of Congress Cataloging-in-Publication Data
Vaughan Williams, Leighton.
The Oxford handbook of the economics of gambling/Leighton Vaughan Williams
and Donald S. Siegel.
pages cm. — (Oxford handbooks)
Includes bibliographical references and index.
ISBN 978–0–19–979791–2 (cloth: alk. paper)
1. Gambling. 2. Gambling—Economic aspects. I. Siegel, Donald S., 1959– II. Title.
HV6710.V38 2013
338.4'7795—dc23
2012038792

1 3 5 7 9 8 6 4 2
Printed in the United States of America
on acid-free paper

Contents

Contributors

Josemari Aizpurua is a Professor in the Department of Economics at the Public University of Navarre in Spain.

Gary Anders is an Emeritus Professor of Economics in the W.P. Carey School of Business at Arizona State University and Adjunct Professor of Global Economics at Thunderbird School for Global Management.

John E. Anderson is the Baird Family Professor of Economics at the College of Business Administration, University of Nebraska–Lincoln (USA).

Alistair Bruce is Professor of Decision and Risk Analysis at Nottingham University Business School.

Chava Carter is a Research Assistant at UMBC, the University of Maryland, Baltimore County.

Les Coleman is a Senior Lecturer in the Department of Finance at the University of Melbourne.

Stephen Creigh-Tyte is Visiting Professor and Research Fellow at Durham Business School.

George Diemer is Assistant Professor of Business at Chestnut Hill College.

Greg Durham is Assistant Professor of Finance in the College of Business at Montana State University.

David Edelman is a Senior Lecturer in the School of Business at University College, Dublin.

Scott Farrow is a Professor in the Department of Economics at UMBC, the University of Maryland, Baltimore County.

David Forrest is Professor of Economics at the Salford Business School, University of Salford.

Ramon P. DeGennaro is the CBA Professor in Banking and Finance at the University of Tennessee.

Ann B. Gillette is a Professor in the Department of Economics, Finance, and Quantitative Analysis at Kennesaw State University.

John Goddard is Professor of Financial Economics at Bangor Business School, Bangor University.

Andrew Grant is Lecturer in Finance at the University of Sydney Business School.

Kent Grote is Assistant Professor of Economics and Business in the Department of Economics and Business at Lake Forest College.

O. David Gulley is Professor of Economics at Bentley University.

Robert Hannum is Professor of Risk Analysis and Gaming in the Remain School of Finance at the University of Denver.

Brad R. Humphreys is Professor of Economics and Chair in the Economics of Gaming in the Department of Economics at the University of Alberta.

Johnnie Johnson is Professor of Decision and Risk Analysis and Director of the Centre for Risk Research at the University of Southampton.

David Johnstone is Professor in Finance at the University of Sydney Business School.

Martien Lamers is a doctoral student at the Department of Financial Economics, Ghent University.

John Lepper is Adjunct Professor, Alfred Deakin Research Institute, Deakin University.

Loreto Llorente is Associate Professor in the Department of Economics at the Public University of Navarre in Spain.

Leonard MacLean is Herbert S. Lamb Chair in Business Education at Dalhousie University.

David McDonald is a Research Fellow at the Centre for Risk Research in the School of Management, University of Southampton.

Vasiliki Makropoulou is Research Associate at Athens University of Economics and Business.

David Marginson is Head of the Accounting and Finance Section and Professor of Management Accounting at Cardiff Business School.

Victor Matheson is an Associate Professor in the Department of Economics at the College of the Holy Cross in Worcester, Massachusetts.

John C. Navin is Professor and Chair of the Department of Economics and Finance at Southern Illinois University Edwardsville.

Mark W. Nichols is Professor of Economics at the University of Nevada, Reno.

David Paton is Professor of Industrial Economics at Nottingham University Business School.

Rodney J. Paul is a Professor in Department of Sport Management at the David B. Falk College of Sport and Human Dynamics at Syracuse University.

Levi Pérez is an Associate Professor in the Department of Economics at the University of Oviedo and a Member of the Research Staff of the Fundación Observatorio Económico del Deporte.

Javier Puértolas is Associate Professor in the Department of Economics at the Public University of Navarre in Spain.

J. James Reade is Lecturer in Economics in the Department of Economics at the University of Birmingham.

Paul Rhode is a Professor in the Department of Economics at the University of Michigan.

Warren Richards is an instructor in the Department of Economics and Finance at Southern Illinois University Edwardsville.

Ryan Rodenberg is an Assistant Professor in the Department of Sport Management at Florida State University.

Adi Schnytzer is Associate Professor of Economics at Bar Ilan University.

Donald S. Siegel is Professor and Dean of the School of Business at the University at Albany, SUNY.

Ricardo Siu is Associate Professor of Business Economics in the Department of Finance and Business Economics at the University of Macau.

Michael Smith is a Senior Lecturer in Economics at Leeds Business School.

Koleman Strumpf is Koch Professor of Economics at the University of Kansas School of Business.

Timothy S. Sullivan is an instructor in the Department of Economics and Finance at Southern Illinois University Edwardsville.

Ming-Chien Sung is Professor of Risk and Decision Sciences and Director of MSc Risk Management at the School of Management, University of Southampton.

Richard Thalheimer is President of Thalheimer Research Associates, Inc. in Lexington, Kentucky.

Mehmet Serkan Tosun is an Associate Professor and Director of Graduate Programs in the Department of Economics at the University of Nevada, Reno.

Leighton Vaughan Williams is Professor of Economics and Finance and Director of the Betting Research Unit at Nottingham Business School, Nottingham Trent University.

Douglas M. Walker is Professor of Economics at the College of Charleston in Charleston, South Carolina.

Andrew P. Weinbach is the Colonel Lindsey H. Vereen Professor in the Wall College of Business at Coastal Carolina University.

William Ziemba is Alumni Professor of Financial Modeling and Stochastic Optimization, Emeritus at the University of British Columbia. He is also a Visiting Professor at the Mathematical Institute of the University of Oxford.

INTRODUCTION

In recent years, there has been a substantial rise in interest among academics and policy makers in the economics of gambling. A concomitant trend in several nations has been the implementation of major regulatory changes and modifications to the taxation of gambling markets. Examples include a fundamental change in the United Kingdom in 2001 from a turnover-based tax on betting operators to a tax based on gross profits, resulting in the effective abolition of taxation levied directly on bettors (Paton, Siegel, and Vaughan Williams 2002, 2004), followed in 2005 by extensive reforms to the gambling sector introduced in the Gambling Act. In the United States, passage of the Unlawful Internet Gambling Enforcement Act of 2006 had profound implications for the global online gambling sector. There have also been numerous regulatory changes to gambling in Europe, Asia, and Australia. These changes, and an increase in attention paid to revenue generated from this activity, have heightened interest in understanding the economics of this sector.

Despite growing interest in the economics of gambling, there is no comprehensive source of pathbreaking research on this very broad topic. The purpose of this handbook is to fill this gap. We commissioned chapters from leading academics on all aspects of gambling research. Topics covered include the optimal taxation structure for various forms of gambling, factors influencing the demand and supply of gambling services, forecasting of gambling trends, regulation of gambling, wagering on sports, horses, politics, gambling in casinos and on the Internet, the efficiency of racetrack and sports betting markets, gambling prevalence and behavior, modeling the demand for gambling services, the economic impact of gambling, substitution and complementarities among different types of gambling activity, and the relationship between gambling and other sectors of the economy. These are all important issues, with significant global implications. Specifically, we divide the handbook into sections on casinos; sports betting; horse race betting; betting strategy; motivation, behavior, and decision-making in betting markets; prediction markets and political betting; and lotteries and gambling machines.

I Casinos

The first section explores the economic effects of casinos. Gary Anders of Arizona State University examines the employment impact of casino gambling in the United States in eight key industries. This is a critical issue due to the recent recession and persistently high unemployment. Anders reports that expansion of casinos results in employment increases in arts and entertainment, hotels, and food and beverage industries but leads to a decline in employment in management and professional services, technology, and manufacturing (generally high-paying jobs). The negative employment effects of casino expansion are exacerbated in states with competing gambling venues.

In the next chapter, John Anderson of the University of Nebraska examines how national governments tax casino gambling, including wagering taxes, admissions taxes, and fees. He also considers the effects of taxation on equity and efficiency. An interesting aspect of Professor Anderson's analysis is that he highlights major gambling locations, such as Las Vegas, Macau, and Singapore. Mark Nichols and Mehmet Serkan Tosun of the University of Nevada, Reno, provide important theoretical and empirical evidence on price and income elasticities of demand for casino gambling. Ricardo Siu of the University of Macau analyzes the growth and evolution of Asian casino gambling. In addition to examining the economic aspects of this activity, he also focuses on the significance of the unique features of Asian culture and the related institutional structure of gaming industry performance. Finally, Professor Siu assesses the social benefits and costs of casino gaming in Asia.

John Navin, Timothy Sullivan, and Warren Richards of Southern Illinois University Edwardsville review the literature on gaming and the restriction of smoking. They then present an empirical analysis of the effects of a smoking ban on the riverboat casino market in the Illinois portion of the St. Louis Metropolitan area. The section on casinos concludes with a chapter by Douglas Walker of the College of Charleston, who summarizes empirical research on the economic and social impacts of gambling. Issues considered in this chapter include the effects of casino gambling on economic growth, relationships among gambling industries and the implications of these relationships for net government tax revenue, the social costs of gambling, casinos and crime, casinos and political corruption, and problems with applying cost-benefit analysis to gambling.

II Sports Betting

The second section begins with a chapter by George Diemer of Chestnut Hill College and Ryan Rodenberg of Florida State University on a growing sector of the gambling industry: online sports betting. Diemer and Rodenberg consider this sector from a law and economics perspective and compare the efficiency of online/offshore sports books relative to conventional sports books in Las Vegas and London.

David Forrest of the University of Salford and Levi Pérez of the University of Oviedo examine football pools, defined as any pari-mutuel wagering concerning the outcomes, or any other aspects, of football (soccer) matches. These are long odds, high-prize games where a share of the jackpot is linked to football results. Salford and Pérez assert that this type of wagering closely resembles lotto, where the difference between the two types of games is that one depends on football results while the other is based on numbers drawn randomly. They review the literature on football pools, focusing on the United Kingdom and Spain.

In the next chapter John Goddard of Bangor University examines the efficiency of soccer betting markets. Specifically he evaluates whether certain types of forecasting models can be used to develop profitable fixed-odds betting strategies. A key finding is that inefficiencies in the market have been largely eliminated by the increased sophistication of contemporary sports betting markets, greatly enhanced by advances in information technology.

Loreto Llorente, Jose Maria Aizpurua, and Javier Puértolas of the University of Navarra examine the efficiency of a special type of betting market: wagering on pelota matches. These are games with two mutually exclusive and exhaustive outcomes, where wagers are made among viewers via a middleman who receives 16 percent of the prize. Based on field data, the authors analyze three different concepts of market efficiency widely utilized in the literature and also provide some insights for future research on hedging strategies in these markets.

Rodney Paul of Syracuse University, Andrew Weinbach of Coastal Carolina University, and Brad Humphreys of the University of Alberta study the baseball betting market. The authors ask a very specific research question: How do elite starting pitchers affect gambling behavior and volume? Using comprehensive data, they report that games involving an elite pitcher attract more bettors, especially on the "under" wager. J. James Reade of the University of Birmingham and John Goddard of Bangor University analyze information efficiency in high-frequency betting markets. Their analysis is based on "betting exchanges," which enable traders to either buy or sell bets on many sporting events. Such continuously operating online betting markets have ensured the transition of the use of high-frequency data from the financial setting into the betting market context. The authors review recent academic research on the topic of information efficiency in high-frequency, in-play betting markets for football (soccer).

III Horse race Betting

David Edelman of University College Dublin examines the implications of the alleged "takeover" of tote (totalizator) betting markets by "sophisticated" gamblers. Based on a mathematical analysis, he concludes that their activity could not in itself seriously or irreparably damage pari-mutuel markets. Richard Thalheimer of Thalheimer Research

Associates examines an important trend: the rise of casino-style gambling at pari-mutuel racetracks. Specifically he analyzes the impact of "racino" betting on pari-mutuel racing as well as its effects on state lotteries and casino gaming.

Ramon DeGennaro of the University of Tennessee and Ann Gillette of Kennesaw State University provide a comprehensive economic analysis of racetrack betting. The authors examine the effects of technological change on this industry, the growth and evolution of betting options, and the efficiency of the betting market. They also consider the antecedents and consequences of subsidies for this sector, which many view as a declining industry. David Marginson of Cardiff Business School examines starting price–based overrounds on Betfair, the leading person-to-person Internet betting site. His empirical analysis is based on 2,184 horse races that took place in the United Kingdom between 2008 and 2010. He reports a positive relationship between grade of race and Betfair overround (the higher the grade, the higher the overround). His results imply that microstructure analysis of order-driven betting markets, such as Betfair, constitutes a fruitful area of research for those interested in understanding market efficiency.

Adi Schnyzter of Bar-Ilan University examines the incidence of insider trading in the market for horse and greyhound racing. He analyzes whether the presence of betting insiders at the track implies that their presence is easily detected. It is shown that this is a function of the microstructure of the particular betting market. In some markets the impact of insider trading is readily measured, while in others it has hitherto proven virtually impossible to detect. The different microstructures of betting markets and the implications for insider trading and its measurement are considered.

In the following chapter, Schnytzer, Vasiliki Makropoulou of Utrecht University, and Martien Lamers of Ghent University analyze pricing decisions and insider trading in fixed-odds horse racing. This chapter conceptualizes fixed-odds horse betting markets as implicit call option markets. The decision-making process of a bookmaker is modeled as a decision-maker who sets prices under uncertainty. The authors show that when a bookmaker adopts this pricing process built on implicit options, the returns will exhibit a favorite-longshot bias. By performing Monte Carlo simulations, option values are generated and a measurement is made of the degree of insider trading.

IV Betting Strategy

The section on betting strategy begins with a chapter by Andrew Grant of the University of Sydney, who examines the predictability of sports betting markets. Specifically he analyzes wagering on simultaneous events and "accumulator gambles." These are situations in which multiple games occur simultaneously. In this case the gambler must consider the problem of allocating capital across different games or events, similar to the situation when an investor allocates capital to different stocks in a portfolio. The author explores the use of accumulator bets (parlays) as part of a portfolio betting strategy.

Robert Hannum of the University of Denver presents a primer on the mathematics of gambling. He demonstrates how the mathematics behind the games generates revenues and drives the economics of gambling. In the following chapter, Hannum describes the science and economics of poker, also from a mathematical standpoint. He also discusses the history of this popular table game. Hannum asserts that poker is unique among gambling activities for two reasons. First, it is not house-banked. Second, there is a considerable amount of skill involved in this game, which is not the case for other casino and lottery games.

Leonard MacLean of Dalhousie University and William Ziemba of the University of British Columbia describe an interesting betting strategy known as the Kelly strategy, where the expected logarithm of final wealth is maximized. The authors consider the advantages and disadvantages of this strategy, with reference to its application in blackjack, horse racing, lotto, and anomalies in the index futures markets. They also discuss the use of Kelly-type strategies by what they term "great investors."

The section on betting strategy concludes with a chapter by Michael Smith of Leeds Metropolitan University, who explores the performance of "experts," or media forecasters, in selecting winners. He also provides a more comprehensive examination of the degree of information efficiency with respect to these events. His empirical analysis considers whether there are any betting strategies based on these "expert" picks that can systematically beat the market.

V Motivation, Behavior, and Decision-Making in Betting Markets

Alistair Bruce of the University of Nottingham examines individual motivations for betting, synthesizing perspectives from economics, psychology, and sociology. His results have important implications for policy makers designing legal and regulatory regimes for betting as well as for those interested in treating the negative effects of "excessive" exposure to betting. The next chapter, by Les Coleman of the University of Melbourne, examines a variety of characteristics of betting markets. He focuses on comparing betting markets to financial markets and then assesses the motivations for gambling.

In a similar vein David McDonald, Johnnie Johnson, and Ming-Chien Sung of the University of Southampton present evidence of biased decision-making in betting markets. An interesting aspect of their research is that they demonstrate how systematic biases that were first identified in the laboratory are reflected in real-world gambling behavior. Greg Durham of Montana State University examines sports betting, and specifically point spread wagering, through the lens of "behavioral finance," which draws heavily from psychology.

VI Prediction Markets and Political Betting

David Johnstone of the University of Sydney examines the question of the predictability of sports betting markets based on a simple automated "market maker" for prediction markets. He notes that there is great potential for the use of automated "robot" market makers in prediction markets and market simulation games, based on research in experimental economics and behavioral finance. He outlines the benefits of adopting such an approach in terms of two key advantages: (1) the model is easy to derive and (2) the opening security price can be set arbitrarily between zero and one, so as to match the market maker's prior beliefs.

The next chapter, by Paul Rhode of the University of Michigan and Koleman Strumpf of the University of Kansas, describes the long history of political betting markets. Contrary to popular wisdom, political futures markets are not a recent invention. The authors trace the operation of political futures markets back to sixteenth-century Italy, eighteenth-century Britain and Ireland, nineteenth-century Canada, and twentieth-century Australia and Singapore. They also note that election wagering was quite popular in the United States in the pre-1860 period but during the post-1860 period became increasingly concentrated in the organized futures markets in New York City.

VII Lotteries and Gambling Machines

David Forrest of the University of Salford and O. David Gulley of Bentley University examine the efficiency of lottery markets. The authors conclude that lotto players act as if they understand the "rules of the game" and appropriately use relevant information about the games. Exceptions to efficiency are found, but these "inefficiencies" cannot be easily exploited by bettors.

John Lepper of Deakin University and Stephen Creigh-Tyte of Durham Business School provide an historical analysis of the U.K. National Lottery. Although the first National Lottery draw of the modern era occurred in November 1994, state-sponsored lotteries had been common since the reign of Queen Elizabeth I. The authors describe the development of the National Lottery and consider the introduction, nature, and performance of the modern Lottery while also describing the economics of the National Lottery.

Scott Farrow and Chava Carter of the University of Maryland, Baltimore County, assess the costs and benefits of slot machine gambling. This chapter begins by defining the slot machine segment of the gambling industry and then reviews the economics of these machines based on benefit-cost analysis. They also review illustrative empirical studies and provide suggestions for additional research.

Kent Grote of Lake Forest College and Victor Matheson of the College of the Holy Cross survey the literature on the economics of lotteries in terms of two central themes. The first section examines the microeconomic aspects of lotteries, including consumer decision-making under uncertainty, price and income elasticities of demand for lottery tickets, cross-price elasticities of lottery tickets to each other and to other gambling products, consumer rationality and gambling, and the efficiency of lottery markets. The second section covers topics related to public finance and public choice, including the revenue potential of lotteries, the tax efficiency and dead-weight loss of lottery games, the horizontal and vertical equity of lotteries, earmarking and the fungibility of lottery revenues, and individual state decisions to participate in public lotteries.

Leighton Vaughan Williams of Nottingham Trent University and David Paton of the University of Nottingham consider the taxation of gambling machines. Although there has been considerable research on the economic impact of gambling on regional economies in the United States and the United Kingdom, relatively little research has focused on the optimal taxation of gambling machines within these facilities. The authors seek to fill this gap by examining the theoretical arguments for taxing gambling machines by means of a levy on machine takings rather than by means of a licence fee levied per machine. Recent tax debates in the United Kingdom provide an ideal context for such a discussion.

References

Paton, David, Donald S. Siegel, and Leighton Vaughan Williams. 2002. A policy response to the e-commerce revolution: The case of betting taxation in the UK. *Economic Journal* 112(480): F296–F314.

——. 2004. Taxation and the demand for gambling: New evidence from the United Kingdom. *National Tax Journal* 57(4):847–861.

This page intentionally left blank

SECTION I

CASINOS

This page intentionally left blank

CHAPTER 1

THE EMPLOYMENT IMPACT OF CASINO GAMBLING IN THE U.S.

GARY ANDERS

The Growth of Gambling in the United States

Over the past twenty years, gambling has become a growth industry in the United States, the United Kingdom, Canada, Australia, Macau (Macao), and other nations.[1] Once restricted to Nevada and, later, Atlantic City, casinos now operate in 33 U.S. states. Since Congress passed the Indian Gaming Regulatory Act (IGRA) in 1988, tribes in 28 states have established casinos, and others are seeking federal recognition so that they can do the same. According to the National Indian Gaming Commission, 236 Indian tribes are currently operating casinos.[2] In 2010, Indian casinos earned approximately \$26.5 billion (National Indian Gaming Commission 2011). California alone has 56 tribal casinos, making it one of the most lucrative gambling markets in the world. Starting in 1989 with Iowa, states began to legalize land-based riverboat casinos, horse and dog racetrack casinos (racinos), and urban casinos.[3] Altogether these non-Indian commercial gambling casinos earned over \$34.6 billion in gross gambling revenues in 2011 (American Gaming Association 2011). Including those operated by Indian tribes and regulated by the states, casinos generated over \$61 billion in annual revenues.

Over a decade ago, the National Gambling Impact Study Commission (NGISC) was established to investigate various issues regarding the legalization of gambling. After sponsoring several major studies and conducting hearings in different cities across the country, the NGISC issued a final report that called for a moratorium on new gambling venues (NGISC 1999). Despite this recommendation, new casinos continue to be established, with urban land-based casinos and Indian casinos being the most prevalent. The rapid growth of commercial gambling has been striking not only in the United States but also worldwide. The unprecedented growth of government-sanctioned gambling raises a number of policy issues related to the social and economic impacts of gambling. The purpose of this study is to examine how the introduction of commercial gambling casinos impacts a state's economy by focusing on changes in industry employment.

This chapter is divided into four parts. The first provides an overview of the literature on the economic and social impacts of commercial gambling. This is followed by a set of hypotheses that are used to examine the effects of casino gambling on state employment. Next, a regression analysis of the relationship between changes in industry employment and gambling is presented. The conclusion summarizes the findings and discusses the limitations of the study and future research directions.

Literature Review

This section summarizes some of the most relevant research on the economic and social impact of casinos on local communities. A useful starting point is Rose (2001). Based on a survey of more than 100 published studies, it concluded that, as a general rule, a new casino provides positive economic benefits to its host economy. Among these benefits are direct and indirect employment from the construction of the casino, taxes collected by the state, the capture of gambling revenues from residents who would otherwise gamble in out-of-state casinos, the increase in consumer utility from additional recreational choices, and employment from casino jobs. Rose also mentions such social costs as increased gambling addictions, congestion, and profits going to outside interests. He claims that most of the economic impact studies he reviewed suffer from a number of critical flaws, major omissions, or biased assumptions.

Various gambling interests have sponsored numerous studies to support the adoption of casinos. Representative of this type of work is the Arthur Andersen study conducted for the American Gaming Association (AGA). Without mentioning any negative impacts, this study lauds the economic benefits of casinos, asserting that "the introduction of casinos leads to growth in almost all other areas: retail sales, commercial and housing construction, restaurants, etc." (Arthur Andersen 1997, 45) and claiming that casinos are responsible for increased taxes, employment growth, and reductions in the number of families on welfare. Similarly, a recent AGA-funded study by the Brattle Group (2011) provides analysis supporting the positive impact of casinos and includes direct, indirect, and induced multiplier effects of casinos. What is interesting is that this study expands the scope of casino operations to include hotels, food and beverage services, and other business lines.

William R. Eadington (1998), a leading scholar in the field, includes tourism development, economic revitalization, tax revenue, jobs, new investment, and employment opportunities for minorities as expected benefits from casinos. He argues that destination resort casinos are the strongest in job creation, due to the fact that these casinos are able to export the activity to nonresidents while capturing significant tax revenue for the local jurisdiction (Eadington 1999).

Douglas Walker and John Jackson's (1998) article is one of the more rigorous empirical studies of the relationship between legalized gambling and local economic growth. This study employs regression analysis to address two research questions: (1) Does legalized gambling contribute to state economic growth? And (2) Must gambling be exported for economic growth to occur? They found evidence that casino gambling and greyhound racing do increase state income but that it is not necessary for the state to export gambling in order for these results to hold. A follow-up to this study (Walker and Jackson 2007) employed Granger causality analysis to examine the relationship between casino gambling and state-level economic growth as measured by Consumer Price Index–adjusted personal income. The authors found that there does not appear to be a significant relationship between the introduction of casino gambling and per capita income.

Daniel Felsenstein, Laura Littlepage and Drew Klacik (1999) use the prisoner's dilemma to explain the conundrum faced by many states regarding the legalization of casinos. They believe that competitive bidding by states to attract gambling casinos undercuts the potential economic development benefits. The authors identify positive economic impacts in terms of new investment, jobs, higher incomes, and consumer choice. The negative effects are increases in compulsive gambling pathologies and the regressive economic impact on those in lower income groups who are the primary customers of casinos.

William Thompson, Ricardo Gazel, and Dan Rickman (2000) identify some of the ways that compulsive gambling affects the economy. These include theft, forgeries, bad debts, insurance fraud, credit card fraud, loan sharking, increased criminal justice costs, civil court costs, divorce, bankruptcy proceedings, treatment costs, public assistance, and suicide. They assert that, taken in full measure, these costs outweigh the benefits of casino gambling.

Earl Grinols, a professor of economics at the University of Illinois, is a leading critic of legalized gambling. He argues that the introduction of riverboat casinos in Illinois did not create the jobs that were promised and had little impact on local unemployment (Grinols 1994). Grinols (1996) further reported that only when gambling is able to tap outside markets will casinos have a positive economic impact; otherwise gambling results in inefficient transfers from one business to another. Exacerbating this problem are the costs of problem gamblers who constitute the largest share of the casino client base. Grinol's book *Gambling in America* (2004) discusses the casino industry and presents a theoretical cost-benefit framework for assessing the net economic impact of gambling. As with his other writings (Grinols and Mustard 2001, 2006), there is strong anti-gambling sentiment reflecting his concern regarding the social costs associated with gambling.

Louise Simmons (2000) found that while casinos do create jobs in related industries, they also tend to divert consumer expenditures from other businesses. Furthermore she asserts that jobs resulting from casino development are lower paying positions. She believes that gambling exacerbates social problems among lower income and minority groups. She argues that casinos are not a sound economic development alternative for many local communities seeking to generate increased tax revenues, especially given the possibility of market saturation from increased numbers of casinos.

Several studies by Gary Anders and Donald Siegel have examined the possibility of economic displacement as a result of gambling. Anders, Siegel, and Munther Yacoub (1998) found that the operations of Indian casinos in Arizona were strongly correlated with structural changes in state tax revenue. Siegel and Anders 1999 found strong statistical evidence of displacement between riverboat casinos and other sectors in Missouri. Siegel and Anders 2001 reported evidence of substitution between casinos and state lotteries. Anthony Popp and Charles Stehwien (2002) have found further evidence of a negative correlation between Indian casino gambling and state revenues.

Focusing on the employment impacts, Thomas Garrett (2003) analysed monthly data from the Bureau of Labor Statistics (BLS) for six counties: two in Mississippi, two in Illinois, one in Iowa, and one in Missouri to examine the relationship between employment and the opening of a casino. He found evidence that in three of the four rural counties, establishing a casino did increase household employment. In the case of the two urban counties, however, he found that it was much harder to detect a significant impact of casinos on either household employment or payroll.

Garrett's approach uses a comparison of differences between a forecast of selected county employment versus actual county employment after the introduction of a casino. The five counties selected were Warren and Tunica counties, Mississippi; Massac and St. Could counties, Illinois; Lee County, Iowa; and St. Louis County, Missouri. His analysis produced mixed results. In three counties (Warren, Tunica, and Massac) actual household employment after the introduction of a casino was greater than forecasted. In two counties (St. Clair and Lee County), actual household employment after the introduction of a casino was less than forecasted household employment. In St. Louis County, the underlying cyclical nature of the economy undermined a comparison. Garrett's study includes a comparison of six employment sectors for each county. These include manufacturing, retail trade, services, financial, construction, and casino employment.

Since the data Garrett used date from the introduction of casinos until December 2001, the study presents only limited consideration of the long-term employment effects. Moreover, differences in the size of the local economies make it difficult to draw meaningful conclusions. One significant aspect of this study worth noting is that Garrett finds rural counties more likely than metropolitan counties to benefit in terms of increased household employment. The reason for this may have to do with the relative size of gambling employment compared to total employment, which would be smaller in metropolitan counties than in less populated rural areas of the state.

In his pathbreaking book *The Economics of Casino Gambling* (2007), Douglas Walker concludes his brief discussion of the impact of casinos on employment and wages with the following observation. "Unfortunately, there has been relatively little research on the labor market effects of the casino industry. This is an important issue that deserves more attention from independent researchers" (8).

Discussion

This brief review of the existing literature presents a mixed set of results regarding the economic impact of commercial gambling. On the one hand, proponents of gambling argue that casinos create new jobs, increase local tax revenues, and stimulate economic growth through induced consumption and employment multipliers. Critics of gambling, on the other hand, assert that casinos displace consumer expenditures from other businesses and largely result in transfers from one sector to another while increasing social problems. The literature also reflects heightened concern about the social costs associated with pathological gambling that increase both public and private costs as well as the way in which these costs have been distorted by inappropriate methodologies (Walker 2003).

Interest in the field of gambling studies has increased due to the growing recognition of the economic significance of gambling. New researchers are entering the field, and highly respected academic journals are more interested in publishing gambling-related research. Along with this have been efforts to apply rigorous empirical analytical approaches to gambling-related policy issues. My own research in this area has greatly benefited from collaborations in the United States and the United Kingdom, where there has been an even greater recognition of the importance of gambling.

The following presents four testable hypotheses that will be used to structure a regression analysis of the employment impact of casinos. These are:

Hypothesis 1: The construction and operation of casinos are expected to increase the total number of new jobs and through income and expenditure multipliers will lead to employment growth in other sectors of a local economy.

Hypothesis 2: The operation of casinos is expected to have a positive impact on employment in businesses related to entertainment and recreation as well as business services. The jobs created by new casinos are likely to include higher paying professional/managerial positions as well as lower paying jobs in the food and beverage industry.

Hypothesis 3: Assuming the possibility of cannibalization arising from the competition between casinos and other recreational establishments, the growth in employment in competing entertainment businesses and retail establishments, including restaurants and bars, will be negatively correlated with casinos drawing a significant number of local customers.

Hypothesis 4: Used as an economic development tool, casinos will increase state revenue and per capita income. This will occur because states with legalized commercialized gambling and Indian casinos will collect additional revenues as a result of revenue sharing arrangements with the gaming establishments. In addition, jobs created by casinos will reduce the state's unemployment rate.

Regression Model

A regression model for 11 states with casino gambling was used to test these hypotheses. The states included in the study are California, Connecticut, Illinois, Iowa, Louisiana, Michigan, Minnesota, Mississippi, Missouri, Nevada, and New Jersey. These states were selected for two reasons. First, they have the longest history of legalized gambling. Second, data were not available for other states, for reasons that will be delineated below. Monthly employment data for 8 industries in these 11 states were collected from 1990 to 2004 from Bureau of Economic Analysis (BEA) website (www.bea.gov).

According to the AGA (2011), the global financial crisis of 2008 resulted in a significant decrease in gross revenues for casinos, which have yet to return to their peak levels of 2006–2007. Given that 8 out of 22 states with commercial casinos are continuing to experience decreases in income, employment, and tax revenues, I decided to utilize a time period that captures an entire national business cycle, starting with the contraction that began in the early 1990s and continuing through the recovery and expansion that lasted into the early years of the George W. Bush administration. Among the various possibilities, eight industries that previously have been identified as most likely to be directly affected by the opening of a casino were selected for analysis. These industries are

(1) Professional scientific and technical services
(2) Management of companies and enterprises
(3) Depository and non-depository institutions
(4) Construction
(5) Arts and Entertainment;
(6) Gambling
(7) Accommodations
(8) Food services and drinking establishments.

As noted above, it was not possible to include more states in the study due to the lack of data. A major problem in collecting state employment data is one of standardizing the series. Starting in 2001 the U.S. government switched the industry classification system from the Standard Industrial Classification (SIC) to the North American Industry Classification System (NAICS) (For a discussion of the switch see www.census.gov/epcd/www/naicsdev.htm.) As of early 2013, many states had yet to reconcile their industry employment data with the new classifications. Because the industries in the NAICS were changed, only the same industries included in the old SIC

Table 1.1 NAICS and SIC Industry Classification Comparisons

Industry Name	NAICS Code	SIC Code
Professional Scientific and Technical Services	1200	875
Management of Companies and Enterprises	1300	400
Construction	0400	300
Depository and Non-Depository Institutions	1002	710
Real Estate	1101	734
Gambling	1703	835
Accommodations	1801	805
Food Services and Drinking Establishments	1802	627

were used. For example, the professional, scientific and technical services classification was SIC 875 but after 2001 became NAICS 1200. Thus because of omitted observations it not possible to include more industries or states within the current framework. The NAICS and SIC codes for the industries used are given in table 1.1.

Another issue that needs to be mentioned is the problem of collecting data on Indian casinos. While data on gaming revenues by casino are available for commercial casinos from state gaming commissions, actual revenue data for most Indian-run casinos cannot be obtained.[4] As of this writing, only highly aggregated tribal gaming revenue data are available from the National Indian Gaming Commission. Because of this, proxies in the form of additional measures of gambling activity were used. These variables include the number of casinos, the number of admissions into casinos, the total number of tables, the total number of slot machines, the number of wins per admission, the total square footage of the casinos, and the number of tables, information available from industry publications, such as *International Gaming & Wagering Business.*

Additional explanatory variables were created using this data set. For gambling these variables include win per admission, win per table, win per slot machine, and win per square foot.[5] Also, data from the U.S. Census Bureau provided another a set of other possible explanatory variables. Data for all 11 states for the years 1990 to 2004 were collected on the following variables: industry revenue, state population, state per capita income, total employment, state and local revenue, and average earnings. Also included were changes in both state and local revenues as well as selected industries. A complete list of the explanatory variables is presented in table 1.2.

The original approach was to use a regression framework to determine how industry employment was affected by the introduction of casino gambling in a particular state by using gambling data only for that state. However, it became clear that gambling is a regional as opposed to a discrete state economic activity. For example, riverboat casinos in Illinois compete with casinos in Missouri and Indiana and, to a lesser extent, with casinos in Nevada and Atlantic City. Subsequent regression models for each industry were therefore constructed using a panel of state data.

Table 1.2 Independent Variables

Variable Name	Explanation
NSM	Total number of slot machines in the state in a particular year
Year	A time trend to reflect exogenous growth
CMR	Year-to-year change in manufacturing industry revenue of a state
CSLR	Year-to-year change in state and local revenue of a state
CP	Year-to-year change in population of a state
TNT	Total number of gambling tables in a state in a particular year
WPA	Win per admissions in gaming in a state in a particular year
CGR	Year-to-year change in gaming revenue of a state
CDR	Year-to-year change in depository and non-depository revenue of a state
CCR	Year-to-year change in construction revenue of a state
TSFA	Total square footage of casinos present in a state in that year
NA	Total number of admissions into casinos in a particular state in that year
CFDSR	Year-to-year change in food and drinking services revenue of a state
CPCPI	Year-to-year change in per capita personal income of a state

The first set of preliminary regressions suffered from low Durbin-Watson (DW) values (0.2–0.5) due to auto correlation between the independent and dependent variables. In an attempt to correct the problem year-to-year changes in the variables were used. This specification increased the DW statistic to about 1 but reduced the R^2 value significantly. Next, other independent variables, such as per capita personal income and state and local revenues, were introduced. Also, the specification of the dependent variables was revised to measure the change in the ratio of employment in each industry to total employment in the state. This allowed the model to capture the resulting change in the composition of state employment rather than the absolute change and made it easier to isolate the effect of changes in the independent variables. With these changes the test statistics for the regressions improved; however, the DW statistic still indicated the existence of serial correlation.

Various alternative approaches and specifications were used in an attempt to correct this problem (i.e., taking the ratio of two variables or taking the first difference and dividing it by another variable). After some trial and error, the Generalized Least Squares (GLS) produced the best results. No doubt there are other methodological issues, relating to stationarity of the data. However, this issue is complicated by the profound effect that technology played in restructuring certain segments of the economy over this time period. To capture some of this effect a technology variable proxy (personal computers per capita) was included, as suggested by David Card and John DiNardo (2003).

The GLS method can be performed in one of two methods: Prais-Winstein or Cochrane-Orcutt. Prais-Winstein is preferred over Cochrane-Orcutt because it provides greater flexibility to reduce serial correlation by allowing weights to be assigned to the independent variables. Using this method the dependent and the independent variables were modified according to the following equations:

$$Y^* = Y_t - \rho \times Y_{t-1}$$

$$X^* = X_t - k \times X_{t-1}$$

Where ρ and κ are assigned weights depending on the degree of serial correlation in the model. The Prais-Winstein method involved taking the ratio in dependent variable (industry sector employment in the state (E_{IJ}) to (T_J), the total employment in the state.) Then the regression model becomes:

$$\left\{\left(\frac{E_{IJt}}{T_J}\right)_t - \rho \times \left(\frac{E_{IJt}}{T_J}\right)_{t-1}\right\} = (X_t - k \times X_{t-1}) \times \beta + \varepsilon_t$$

$$=> \left(\frac{E_{IJ}}{T_J}\right)^*_t = X_t^* \times \beta + \varepsilon_t$$

Where X_t is vector of independent variables.

This specification produced an acceptable Durbin-Watson. However, by using the Prais-Winstein method the first observation is lost, which reduces the R^2 value. To recover this observation the first observation was weighted by $1/\sqrt{1-p^2}$ for dependent variables and by for independent variables. Then the regression model becomes

$$X^* = \begin{bmatrix} \sqrt{1-\rho^2} & \sqrt{1-\rho^2}x_{11} & \sqrt{1-\rho^2}x_{21} \\ 1-\rho & x_{12}-\rho x_{11} & x_{22}-\rho x_{21} \\ 1-\rho & x_{13}-\rho x_{12} & x_{23}-\rho x_{22} \end{bmatrix}$$

$$\left(\frac{E_{IJ}}{T_J}\right)^* = \begin{bmatrix} \sqrt{1-\rho^2}\left(\frac{e_{ij}}{t_j}\right)_1 \\ \left(\frac{e_{ij}}{t_j}\right)_2 - \rho\left(\frac{e_{ij}}{t_j}\right)_1 \\ (\frac{e_{ij}}{t_j})_3 - \rho\left(\frac{e_{ij}}{t_j}\right)_2 \end{bmatrix}$$

$$\left(\frac{E_{IJ}}{T_J}\right)^*_t = X_t^* B + \varepsilon_t$$

Results

This specification produced acceptable results for all eight industries. Table 1.3 presents the results, including T-statistics, R^2 values, and level of statistical significance. The R^2 statistic ranges from .625 to .295, indicating that the model accounts for up to two-thirds of the variation in industry sector employment changes. This is acceptable given the data limitations and the fact that only a small number of states had complete time series for these industries. The results indicate that some of the gambling variables (i.e., number of machines, numbers of tables, and gambling revenue) are positively correlated with changes in employment for some sectors, such as employment in food

Table 1.3 Regression Results

Dependent Variable Independent Variable GLS in Parentheses	(A) Change in Ratio of Science and Technology Employment to Total Employment	(B) Change in Ratio of Management Employment to Total Employment	(C) Change in Ratio of Credit Employment to Total Employment	(D) Change in Ratio of Construction Employment to Total Employment	(E) Change in Ratio of Arts Employment to Total Employment	(F) Change in Ratio of Gambling Employment to Total Employment	(G) Change in Ratio of Accommoda-tion Employment to Total Employment	(H) Change in Ratio of Food and Drinking Places Employment to Total Employment
Constant	−1280.935 (−3.194)**	−1257.983 (−3.004)**	−0.082 (−0.201)	1.604 (3.039)**	−1074.43 (−2.360)*	613.541 (2.443)*	−446.923 (−1.423)	1.245 (1.454)
No. of slot machines (NSM) (0.95)++	−0.000432 (−2.889)**	–	–	–	0.00001 (1.980)*	–	–	–
Year (Y) (1)++	0.641 (3.195)**	0.634 (3.016)**	–	–	0.536 (2.354)*	−0.308 (−2.446)*	0.224 (1.823)	–
Change in state and local revenue (CSLR) (0.92)++	−0.0000288 (−1.729)	–	–	−0.000367 (−3.198)**	–	0.002133 (2.440)*	−0.000414 (−3.523)**	–
Change in population (CP) (0.9)++	0.116 (3.883)**	–	−0.032 (−2.948)**	0.03722 (2.903)**	0.05138 (3.192)**	−0.021 (−1.516)	0.08399 (4.401)**	0.02673 (2.323)*
Total number of tables (TNT) (1)++	–	–	–	–	0.006178 (1.813)	–	–	–

Win per admission (WPA) (1)++	–	−0.037 (−1.977)*	–	–	–	–	–	–
Change in Gaming Revenue (CGR) (1)++	–	-0.003 (2.543)*	0.0229 (1.974)*	–	0.006438 (2.544)*	0.08096 (4.786)**	0.01186 (5.362)**	0.036786 (2.394)*
Number of admissions (NA) (1)++	–	–	–	−0.000275 (−1.984)*	–	–	–	–
R^2	0.625	0.558	0.295	0.314	0.350	0.393	0.541	0.314
Standard Error	4.433	1.548	1.980	1.382	3.206	2.635	3.546	2.572
Durbin-Watson	1.475	1.926	1.743	1.635	2.017	1.622	1.587	1.456
DF	72	64	83	58	58api	67	74	74

* (*T*-statistic) * at 95% confidence level, ** at the 99% confidence level.
++ Auto Correlation factor for each variable used in the Prais-Winstein method.

and drinking establishments and accommodations, and arts industries, but negatively correlated with employment changes in other sectors, such as science and technology, and management.

An examination of the relationships among variables reveals some interesting trends. First, an obvious demographic change involving an aging population of baby boomers has played a strong role in the growth of recreation industries that is reflected in the growth in hotel and restaurant employment. Second, beginning in the early 1990s, an economic contraction and increased global competition led companies to downsize their workforce. Related to this were reductions in the number of managerial and supervisory positions. Third, changes in technology and the use of personal computers resulted in negative employment trends in such industries as banking and credit intermediation. Thus increases in state population had a strong positive effect on employment in most industries except those experiencing structural changes. Finally, changes in state and local government revenues are positively correlated with the change in gambling employment but negatively correlated with employment changes in science and technology, construction, and accommodations industries. It is not obvious why this is the case because the results appear to be counterintuitive, but previous gambling research indicates that revenue displacement may have affected state revenues to some degree. Additional research is needed to examine this issue more carefully.

Conclusions

This chapter attempts to shed light on the impact of casinos on state economies. Many of the economic impact studies written to support the adoption of casinos use expenditure and employment multipliers, which demonstrate that casinos benefit regional economics through direct purchases and employment or through indirect multiplier effects (Rose 2001). The prevailing wisdom is that the induced impacts associated with gambling employment stimulate the creation of additional jobs in related industries. Given that employees in the casinos and related businesses pay state income taxes and sales taxes on their purchases of goods and services, states are expected to benefit from additional tax revenue. Thus the total fiscal impact of casinos is often thought to be positive. Although limited by data availability, these preliminary finding suggest that the expansion of commercial and tribal casinos may have a positive local employment effect in some sectors, but taken regionally where states have legalized competing gambling venues, the overall employment effect for other sectors is quite possibility negative. The evidence suggests that casinos may not have as significant a positive effect on higher paying jobs (e.g., jobs in science and technology, and management) due to the types of jobs created by casino expansion and possible job losses that occur from competition with other sectors of the economy.

The literature indicates that there are several factors that significantly increase or reduce the positive impacts of casinos. First, to a certain extent, the economic benefits of casinos depend on whether gambling is exported to residents of other states or whether opening local casinos encourages residents not to gamble in other jurisdictions (import substitution). For maximum benefit local economy casinos should bring in new money rather than displace existing consumer expenditures. For example, when Mississippi legalized riverboat casinos, the state captured some of the gambling revenue that would have been spent by its residents in Louisiana. Likewise, Mississippi gamblers now substitute gambling at local casinos for trips to Las Vegas or Atlantic City, and this increases the economic contribution to the state. But as more and more states adopt casino gambling the opportunities for import substitution decrease, and the only way that growth can occur is through increasing the amount gambled per visit or increasing the number of gamblers.

A major unresolved issue is the nature of source of money spent in casinos . If some portion of the gaming dollars comes from savings, or substitutions from other types of expenditures, the displacement effect would be smaller. At this time there is not enough specific household information on gambling behavior to draw conclusions. The existing literature does, however, suggest that a large segment of the gambling market is from lower income households where the marginal propensity to consume is high and saving is conversely low (Borg, Mason, and Shaprio 1990).

Given the favorable public attitude toward legalized gambling, I anticipate continued expansion in new markets (Anders 2003). In calculating the net economic impact of gambling, policy makers should take into consideration the reductions taking place in the proportion of higher paying jobs that are lost against the economic value of lower wage jobs generated in the entertainment, hotel, and food and beverage sectors. Further expansion of casino gambling also raises the possibility of market saturation. Unless the growth of gambling casinos is met with an increase in the demand for gambling, then as with other industries, there could be overcapitalization and eventually reduced profitability (Eadington 2007). Furthermore, on-line gambling presents a new competitive alternative that could significantly affect the demand for casino gaming. Internet gambling is already estimated at over $50 billion worldwide and is growing rapidly. Given the ease and convenience of this form of on-line gambling it is certain to become a serious competitor to casinos. It will be interesting to see how this phenomenon, which is quite popular among younger gamblers, will offset current employment trends caused by the increased demand for recreation and tourism by an aging population of baby boomers.

Notes

* The author wishes to acknowledge the research assistance of Sandeep Kumar Somavarapu and the helpful comments of Donald Siegel and William Eadington.

1. For a discussion: Kurlantzick (2007).
2. Some of the largest and most profitable casinos are tribally owned and located in areas that may have had only pari-mutuel gambling or state lotteries.

3. In 1989 Colorado and South Dakota legalized historical limited stakes casinos. Starting in the early 1990s Illinois, Indiana, Iowa, Louisiana, Missouri, and Mississippi legalized riverboat casinos. Other states have legalized racetrack casinos.
4. Because the Indian Gaming Regulatory Act specifically exempts tribes from the Freedom of Information Act, it is not possible to acquire data on tribal gambling revenues unless a tribe makes that information available.
5. Gambling industry data are collected from issues of Christiansen Capital Advisors' *International Gaming & Wagering Business*, Bear Stearns' *North American Gaming Almanac*, and gambling websites for various states.

References

American Gaming Association. 2011. *State of states: The AGA survey of casino entertainment.* Washington, D.C.: American Gaming Association; http://www.americangaming.org/newsroom/op-eds/state-states-2011-turning-corner.

Anders, Gary C. 2003. Reconsidering the economic impact of Indian casino gambling. In *The Economics of Gambling*, edited by Leighton Vaughan Williams. London: Routledge, 204–223.

Anders, Gary C., Donald Siegel, and Munther Yacoub. 1998. Does Indian casino gambling reduce state revenues? Evidence from Arizona. *Contemporary Economic Policy* 16(3): 347–355.

Arthur Andersen. May 1997. *Economic impacts of casino gambling in the United States, Volume 2: Micro Study.* Study conducted for the American Gaming Association. Las Vegas, Nev.: Arthur Andersen.

Brattle Group. 2011. *Beyond the casino floor: Economic impacts of the commercial casino industry.* Study conducted for the American Gaming Association.

Bear Stearns. 2002–2003. *North American gaming almanac.* Las Vegas, Nev: Huntington Press.

Beyond the casino floor: Economic impacts of the commercial casino industry. Study conducted for the American Gaming Association.

Borg, Mary, Paul Mason, and Stephen Shapiro 1990. "An economic comparison of gambling behaviour in Atlantic City and Las Vegas." *Public Finance Quarterly*, 18(3): 291–312.

Card, David, and John E. DiNardo. 2003. Technology and U.S. wage inequalities: A brief look. In *Technology, growth, and the labor market*, edited by Donna Ginther and Madeline Zavodny. Boston: Kluwer Academic Publishers, 131–160.

Eadington, William R. 1998. Contributions of casino-style gambling to local economies. *Annals of the American Academy of Political and Social Science* 556:53–65.

——. 1999. The economics of casino gambling. *Journal of Economic Perspectives* 13(3): 173–192.

——. 2007. "Trends in commercial gaming industries globally: Issues that might interest economists." Paper presented at the conference The Growth of Gambling and Prediction Markets: Economic and Financial Implications, University of California Riverside, Desert Palms, May 21.

Felsenstein, Daniel, Laura Littlepage, and Drew Klacik. 1999. Casino gambling as local growth generation: Playing the economic development game in reverse? *Journal of Urban Affairs* 21(4):409–421.

Garrett, Thomas A. 2003. *Casino gambling in America and its economic impacts.* St. Louis, Mo.: Federal Reserve Bank of St. Louis.

Grinols, Earl L. 1994. Incentives explain gambling's growth. *Forum for Applied Research and Public Policy* 11(2):119–124.

——. 1996. Who loses when casinos win? *Illinois Business Review* 53(1):7–11.

——. 2004. *Gambling in America.* Cambridge: Cambridge University Press.

Grinols, Earl L., and David B. Mustard. 2001. Business profitability versus social profitability: Evaluating industries with externalities, the case of casinos. *Managerial and Decision Economics* 22:143–162.

——. 2006. Casinos, crime, and community costs. *Review of Economics and Statistics* 88(1): 28–45.

Kurlantzick, Joshua. 2007. Raising the stakes. *Foreign Policy*, May/June.

National Gambling Impact Study Commission (NGISC). 1999. *National Gambling Impact Study Commission final report.* Washington, D.C.: NGISC; http://govinfo.library.unt.edu/ngisc.

National Indian Gaming Commission. 2011. National Indian Gaming Commission announces 2010 industry gross gaming revenue; Indian gaming revenues remain stable. Press release No. 175 07-2011. Washington, D.C. National Indian Gaming Commission, July 18; http://www.nigc.gov/Media/Press_Releases/2011_Press_Releases/PR-175_07–2011.aspx.

Popp, Anthony V., and Charles Stehwien. 2002. Indian casino gambling and state revenue: Some further evidence. *Public Finance Review* 30(4):320–330.

Rose, Adam Z. 2001. The regional economic impacts of casino gambling. In *Regional science perspectives in economic analysis*, edited by Michael L Lahr and Ronald E. Miller. Amsterdam: Elsevier, 345–378.

Siegel, Donald, and Gary Anders. 1999. Public policy and the displacement effects of casinos: A case study of riverboat gambling in Missouri. *Journal of Gambling Studies* 15(2):105–121.

——. 2001. The impact of Indian casinos on state lotteries: A case study of Arizona. *Public Finance Review* 29(2):139–147.

Simmons, Louise. 2000. High stakes casinos and controversies. *Journal of Community Practice* 7(2):47–69.

Thompson, William N., Ricardo Gazel, and Dan Rickman. 2000. Social cost of gambling: A comparative study of nutmeg and cheese state gamblers. *UNLV Gaming Research & Review Journal* 5(1):1–15.

Walker, Douglas M. 2003. Methodological issues in the social cost of gambling. *Journal of Gambling Studies* 19(2):149–18.

——. 2007. *The economics of casino gambling.* New York: Springer.

Walker, Douglas M., and John J. Jackson. 1998. New goods and economic growth: Evidence from legalized gambling. *Review of Regional Studies* 28(2):47–69.

——. 2007. Do casinos cause economic growth? *American Journal of Economics and Sociology* 66(3):593–607.

CHAPTER 2

THE ECONOMICS OF CASINO TAXATION

JOHN E. ANDERSON

1 Introduction

In this chapter, I analyze the various ways that governments tax casino gambling around the world and consider the economic effects of that taxation. In the first section, I examine the types of casinos and casino gambling that takes place along with the forms of taxation that governments apply to casino operations, including wagering taxes, admissions taxes, fees, and other taxes. In section 2 I examine the economics of casino taxation, including market analysis of a casino game, efficiency effects of casino game taxation, equity impacts or the incidence of casino taxation, and optimal government tax policy regarding casinos. Section 3 provides a summary of the forms of taxation used around the world, highlighting major gambling locations, such as Las Vegas, Macau (Macao), and Singapore, among others.

2 Economic Analysis of Casino Taxation

In order to analyze casino taxation, the tax base, or that which is taxed, must first be identified so that the tax rate(s) and any exemptions, deductions, or credits can be considered. For general background on economic analysis of casino gambling, see Suits (1979b), Eadington (1999), Walther (2002), Benar and Jenkins (2008), and Hoffman et al. (1999).

The most common form of casino taxation around the world is a wagering tax, based on the adjusted gross receipts (AGR) collected by casinos on all forms of games that they offer (table games, roulette, slot machines, etc.). AGR is generally defined as gross gambling receipts minus payouts for prizes. Some governments also subject casinos to admissions taxes and fees of various sorts. Admissions taxes are imposed by several states that permit casinos in the United States, as indicated in Anderson (2005), although these taxes are not frequently used in the rest of the world. Fees are often charged to support social purposes in the jurisdictions where casinos are permitted, as in Macau, for example, where two fees are levied for social and economic purposes.

2.1 Market Analysis of a Casino Game

Following the analysis in Anderson (2005), we can consider basic market analysis of a casino game. A unique terminology is used in the world of gambling, but the economic analysis of casino game taxation is relatively straightforward. Bettors place wagers on a game, with the total amount wagered called the handle H, and the casino withholds a fraction, w, of the handle. That fraction is called the takeout rate, and it determines the price of the casino game. The total amount of prizes paid out to bettors P can be written as $P = H - wH = (1 - w)H$. Then, if we solve this equation for the takeout rate w, we obtain $w = 1 - P/H$. Hence, the takeout rate is one minus the ratio of total prizes paid to bettors divided by the total amount wagered, or, the handle.

Consider how the relationship between the total amount of prizes awarded and the handle affect the takeout rate. At one extreme, suppose the casino was entirely benevolent and paid out all the money wagered in prizes. In that case, $P = H$ and the takeout rate is $w = 0$. At the other extreme, suppose the casino pays out no prizes, so $P = 0$ and the takeout rate is unity: $w = 1$. In general, the derivative of the takeout rate w with respect to the prize amount P is, $\frac{\partial w}{\partial p} = -\frac{1}{H}$, indicating that as the prize amount rises the takeout rate falls in inverse proportion to the total amount wagered H. Consequently, the price of the casino game varies inversely with the prize amount.

The demand for casino gambling is inversely related to prices of the casino games. At lower (higher) relative prices, other things being equal, we expect a larger (smaller) quantity demanded. In this regard, there is nothing different about the demand for casino gaming as compared to other goods and services. With full information, a fundamental requirement of a well-functioning market, casino gamblers will have well-behaved demand curves. The demand for a casino game can be written as $H(w)$, where $H'(w) < 0$, reflecting the usual situation where the higher the price, measured by the takeout rate applied by the casino, the smaller the amount wagered (handle). We cannot assume that gamblers are perfectly informed about the price of each game they play, but there is evidence that the price is generally well known by gamblers. For example, there is empirical evidence in the lottery literature reported by Victor Matheson and Kent Grote (2004) indicating that lottery ticket purchasers exhibit a high degree of rationality as the effective price of a lottery ticket changes with the size of the jackpot. It should be noted, however, that there is also evidence of non-rational lottery play, such as overuse of birthday numbers (1–31), as in Anderson and Schmidt 2002.

The market for a casino game is illustrated in Figure 2.1, where the price or the takeout rate w is measured on the vertical axis and the total quantity of bets or handle

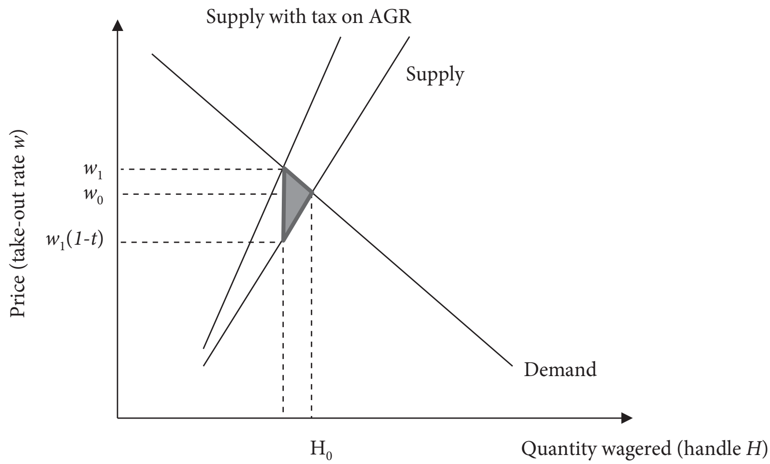


FIGURE 2.1 Market analysis of a casino game with a tax on AGR

H is measured on the horizontal axis. The demand function is downward sloping and can written as $H(w)$ where $H'(w) < 0$. On the supply side, we assume that there is an upward sloping supply. As the price of the casino game rises, the quantity supplied increases. Because the market for casino games is highly regulated in most governmental jurisdictions, often with limits on the number of games and gaming outlets, we illustrate the supply curve as relatively inelastic. The interaction of market supply and demand determines the equilibrium price (takeout ratio) w_0 and quantity (handle) H_0.

The precise nature of the demand curve deserves further attention. While some individuals gamble purely as a form of entertainment and can readily walk away from a casino after a gambling session regardless of their winnings or losses, pathological gamblers cannot. Hence, the overall demand for casino gambling may comprise two quite different groups of gamblers, each with distinct demand characteristics. Pathological and problem gamblers, not being very responsive to price given their addictions and compulsions, may have very inelastic demand. Other gamblers without such addictions and compulsions may have much more elastic demand. Evidence reported in Grinols 2004 indicates that in areas near casinos, pathological gamblers constitute one or two percent of the population, with problem gamblers making up another two or three percent. Of course these problem and pathological gamblers may constitute a much larger share of the population of gamblers who enter a casino. When the demands of the two groups are aggregated to obtain total market demand for a casino game, the overall demand will reflect the shares of the market demand that come from the two groups. The overall sensitivity of demand to price will reflect the dominant group of gamblers. Dean Gerstein et al. (1999) provide an overview of gambling impacts and human behavior, and the General Accounting Office (2000) provides insight on measurement of both economic effects and social effects of gambling.

There is some evidence that slot machines take in nearly 60 percent of their revenue from problem and pathological gamblers. There is also evidence that a typical casino derives about 80 percent of its revenue from slot machines. Consequently, nearly half of the casino revenue may be derived from problem and pathological gamblers. As a result, the demand for casino gaming is relatively inelastic. This is especially so if the casino has a local geographic monopoly due to restrictive regulations. If this is the case, the demand curve will be relatively inelastic, and so will the supply curve. In the extreme case of a local monopoly, the supply curve will be vertical and the demand curve will be relatively steep, both of which will have an impact on the incidence of taxation.

2.2 Effects of a Tax on Adjusted Gross Receipts (AGR)

Now consider the effect of introducing a tax on this casino game. The most common form of tax applied to casino games defines the tax base as the adjusted gross receipts (AGR) of the game. Tax revenue T generated by the tax determined by the product of the nominal marginal tax rate $t$and the tax base AGR: $T = tAGR$. Recognizing that the tax base AGR is the product of the handle and the takeout rate, $AGR = wH$, we can write the tax revenue as, $T = twH$. This expression reveals that the effective marginal tax rate applied to the handle is tw, the product of the nominal tax rate t and the takeout rate w. The higher the takeout rate the larger the effective tax rate. The supply curve shifts upward in a non-parallel manner. The reason for the non-parallel shift is due to the form of the tax. The tax revenue T is obtained by multiplying the marginal tax rate t times the AGR: $T = tAGR$. Because the AGR is the product of the handle and the takeout rate, $AGR = wH$, we can rewrite the tax as $T = twH$. This expression indicates that the marginal tax rate applied to the handle H is $\frac{\partial T}{\partial H} = tw$. Julie Smith (1999) and Jim Johnson (1985) identify three potential measures of the effective tax rate, depending on the definition of the tax base that is used in the computation.

A tax is applied to the casino's AGR, but we can analyze the market effects in terms of the handle H. The tax shifts the supply curve upward as indicated and results in a higher equilibrium takeout rate w_1 and lower equilibrium handle H_1. A tax of tw on handle H_1 raises the price paid by gamblers from w_0 to w_1 and lowers the price received by the casino from the original takeout rate w_0 to the new rate $w_1(1 - t)$. Revenue generated by the tax is tw_1H_1. Of that amount, the gamblers bear a tax burden of $(w_1 - w_0)H_1$ while the casino bears the remaining burden of $[(w_0 - w_1(1 - t)]H_1$.

The incidence of the tax can be analyzed in terms of the changes in takeout rate w and handle H. The tax shifts the supply curve upward as illustrated, resulting in a higher equilibrium price for the casino game w_1 and a smaller equilibrium quantity H_1. The tax raises the price paid by gamblers from w_0 to w_1 and lowers the price that the casino receives from w_0 to $(1 - t)w_1$. The tax generates revenue of tw_1H_1, of which the gambler bears the burden $(w_1 - w_0)H_1$ and the casino bears the burden

$[(w_0 - w_1)(1 - t)]H_1$. As is typical of tax incidence analysis, the tax has a statutory incidence falling entirely on the casino, but market forces result in economic incidence that differs from that. As usual, the economic agent with the less elastic behavior bears the greater share of the economic incidence of the tax. A regulatory environment that limits the supply of casino games causes the elasticity of supply to be relatively inelastic in relation to the elasticity of demand, thereby causing the incidence of the casino game tax to fall primarily on the casino. In fact, if the government jurisdiction has legal limits on the number of casinos, slot machines, or table games, then as those limits are reached the supply curve turns vertical. In the extreme with a vertical supply curve, the incidence of the tax falls entirely on the casino. With the proliferation of gaming opportunities in recent years, however, the supply curve is becoming increasingly elastic.

The tax causes a welfare loss illustrated as the shaded triangle. This is the reduction in total welfare over and above the tax revenue collected by the tax, or the excess burden of the tax. As usual, the size of this excess burden depends on the magnitude of the tax rate and elasticities of demand and supply. The greater the tax rate or either of the elasticities, the larger the excess burden. Figure 2.1 illustrates the excess burden as the usual shaded triangle with height t and base width $H_0 - H_1$. The magnitude of the excess burden of the tax depends on the tax rate as well as on the compensated elasticities of demand and supply. In the typical case of linear demand and supply, the excess burden rises with the square of the tax rate. Doubling the tax rate quadruples the excess burden, for example.

What do we know about the elasticity of demand? In a classic study of gambling, Daniel Suits (1979a) estimated elasticities of demand for several types of gambling. He found price elasticities substantially in excess of unity for legal bookmaking establishments in Nevada and also for wagering at thoroughbred racetracks. More recently, David Paton, Donald Siegel, and Leighton Vaughan Williams (2004) confirmed that the demand for bookmaking is price elastic in the United Kingdom. They used recent changes in U.K. tax policy with regard to bookmakers in order to estimate the elasticity. Both of these studies indicate elastic demand. David Forrest, David Gulley, and Robert Simmons (2000) estimated the elasticity of demand for U.K. national lottery tickets during the first three years of that lottery's existence and found that their estimated elasticity of -1.03 was not significantly different from -1, indicating unitary elastic demand. Richard Thalheimer and Mukhtar Ali (2003) have provided evidence on the demand for gaming at riverboat and racetrack casinos in the United States, finding that the demand for slot machine gaming was price elastic at the beginning of their study period (1991–1998) but that it declined to be approximately unitary elastic at the end of that period. They also found that demand for table games and slot machines are substitutes. Paton, Siegel, and Vaughan Williams (2003) provide a review of studies on the demand for gambling.

These estimates raise several important policy issues. First, high-price elasticities limit the effectiveness of gambling taxes, in terms of providing additional revenue for governments. Tax revenue is maximized at the unitary elastic point on the demand curve (where marginal revenue is zero). A monopolist will operate above that point on the demand curve where the price elasticity of demand exceeds unity (in absolute value). As a consequence, any tax rate increase that raises the price will reduce revenue. If governments have already set tax rates on casinos to maximize revenue, any further tax rate increases will reduce revenues. Beyond that, the finding of relatively high-price elasticities likely points to the availability of illegal gambling alternatives. That being the case, any further tax increase may simply drive gambling activity out of the legal casino sector into the illegal gambling sector. While these two studies do not estimate price elasticities for casino gambling as such, they do suggest that gambler responsiveness in general may be elastic.

2.3 Optimal Takeout Rate

Charles Clotfelter and Philip Cook (1987) were the first to consider the question of the optimal takeout rate in the context of state-run lotteries in the United States. More recently Herbert Walther (2004) examined the issue of optimal taxation of several forms of gambling, with a primary focus on lotteries, within an inter-temporal, state-dependent expected utility model. Using that type of model, he demonstrates that (1) optimal tax rates are higher for larger lotto communities, (2) jackpots induce overshooting bubbles, and (3) taxes on lotto and fix-prize gambles are regressive.

Paton, Siegel, and Vaughan Williams (2001) provide a helpful note in which they analyze the difference between a tax on stakes (the quantity wagered) and a general goods and services tax (GST) applied to the net revenue earned from gambling activity. They rightly indicate that a tax on stakes/wagers is the equivalent of an excise tax while the GST is ad valorem in its nature. Consequently, a policy change from an existing tax on stakes to a GST will potentially have efficiency gains.

2.4 Incidence of Casino Taxes

In popular discussion of casino taxation, it is often assumed that the tax burden falls entirely on the gambler, but that is not necessarily true unless the gambler's demand is price inelastic—that is, the gambler is completely unresponsive to price. Such is unlikely to be true except for the most pathological problem gambler. Consequently, we expect the tax burden to be shared by the gambler and the casino.

Taxes on gambling have generally been found to be regressive with respect to income. The seminal study finding this result is Suits 1977, which analyzed survey data on horse tracks, state lotteries, casino games, numbers, sports cards, off-track betting, and sportsbooks. Suits's estimates indicate that overall gambling taxes are somewhat more regressive than general state sales taxes (the most regressive of major state revenue sources). His early findings for Nevada casino gambling, in particular, were progressive. However, at that time casino gambling was limited to isolated locations, such as

Las Vegas, where the time cost of travel plus the out-of-pocket expense meant that only relatively high-income gamblers could afford to travel to Vegas to gamble. Since then, casinos have multiplied all over the United States, making it possible for gamblers to travel to a casino at a much lower cost. Consequently, more recent studies have found that casino taxes are regressive. For example, Paul Mason, Stephen Shapiro, and Mary Borg (1989) studied three groups of Las Vegas gamblers and found gaming taxes to be regressive for all three groups but especially so for local residents as compared to other Nevada residents and non-Nevada residents. A later study (Borg, Mason, and Shapiro 1991) also found regressive incidence. William Rivenbark (1998) found regressive incidence for casino gamblers in Mississippi as well.

Another aspect of casino tax incidence is obtained if we view these taxes as taxes on the economic rents earned by casino operators. In most cases, the casino operators have been granted a franchise by way of restrictive government regulations on casino gambling. Government legalization of casino gambling, when combined with restrictions on the number of casinos, table games, and slot machines, results in economic rents for the casino operators. The casino taxes, to the extent they fall on the owners and operators of casinos, may then be progressive. The casino taxes are paired with the liberalized regulations permitting casinos to operate. States grant the casino franchise then tax away part of the economic rents generated by the regulations they create.

2.5 Taxation of Casinos in Relation to other forms of Gambling

Casinos are not the only form of gambling that governments tax. Consequently, we must consider the ways in which taxation of casinos may affect other forms of gambling and the revenue derived from them. Lisa Farrell and David Forrest (2008) draw an important distinction in this regard. Their review of the literature on the relationship between alternative forms of gambling points out that the concepts of substitutes and complements are often misapplied. To be correct, the precise question is whether the cross-price elasticities for alternative forms of gambling are negative or positive. Authors in the gambling literature tend to be more interested in the question of whether a new form of gambling reduces sales from existing forms of gambling—a distinctly different, although important, issue.

Farrell and Forrest (2008) examined whether the introduction of casinos in Australia caused a change in lottery sales, displacing those sales. Using state-level panel data to analyze intra-state differences in the portfolio of games available, the authors report mixed results. The introduction of new super casinos appears to have reinforced lottery sales, while the addition of more slot machines in the network of local gaming venues appears to have diverted sales away from lottery games. The evidence provided in Paton, Siegel, and Vaughan Williams 2004 indicates that the demand for bookmaker gambling in the United Kingdom is highly sensitive to the tax rate principally because that form of gambling is a strong substitute for lottery and other forms of gambling. Forrest, Gulley, and Simmons (2010) found evidence from Britain that bettors substitute away from horse race, soccer, and numbers betting as the price of lottery tickets is unusually low.

2.6 Monopoly Regulatory and Pricing Rules

There are several aspects of the monopoly position often granted to casinos by governments that deserve more attention in the literature. For example, in cases where the government applies both an admissions tax and a wagering tax, as is common in the United States, the question is why both taxes are applied. This combination may entail an optimal two-part tariff, as is often implemented in rate-of-return regulatory contexts. Or it may reflect risk sharing on the part of the state government and the casino operator, with the state receiving a certain return for each gambler and the casino operator receiving a return based on performance.

Pursuing a monopoly regulatory point of view, we could consider how the government may regulate a monopoly casino that has been granted an exclusive franchise to operate in a given geographic area. The government could set the takeout rate to extract the monopoly rent earned by the monopoly casino. The casino would maximize its AGR net of its cost of operation: $AGR - C(H)$, where $C(H)$ is the casino's total cost of operation at the handle amount H. We can assume a concave cost function, with $C'(H) > 0, C''(H) < 0$. The usual monopoly pricing solution yields the optimal takeout rate $w_m = C'(H_m)\left[\frac{s_d}{(1+s_d)}\right]$, where the marginal cost at the monopolist's optimal handle $C'(H_m)$ is multiplied by the ratio of the elasticity of demand divided by one plus that elasticity. Then the government could allow the casino to retain a share of the handle equal to $\frac{C(H_m)}{H_m}$ and take the remaining share, $w - \frac{C(H_m)}{H_m}$.

The former monopoly position of casinos being granted exclusive rights to operate in isolated geographic markets has in recent years given way to a more competitive situation, however. At present there is more competition as more markets have been opened with liberalized gaming regulations. Furthermore, the market is more contestable than ever with a growing international market and computer-based means of gaming. Raymond Sauer (2001) examined the industry structure of gambling markets in the United States and how they have changed as government regulations have been altered in response to political and economic conditions. He used the interest group model to explain how government regulations have changed and become more liberalized. Smith (2000) examines the general issue of the government's stake in the gaming industry.

2.7 Multiple Taxes and the Second Best

The introduction of casino gambling in a region yields a new source of tax revenue, but that revenue is not entirely new. There are important interactions among the various preexisting taxes that are collected by state and local governments. Thomas Garrett (2003) has studied casino start-ups and warns that the new tax revenue a new casino generates cannot be considered as new money to the region in which it is located. His review of the evidence on revenue interactions indicates that the effects of casino revenue on other state revenue sources have to be examined carefully with particular attention paid to local conditions. Charles Leven and Donald Phares (1998) estimated the spending displacement caused by casinos introduced in the State of Missouri (USA). Their evidence indicates that spending on casino gaming comes at the expense of reduced spending on other goods and services, including other forms of gambling (dog and horse racetracks in the Missouri context), as well as from reduced savings. Since the displaced spending may have also been taxable, the net revenue gains to the state and local governments are smaller than they first appear.

It is important to disentangle the significant revenue interactions in order to get an accurate picture of the net revenue generated by new casinos. Mason and Harriet Stranahan (1996) modeled several channels by which tax revenue substitution may occur. One channel is through direct substitution of various forms of gambling. Suits (1977) originally provided evidence that other forms of gambling are substitutes for casino gambling, not complements. If that is the case, then a new gambling opportunity will substitute for existing gambling, with the result that the new casino revenue will in part displace existing tax revenue. The other channels of revenue substitution run through sectoral changes in the local economy that occur with the opening of a new casino. Adjustments in income and employment in the regional economy are important to consider. Furthermore, there are effects on the tourism industry. Suits (1982) provides an example of such an analysis for the City of Detroit.

Gary Anders, Donald Siegel, and Munther Yacoub (1998) examined whether the introduction of Native American casinos in Arizona (USA) in 1993 caused a structural shift in state tax revenue sources. They used data for the transaction privilege, use, and severance tax (TPT) collected in Maricopa County (the largest county in the state and the location of the state's largest city, Phoenix). The authors report that the new Native American casinos diverted spending from taxable to nontaxable sectors of the regional economy. Gamblers substituted expenditure on nontaxable Indian gaming for other taxable consumption expenditures. The TPT tax base was reduced as a result. They found that the spending displacement occurred primarily in retail trade, restaurants and bars, hotels and motels, and amusements. Siegel and Anders (2001) also examined the impact of Native American casinos, but their study considered the effect on state lottery revenue in Arizona. They used monthly time series data to estimate a model of lottery game sales. Their estimates indicate that a 10 percent increase in the number of slot machines due to the introduction of new casinos resulted in a 3.8 percent reduction in general lottery revenue and a 4.2 percent reduction in lotto revenue. In addition, Siegel and Anders (1999) estimated the economic displacement effects of riverboat casinos in Missouri (USA) using industry-level, time series data for the eleven counties where the casinos were introduced. Among the various effects they found, the strongest source of economic displacement was between casino spending and spending on entertainment and amusements.

There is limited direct evidence on how net revenue generated by casinos affects other tax collections. A study that examined this issue was conducted by Anthony Popp and Charles Stehwien (2002), who used county-level data for the State of New Mexico (USA) to examine whether Native American tribal casinos had an effect on total taxable gross receipts (TGR), a main source of business tax revenue in that state. Their evidence indicates that the first casino introduced in a county reduces TGR by a small but significant amount. A second casino introduced in a county decreases TGR by approximately 6 percent, reflecting substitution of spending on gambling relative to other taxable goods and services. They further found that the presence of a casino in a county reduces the TGR in contiguous counties, revealing spillover revenue effects. Shonkwiler (1993) examined the impact of new Atlantic City casinos on Nevada gaming revenues. Donald Elliott and John Navin (2002) examined how licensing of new casinos in Missouri (USA) cannibalized revenues from the state lottery. Their conclusion is that the state lottery lost \$0.83 in net revenue for each dollar of additional casino tax revenue. In the lottery literature there is ample evidence that increased lottery revenue comes at the expense of reductions in other tax revenue sources. See, for example, Fink, Marco, and Rork (2004) and Haas, Heidt, and Lockwood (2000).

The excess burden of a set of taxes on k forms of gambling, all of which are subject to taxation, would be written as $-\frac{1}{2}\sum_{i=1}^{k}\sum_{j=1}^{k} t_i t_j S_{ij}$. Arnold Harberger (1974, 37) provided this expression as a general description of the measurement of the welfare cost of a system of excise taxes on several goods. In this expression the $t_{i/j}$ terms are the taxes on each form of gambling and the $S_{i/j}$ terms are the pure substitution effects: $S_{ij} = \partial X_i / \partial P_j$ (using compensated demands, with no income effects). Analysts wishing to estimate the welfare effects of gambling taxes can use this approach and the above expression, to derive empirical estimates of substitution effects. These provide a more accurate measure than would be obtained from a partial analysis, which considers only the casino tax.

2.8 Second-Best Taxation

Suppose that a casino tax t_c is introduced in the presence of an existing tax on lottery tickets t_l. The existing tax on lottery tickets creates an existing excess burden of taxation illustrated in the dark blue triangle in the right-hand panel of Figure 2.2. The introduction of a casino tax creates an excess burden in that market illustrated by the dark blue triangle in the left-hand panel of Figure 2.2. But it is necessary to consider the further effects of the introduction of the casino tax in the market for lottery tickets. Assuming that casino gambling and lottery ticket purchases are substitutes, the new tax on casino games causes the demand curve for lottery tickets to shift rightward. That rightward shift creates a social welfare gain of the amount of the shaded rectangle, including both the existing excess burden triangle and the additional area illustrated. The reason for

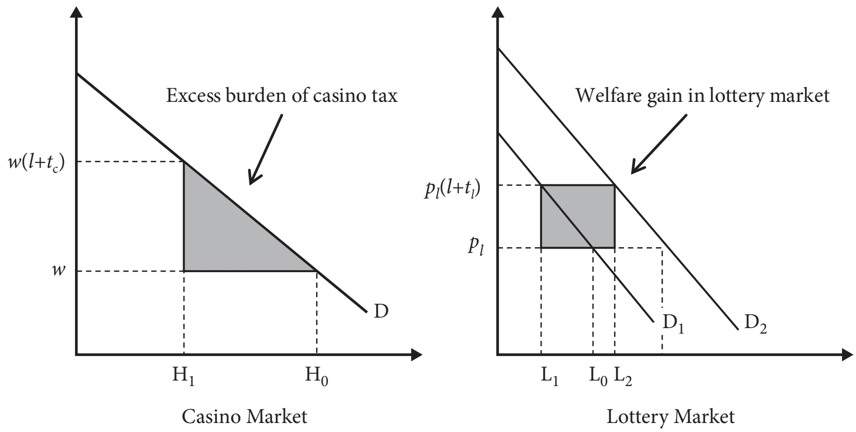


FIGURE 2.2 Effect of a casino tax with a preexisting lottery

this social welfare gain is that the amount that lottery ticket purchasers pay, $p_l\ (1 + t_l)$, for each of the additional L_2–L_1 tickets they buy exceeds the cost of providing those tickets, p_l. Hence, there is a social gain of the amount of the shaded rectangle with height $p_l t_l$ and base L_2–L_1. The net social cost of the new casino tax is then the excess burden triangle in the casino market minus the social welfare gain rectangle in the lottery market. The casino tax introduction in the presence of a preexisting lottery tax is actually less costly in terms of social welfare loss than would have been the case in isolation. In fact, if the welfare gain in the lottery market is sufficiently large, the net cost of the casino tax may actually be negative, or, a welfare gain. This is the standard result from the second-best literature where a policy that may have been inefficient in isolation can be efficient in the face of a preexisting distortion.

The literature to date on casino taxation has not taken a second-best approach to estimating the net excess burden of taxation. Certainly, to examine the social welfare implications of the introduction of a casino tax it is necessary to take into consideration the preexisting distortions in other gaming markets. If that is not done, the estimates of the excess burden of casino taxes will be biased upward. The extent of that bias will depend crucially on the degree of substitutability of the alternate form of gambling. The stronger (weaker) the degree of substitution, the greater (smaller) the extent of the bias.

2.9 Optimal Commodity Taxation (Ramsey Rule)

The classic Ramsey Rule for optimal commodity taxation requires that the tax rates be inversely proportional to the compensated elasticities of demand for the goods being taxed: $t_i/t_j = \vartheta_j/\vartheta_i$. Alternatively, the Ramsey Rule requires that the taxes applied to the goods reduce the demand for each good in the same proportion. Walther (2004) examines the issue of optimal taxation of gambling and lotteries, although his analysis emphasizes lottery games and is confined to optimality for a single form of gambling. Xinhua Gu and Guoqiang Li (2009) consider the factors that may explain why tax rates differ across gaming markets.

2.10 Pigouvian Tax to Correct for Externalities

Concern for the social costs created by casinos has generated proposals to implement corrective taxes. In the tradition of economist A. C. Pigou, who first suggested that a tax (subsidy) can be used to correct for negative (positive) externality, corrective taxes have been suggested as a means of reducing casino gambling and thereby also reducing the spillover or social costs created by casinos. For example, Earl Grinols (2004) has suggested that if a Pigouvian tax were designed to correct for the negative externalities generated by casinos, a tax of 45 to 70 percent of gross casino revenues would be needed. Those estimates do not take into account the excess burden of casino taxation. Douglas Walker (2007) reviewed the social costs of casinos and argued that many of the alleged social costs are inappropriately included or are overestimated. Even so, he does not include the excess burden of taxation in his review or suggest any ways to include this legitimate social cost of casinos.

2.11 Fungibility of Casino Tax Revenue Use

On the expenditure side of the budget, there is the issue of earmarking of gambling revenue and its impact on government budgets. While there is no apparent research on this topic in the literature to date for casinos, experience with lotteries is highly instructive. Garrett (2001) has analyzed the issue of earmarked lottery revenue and found it to be highly fungible. Just because the revenue from a particular source of gambling revenue, including casino games, is dedicated for a particular spending purpose (problem gambler assistance, public education, environmental protection, etc.) does not mean that net spending on that budget category will rise. Borg and Mason (1990) also found that earmarked lottery revenues do not generally benefit the statutory recipients. Charles Spindler (1995) has found that earmarked lottery funds for education in particular are highly fungible.

3 Casino Taxation around the World

Casino gambling has grown substantially in recent years in many parts of the world. While an exhaustive review of the current state of global casino gaming and that growth is beyond the scope of this chapter, it is useful to highlight the forms of taxation used in selected major global casino markets. For a useful overview of casinos around the world, see Thompson 1998. The following section reviews casino taxation in the United States, Singapore, and Macau.

3.1 United States

In the United States casino gambling was first permitted legally in the State of Nevada in 1931 and in New Jersey in 1976. Since those two pioneering states legalized casinos, a number of other states have legalized casino gambling in various forms (riverboat, land-based, and racetrack casinos). Over the very active period 1989–1996 nine states legalized casinos (Colorado, Illinois, Indiana, Iowa, Louisiana, Michigan, Mississippi, Missouri, and South Dakota). Native American casinos operate in 28 states on land sovereignly controlled by the tribes, regardless of state laws regarding the legality of casinos. William Evans and Julie Topoleski (2002) provide a useful overview of the growth of Native American gaming, in particular. The American Gaming Association (AGA) (2011) reports that in 2011 casinos were operating in 38 states, including 438 land-based or riverboat casinos in 15 states, 45 racetrack casinos in 12 states, and 456 tribal casinos in 28 states. The AGA survey also indicates that casinos paid a total of $34.6 billion in taxes in 2010, of which $7.59 billion was paid in direct gaming taxes to state and local governments. For a comprehensive review of casino taxation in the United States, see Anderson 2005. The most recent 50-state review of casino gambling in the United States is provided by the National Conference of State Legislatures (2010).

States apply wagering taxes to casinos along with other taxes on admissions and/or license fees. The most important tax revenue source is the wagering tax. The tax base is some form of adjusted gross receipts (AGR) with some variation in the precise definition across states. In 2011 rates applied to the AGR base varied from a low of 4 percent to a high of 50 percent. Rate structures are flat in some states and graduated in others.

Other than wagering taxes, casinos in the United States may also be subject to other forms of taxation. Riverboat casinos are typically required to charge an admission tax of each gambler. Tax rates generally range from a low of $2 per admission to a high of $5 per admission. In some states, the admissions tax varies with the casino's patronage or the size of the casino facility. In other cases, the admissions tax varies across local government units, as proscribed by the state. Ranjana Madhusudhan (1996, 1999) was the first to document the emerging importance of all forms of gaming revenue among state and local government units in the United States.

Other taxes and fees also may be charged, depending on state statutes. Riverboat casino states, for example, typically impose a licensing fee based on the capacity of the riverboat or a local government licensing fee based on AGR. States permitting land-based casinos have more extensive systems of fees and taxes imposed by both state and local government units. Nevada, home of Las Vegas, has a system of license fees imposed by county governments and a separate slot machine tax. New Jersey, home of Atlantic City, imposes an annual license fee per facility and a slot machine fee. In addition, New Jersey imposes taxes on goods complementary with casino services (such as entertainment, food, hotel rooms, and beverages) provided at reduced or no prices to casino patrons, a tax on gross revenues of companies operating multi-casino progressive slot machines, and a tax on alternative investments applied to casino licenses.

3.2 Singapore

The rise of Singapore as a casino center is a recent phenomenon worthy of attention and analysis. The Inland Revenue Authority of Singapore (2011) reports that casinos are taxed by way of two integrated resorts (IRs): Marina Bay Sands and Resorts World Sentosa. The tax applied to casinos is based on gross gaming revenue (GGR), defined as the difference between the "aggregate of the amount of net wins received on all games conducted within the casino premises of the casino operator" and the goods and services tax (GST) "chargeable to the casino from all gaming supplies made by the casino operator."

Net wins is computed as the total of all bets received by the casino operator on a game minus the amount paid out by the casino operator as winnings on that game. Two tax rates are applied to GGR, depending on the class of players from whom the GGR was generated. A percent tax rate is applied to GGR obtained from premium players, while a 15 percent tax rate is charged from all other players. Premium players are defined as those who maintain a deposit account of at least $100,000 prior to the start of playing any game at the casino. As *The Economist* (2011) reports, "Thanks to low taxes—roughly 17% compared with Macao's 39%—Singapore's casinos are fabulously profitable."

3.3 Macau

Macau (also spelled Macao) is currently the largest casino gambling jurisdiction in the world (Center for Gaming Research 2011). Once a Portuguese colony off the coast of China, it is now a Special Administrative Region (SAR) of the People's Republic of China (PRC), with exclusive rights to have casino gambling, which is illegal in the rest of the PRC. The current gaming regime began in 2002 with an oligopoly of three gaming companies and/or consortia serving the market. Gu and Pui Sun Tam (2011) indicate that the sources of gaming profits in Macau stem from (1) rising demand for gaming opportunities among Chinese consumers, (2) Macau's monopoly position with regard to casinos in China, and (3) the oligopolistic market structure of the casino industry in Macau.

Casinos are taxed using a two-part mechanism in Macau. A variable tax component is based on the gross gaming revenue of the casino, which is taxed at a 35 percent rate. In addition, there are other variable tax components based on GGR requiring 2 and 3 percent contributions for social and economic purposes, respectively. A fixed tax component also is charged of casinos together with a charge per VIP table and other tables and slot machines in the casino. Furthermore, gaming promoters are required to pay taxes on the commissions they receive. In total, the variable tax rate on GGR can be as high as 40 percent.

The UNLV Center for Gaming Research (2011) reports that the effective rate of taxation is 38 to 39 percent. The total effective rate comprises a 35 percent base rate plus a 1.6 percent contribution to the Macau foundation and a 1.4 percent contribution required of SJM casinos (a division of the former monopoly entity permitted to operate casinos in Macau, Societdade de Turismo e Diversões de Macao—STDM) or 2.4 percent required of other casinos. Revenues generated from casinos in Macau were estimated to have been over $15 billion in 2011, with the primary game source being baccarat games (both from VIP players, the major source, and other players). Gu and Zhicheng Gao (2006) have chronicled the rise of Macau as a major gaming center and provided a critique of the development of the gaming industry in that location. Their major policy recommendation is for diversification of the casino industry, relying less on table games and more on other services. Gu and Tam (2011) provide analysis of casino taxation in Macau, with an emphasis on the structure of the industry and its implications for tax policy. William Thompson and Christopher Stream (2006) provide lessons for Asian policy makers based on the American experience with casinos.

4 Summary and Conclusions

In this chapter I have examined the variety of ways that governments around the world tax casino gambling and considered the economic effects of such taxation. The most common form of casino taxation is a wagering tax whose base is the adjusted gross receipts (AGR) collected by the casino. In some cases, governments also apply a casino entry fee. Both efficiency and equity aspects of this form of taxation are considered in theory in this chapter, and empirical evidence from the literature on the elasticities of demand and supply has been presented to provide insights on the economic effects of casino taxation. On the supply side, a regulatory environment that limits the supply of casino games causes the supply to be relatively inelastic, making a casino tax reasonably efficient and causing the incidence of the tax to fall primarily on the casino, not gamblers. But, with the more recent proliferation of gaming opportunities in general, and casinos in particular, the supply is becoming increasingly elastic, which may be altering these initial indications. On the demand side, the evidence reviewed indicates relatively price elastic behavior by gamblers, but as the industry matures this too may be changing with some evidence indicating that casinos may be moving to the unitary elastic point of their demand curve.

This review suggests five areas where more research is needed in order to inform policy decisions regarding casinos. First, the apparent high price elasticities of demand may limit the effectiveness of additional gambling taxes as a source of revenue for governments. If casinos are operating at the unitary elastic point on the demand any tax rate increase that raises the price will reduce tax revenue. If governments have already set tax rates on casinos to generate maximal revenues any further rate increases will reduce revenues. More research is required to know definitively whether the maturing of the casino industry in various countries together with the regulatory and taxing environment has led to both maximal casino and tax revenue. Second, the evidence of relatively high price elasticities points to the availability of illegal gambling alternatives. With that, further tax increases may simply drive gambling activity out of the legal casino sector into the illegal gambling sector. Further research is needed on the general equilibrium effects of taxation in the legal casino industry. Third, the issue of optimal casino taxation is deserving of further attention. To date, efforts to consider optimal gaming taxation have been limited primarily to lottery games. Fourth, the gaming taxation literature to date has not taken a second-best approach to estimating the net excess burden of taxation, recognizing that a casino tax may be adopted in the presence of existing gaming taxes (e.g. lottery tax). Fifth, viewing a casino tax as a Pigouvian tax to correct for externalities, more research is needed to accurately estimate the size of the externalities involved and design optimal corrective taxes.

References

American Gaming Association (AGA). 2011. *State of the states: The AGA survey of casino Entertainment*. Washington, D.C.: American Gaming Association; http://www.americangaming.org/sites/default/files/uploads/docs/sos/aga-sos-2011.pdf.

Anders, Gary C., Donald Siegel, and Munther Yacoub. 1998. "Does Indian casino gambling reduce state revenues? Evidence from Arizona." *Contemporary Economic Policy* 16(3): 347–355.

Anderson, John E. 2005. "Casino taxation in the United States." *National Tax Journal* 58: 303–324.

Anderson, John E., and James R. Schmidt. 2002. "Playing powerball?" National Tax Association—Proceedings of the Ninety-Fifth Annual Conference:377–382.

Benar, Hasret, and Glenn P. Jenkins. 2008. "The economics of casino taxation." *Applied Economics* 40:63–73.

Borg, Mary, and Paul Mason. 1990. "Earmarked lottery revenues: Positive windfall or concealed redistribution mechanism?" *Journal of Education Finance* 15:289–301.

Borg, Mary, Paul Mason, and Stephen Shapiro. 1991. The incidence of taxes on casino gambling: Exploiting the tired and poor." *American Journal of Economics and Sociology* 50:323–332.

Clotfelter, Charles T., and Philip J. Cook. 1987. Implicit taxation in lottery finance. *National Tax Journal* 40:533–546.

Eadington, William R. 1999. The economics of casino gambling." *Journal of Economic Perspectives* 13(3):173–192.

The Economist. 2011. Sin galore: A strait-laced state puts the vice into "service economy." February 26; http://www.economist.com/node/18231454.

Elliott, Donald S., and John C. Navin. 2002. Has riverboat gambling reduced state lottery revenue? *Public Finance Review* 30(3):235–247.

Evans, William N., and Julie H. Topoleski. 2002. *The social and economic impact of Native American casinos."* NBER Working Paper 9198. Cambridge, Mass.: National Bureau of Economic Research.

Farrell, Lisa, and David Forrest. 2008. Measuring displacement effects across gaming products: A study of Australian gambling markets. *Applied Economics* 40: 53–62.

Fink, Stephen, Alan C. Marco, and Jonathan C. Rork. 2004. Lotto nothing? The budgetary impact of state lotteries." *Applied Economics* 36(21):2357–2367.

Forrest, David O., David Gulley, and Robert Simmons. 2000. Elasticity of demand for UK national lottery tickets. *National Tax Journal* 53:852–863.

——. 2010. The relationship between betting and lottery play. *Economic Inquiry* 48:26–38.

Garrett, Thomas A. 2001. Earmarked lottery revenues for education: A new test of fungibility. *Journal of Education Finance* 26:219–238.

——. 2003. *Casino gambling in America and its economic impacts.* St. Louis, Mo.: Federal Reserve Bank of St. Louis.

General Accounting Office (GAO). 2000. *Impact of gambling: Economics effects more measureable than social effects.* Washington, D.C.: U.S. GAO; http://www.gao.gov/new.items/gg00078.pdf.

Gerstein, Dean, et al. 1999. *Gambling impact and behavior study*. Report to the National Gambling Impact Study Commission; http://govinfo.library.unt.edu/ngisc/reports/gibstdy.pdf.

Grinols, Earl L. 2004. *Gambling in America: Costs and benefits.* Cambridge: Cambridge University Press.

Gu, Xinhua, and Guoqiang Li 2009. Why do various gaming markets adopt different tax rates? *Journal of Gambling Business and Economics* 3:65–87.

Gu, Xinhua, and Pui Sun Tam. 2011. Casino taxation in Macao: An economic perspective. *Journal of Gambling Studies* 27:587–605.

Gu, Zheng, and Jason Zhicheng Gao. 2006. Financial competitiveness of Macau in comparison with other gaming destinations. *UNLV Gaming Research & Review Journal* 10:1–12.

Haas, Mark, Denise Heidt, and Andrew Lockwood. 2000. *Estimating Michigan lottery revenues when new casinos open.* In *Proceedings of the Ninety-Second Annual Conference on Taxation.* Washington D.C.: National Tax Association, 191–197.

Harberger, Arnold C.. 1974. *Taxation and welfare.* Boston: Little Brown and Co..

Hoffman, John. 1999. *Analysis of the casino survey: Report to the National Gambling Impact Study Commission.* Chicago: National Opinion Research Center at the University of Chicago.

Inland Revenue Authority of Singapore, Casino Tax (2011); http://www.iras.gov.sg/irasHome/page.aspx?id=10030.

Johnson, Jim. 1985. Gambling as a source of government revenue in Australia. In *Gambling in Australia*, edited by Geoffrey Caldwell, Brian Haig, Mark Dickerson, and Louise Sylvan. Sydney: Croom Helm, 78–93.

Leven, Charles, and Donald Phares. 1998. Casino gambling in Missouri: The spending displacement effect and net economic impact. In *Proceedings of the 90th Annual Conference on Taxation.* Washington, D.C.: National Tax Association, 431–438.

Madhusudhan, Ranjana G. 1996. Betting on casino revenues: Lessons from state experiences. *National Tax Journal* 49:401–412.

——. 1999. What do we know about casino taxation in the United States? In *Proceedings of the 91st Annual Conference on Taxation.* Washington D.C.: National Tax Association, 85–95.

Mason, Paul M., and Harriet Stranahan. 1996. The effects of casino gambling on state tax revenue. *Atlantic Economic Journal* 24(4:336–348.

Mason, Paul, Stephen Shapiro, and Mary Borg. 1989. Gaming tax incidence for three groups of Las Vegas gamblers. *Applied Economics* 21(9):1267–1277.

Matheson, Victor A., and Kent R. Grote. 2004. Lotto fever: Do lottery players act rationally around large jackpots? *Economics Letters* 83:233–237.

National Conference of State Legislatures (NCSL). 2010. *Gambling Developments in the States, 2010.* Denver: NCSL; http://www.ncsl.org/issues-research/economic-development-trade-cultural-affairs/gambling-developments-in-the-states-2010.aspx.

Paton, David, Donald S. Siegel, and Leighton Vaughan Williams. 2001. Gambling taxation: A comment. *Australian Economic Review* 34:437–440.

——. 2003. The demand for gambling: A review. In *The Economics of Gambling*, edited by Leighton Vaughan Williams. London: Routledge, pp. 247–263.

——. 2004. Taxation and the demand for gambling: New evidence from the United Kingdom. *National Tax Journal* 57(4):847–861.

Perez, Levi. 2009. *The state of empirical research on the demand for lottery*. Economic Discussion Paper 01/2009. Oviedo, Spain: Department of Economics, University of Oviedo; http://www.academia.edu/226171/The_state_of_empirical_research_on_the_demand_for_lottery.

Popp, Anthony V., and Charles Stehwien. 2002. Indian casino gambling and state revenue: Some further evidence. *Public Finance Review* 30(4):320–330.

Rivenbark, William C. 1998. The tax incidence of casino gaming in Mississippi. *Public Finance Review* 26(6):583–598.

Sauer, Raymond D. 2001. The political economy of gambling regulation. *Managerial and Decision Economics* 22(1–3:5–15.

Shonkwiler, J. S. 1993. Assessing the impact of Atlantic City casinos on Nevada gaming revenues. *Atlantic Economic Journal* 21):50–61.

Siegel, Donald, and Gary Anders. 1999. Public policy and the displacement effects of casinos: A case study of riverboat gambling in Missouri. *Journal of Gambling Studies* 15(2): 105–121.

——. 2001. The impact of Indian casinos on state lotteries: A case study of Arizona. *Public Finance Review* 29(2):139–147.

Smith, Julie. 1999. *Australian gambling taxation.* Discussion Paper No. 402. Canberra: Centre for Economic Policy Research, Australian National University.

——. 2000. Gambling taxation: Public equity in the gambling business. *Australian Economic Review* 33:120–144.

Spindler, Charles. 1995. Robbing Peter to pay Paul? *Public Budgeting and Finance* 15:54–62.

Suits, Daniel B. 1977. Gambling taxes: Regressivity and revenue potential. *National Tax Journal* 30(1):19–35.

——. 1979a. Elasticity of demand for gambling. *Quarterly Journal of Economics* 93(1): 155–162.

——. 1979b. Economic background for gambling policy. *Journal of Social Issues* 43():43–61.

——. 1982. Gambling as a source of revenue. In *Michigan's Economic and Fiscal Structure*, edited by Harvey E. Brazer and Deborah S. Laren. Ann Arbor: University of Michigan Press, 828–853.

Thalheimer, Richard, and Mukhtar M. Ali. 2003. The demand for casino gaming. *Applied Economics* 35:907–918.

Thompson, William N. 1998. Casinos de juegos del mundo: A survey of world gambling. *Annals of the American Academy of Political and Social Science* 556:11–21.

Thompson, William N., and Christopher Stream. 2006. Casino taxation and economic development: American lessons for Asian policymakers. In *Law, regulation and control issues of the Asian gaming industry: Proceedings of the first conference of the Asian Academic Network on Commercial Gaming*. Macau: Institute for the Study of Commercial Gaming, University of Macau, 39–82.

UNLV Center for Gaming Research. 2011. "Macau gaming summary." Las Vegas: UNLV Center for Gaming Research; http://gaming.unlv.edu/abstract/macau.html.

Walker, Douglas M. 2007. *The economics of casino gambling*. New York: Springer.

Walther, Herbert. 2002. "Economics of gambling." Unpublished manuscript. Vienna University of Economics and Business Administration.

———. 2004. Optimal taxation of gambling and lotto. Working Paper No. 47. Vienna: Vienna University of Economics and Business Administration.

CHAPTER 3

THE ELASTICITY OF CASINO GAMBLING

MARK W. NICHOLS AND MEHMET SERKAN TOSUN

Introduction

Casino gambling, once an industry dominated by Las Vegas, Nevada, is now available in many developed countries around the world. Indicative of this transformation is the recent surpassing of Las Vegas as the world's largest gambling market, measured by total gross revenue, by Macau (Macao). Singapore, which first opened casinos in 2006, is the world's third largest market, just behind Las Vegas.

A primary factor driving the expansion of gambling is its ability to raise tax revenue. While casino gambling is now legal in many parts of the world, it is frequently restricted to certain locations and a fixed number of licenses. The restriction on casinos is generally justified on grounds of limiting the public's exposure to gambling, but it also generates economic rents due to the associated market power. Thus, as with any industry where competition is restricted, prices may higher than those that exist in a more competitive market.

However, while competition *within* a particular location may be restricted, competition among locations has grown more intensive as gambling has expanded. Indeed, one justification for casino legalization and expansion is competition from nearby locations as jurisdictions attempt to attract gamblers from out of the region or prevent their own citizens from travelling elsewhere to gamble.

The efficacy of casino gambling, in terms of generating tax revenue, depends on many factors, including the size of the market, market structure and the competitive environment, quality of the product offered, and the industry's growth potential over time. One commonality that impacts all of these factors is the price elasticity of demand. Size, market structure, and quality of the product all will impact this elasticity Price elasticity will influence a jurisdiction's ability to generate tax revenue by changing tax rates. Similarly, the long-term growth of the industry and its ability to contribute tax revenue over time will depend on its growth relative to the economy as a whole, that is, the income elasticity of demand. Of course the behavior of individual gamblers, the size of consumer surplus, and the impact of, and effect on, competing forms of gambling also will depend on these factors. A knowledge of the price, cross, and income elasticity of demand for gambling therefore has important implications for the debate about the economic and social impact of casinos.

This review examines the price, cross, and income elasticity of demand for casino gambling. The focus is specifically on casino gambling as opposed to other forms of gambling, such as lottery or pari-mutuel wagering. Nevertheless, in the case of price elasticity, comparisons with other forms of gambling are made to provide a context for the reader. Similarly, in the case of income elasticity comparisons are made between casino revenue versus income and sales tax revenue. As will become clear, despite the importance of elasticity to our knowledge about the revenue generating capacity of casino gambling, gambler behavior, and casino gambling's impact on other forms of gambling and vice versa, very few empirical studies have calculated elasticity estimates for casino gambling. One notable reason for this is the unavailability of accurate data on the quantity and price of casino gambling. Reasons for this and existing studies are reviewed here and recommendations for future research are made.

Price Elasticity of Demand

Walk into any retail location and prices are nearly always published or available by asking a sales clerk. Pay that price and you will have purchased the product. Walk into any casino and what you will see is denominations reported on electronic gaming devices (EGD) such as slot, fruit, or video poker machines equal to, for example, \$1, 25¢, or £1. Table games may similarly provide information on minimum and maximum allowable wagers. Pay the monetary unit necessary to play the EGD or bet on a number or color in roulette and you may walk away with nothing or more than your initial wager. Clearly, then, the information posted on a slot machine or table game is not the price. So what is the price of casino gambling?

The Price of Casino Gambling

The price of casino gambling is known as the house advantage (Eadington 1999) or equivalently as one minus the expected value of a wager (Gulley and Scott 1993; Forest, Gulley, and Simmons 2000; Paton, Siegel, and Vaughan Williams 2004). This is what a customer can expect to lose on average over the long run. The house advantage can vary by numerous factors, including the type of game, player skill, rules of the game, and type of wager being made. The house advantage can vary also between casinos or even between games within a particular casino. Ask a casino dealer the house advantage for placing a numbers bet on roulette and the dealer may tell you it is 2.70 percent for single zero and 5.26 percent for double zero. Ask the same question at a blackjack table and you will likely be told it depends on your knowledge of basic strategy or, less likely, your ability to count cards. Ask what the house advantage is for a particular EGD and it is almost guaranteed that the employee will not know or be unwilling to tell you.

Many casino games are what are known as fixed-odds or pure chance games, where both the outcome and payoff are predetermined or beyond the control of the gambler. Slot machines, one of the most popular EGDs, are an excellent and popular form of fixed-odds games. Once the spin or play button is pushed a computer generates a random number that determines the outcome, which in turn corresponds to a fixed payoff to the gambler. This cannot be manipulated by the gambler and neither can the payout percentage programed into the computer chip. Roulette and craps, where the outcome depends on the spin of a wheel or roll of the dice and payoffs are predetermined, also are fixed-odds games.

Games such as blackjack and EGDs such as video poker are a mix of fixed-odds or pure chance and skill. In these instances players can alter the outcome or change the amount of their wagers as the game progresses. For example, a player dealt two eights in blackjack could keep that hand and stay, take another card, or split the hand, thereby playing two hands and placing a wager on both. Actions taken by the gambler in this case will influence what the player can expect to lose in the long run and hence the price of the game.

Complicating matters further is the fact that the house advantage, while frequently unknown, is almost certain to differ from the amount gamblers lose in the short run. The total amount wagered is known as handle or turnover.[1] The total amount of money lost by the player, or alternatively won by the casino, is known as win or gross revenue. Win as a percentage of handle is frequently used as a measure of the price of gambling, particularly for EGDs, but clearly may not equal the house advantage. Complicating matters even further is the fact that handle is unknown for table games. Rather, what is known is the total amount of money exchanged for chips at the table, known in the industry as drop. For example, if a gambler purchases \$20 worth of chips at a black jack table, makes and loses four wagers of \$5 and decides to stop playing, the drop from that gambler will equal \$20, as will handle. If however, the gambler wins a few hands and it takes, say, eight hands to lose the initial \$20, drop will still equal \$20, but handle will have been \$40 (8 hands × \$5 bet per hand). Thus, drop will differ from handle, and clearly hold percentage, defined as the percent of the drop won by the casino, also will differ from the house advantage.

Table 3.1, taken from Eadington (1999), demonstrates the house advantage for selected games as well as the standard deviation from that advantage after 1,000 wagers. For example, after 1,000 number wagers on European roulette a player, on average, should be down 27 units (0.027*1000), but the standard deviation is 182.1. Thus, after 1,000 wagers there is a 95 percent probability that the player will be somewhere between

391 units down and 337 units up ($-27+/-2^*182.1$). For a typical slot machine holding 5 percent, a player, on average, would be down 50 units after 100 wagers but might be as many as 720 units down or 670 units ahead.

When measuring price in a casino, the researcher is unlikely to have data on the true price. The price of EGDs, determined by a computer chip, is unknown to both researcher and gambler. Similarly, the price of such games as blackjack varies by player skill and rules of the game. What generally is available to the researcher is information on total win (revenue won by the casino or lost by the gambler) and handle (total wagers made), although for table games only drop rather than handle is known. Using this information, the researcher is able to estimate price for game *j* at time t using the following simple formula:

$$\text{Price}_{j,t} = \text{Revenue}_{j,t}/\text{Handle}_{j,t} \tag{3.1}$$

For EGDs the data on both revenue and handle will be accurate as these are accounted for by computer. For table games, as demonstrated in the example above, it is unlikely that handle, which is a function of the number of bets made and the amount wagered during each of those bets, will be known. Moreover, as demonstrated in table 3.1, there is significant variation around the true price (house advantage) in the short run. While in the long run price should converge to the house advantage, at any point in time the estimation of price is potentially quite noisy. Moreover, it is difficult to determine

Table 3.1 Statistical Properties of Select Casino Games and Devices (1 Unit Wager)

Game	House Advantage[d]	Standard Deviation (1 Wager)[d]	Standard Deviation (1,000 Wagers)[d]	Standard Deviation (House Advantage after 1,000 Wagers)[d]
Craps[a]	1.41%	1.0	31.6	3.16%
Blackjack[b]	0.50%	1.1	34.8	3.48%
Roulette (American)[c]	5.26%	5.7	179.8	17.98%
Roulette (European)[c]	2.70%	5.8	182.1	18.21%
Baccarat[a]	1.25%	1.0	31.6	3.16%
Pai Gow Poker[b]	2.50%	1.0	31.6	3.16%
Video Poker[b,e]	2%	2.3	73.7	7.37%
Slot Machines[e]	5%	10.6	335.2	33.52%
Keno	28%	42.3	1336.3	133.63%

[a] Standard wager.
[b] Assuming player plays optimal strategy with typical house rules.
[c] Single number wager.
[d] Approximate.
[e] Typical values. Slot machine and video poker house advantages vary by many factors, including denomination, location, and competition, among others.

Source: Eadington (1999).

whether changes in price over time are due to changes in the house advantage due to new computer chips, new games with different house advantages, differences in player luck, changes in the competitive environment, or variation in the types of games played by gamblers. Higher denomination games generally have lower hold percentages, so if more wagers were made on high denomination machines in one time period relative to another, price may appear lower, even though no changes in actual price took place.

In summary, unlike most other products or services, the posted nominal denomination of the game is not the price of that game. Rather, the price is equal to the house advantage. For some games, particularly EGDs, this price is unknown to the gambler or researcher. For other games, blackjack and video poker, for example, it will vary with player skill. Estimates of the price are nevertheless available using equation (3.1) above. However, it is only an estimate and will vary over time for reasons other than changes in the house advantage. Thus the seemingly simple issue of the price elasticity of casino gambling is in practice rather difficult in large part because price is measured noisily. In cases where it is known, for example, roulette, price may not vary over time or the quantity of bets, that is, handle, is likely unknown. Nevertheless, several studies have estimated the price elasticity of EGDs using the price approximation given by equation (3.1). Those are now reviewed.

Methodology and Estimates of Price Elasticity

There are relatively few studies that estimate the price elasticity of demand for casino gambling, in no small part due to limitations on the availability of price and quantity data described above. The few studies that have estimated the price elasticity of demand have generally limited their investigation to EGDs. To estimate elasticity, variations on the following simple demand equation are estimated:

$$Q_{j,t} = \beta_0 + \beta_1 \mathrm{P}_{j,t} + \beta_2 \mathrm{INC}_{r,t} + \sum \beta_k \mathrm{P}_{k,t} + \sum \theta_c \mathrm{X}_{c,t} + \varepsilon_{j,t} \tag{3.2}$$

where $Q_{j,t}$ represents the quantity of gambling, usually handle, or the natural log of handle, for unit j at time t. This could be estimated for an individual game or casino, or such geographic areas as states or countries. $P_{j,t}$ is the price, generally measured as win percentage, or the natural log of win percentage, as given in equation (3.1) for unit j at time t. $INC_{r,t}$ is a measure of income at time t corresponding to the country, state, or market area where the casino is located. $P_{k,t}$ is price of alternative goods and could include, for example, the price of lottery games, pari-mutuel wagering, or even alcohol and cigarettes, which often are thought of as complements to casino gambling. $X_{c,t}$ is a vector of various other characteristics that researchers frequently include, such as variables to account for operating times, regulatory structure, competitive environment, and so on.

Due to the difficulty obtaining price data described above, there is a paucity of studies, particularly academically refereed papers, that examine the price elasticity of demand.

Table 3.2 Price Elasticity Estimates for Various Forms of Gambling

Author(s)	Location(s)	Elasticity Estimate(s)
Casino (EGDs)		
Thalheimer and Ali 2003	Iowa, Illinois, Missouri (USA)	−1.5 (1991); −0.90 (1998)
Landers 2008	Iowa, Illinois, Missouri, Indiana (USA)	−0.75 to −0.87
BERL 1997	New Zealand	−0.85
Swan 1992	New South Wales (Australia)	−1.9
Lottery		
Gulley and Scott 1993	Kentucky, Massachusetts, Ohio (USA)	−1.15, −1.92, −1.20
Farrell et al. 1999	United Kingdom	−1.05 (short run); −1.55 (long run)
Forrest et al. 2000	United Kingdom	−1.03
Beenstock and Haitovsky 2001	Israel	−0.65
Lin and Lai 2006	Taiwan	−0.142
Yu 2008	Canada	−0.672
Horse Racing		
Suits 1979	Nevada (USA)	−1.59
Morgan and Vasche 1982	California (USA)	−1.30
Thalhiemerand Ali 1995	Ohio, Kentucky (USA)	−2.85, −3.06, −3.09
Betting Shops		
Paton et al. 2004	United Kingdom	−1.59, −1.62

Table 3.2 provides a summary of the few papers that exist, including some prepared by private consultants. Richard Thalheimer and Mukhtar Ali (2003) were among the first to provide estimates of the price elasticity of demand. Due to the unavailability of handle for table games, Thalheimer and Ali restricted their analysis to slot machine handle. They examined the slot machine demand for riverboat casinos in three states in the United States, Iowa, Illinois, and Missouri, for the period 1991–1998. In addition to price (win percent), they also included per capita income, number of slot machines, number of table games, days of operation, government restrictions (betting limits, loss limits, and boarding restrictions[2]), and market conditions (access to customers and customer access to competitors).

Thalheimer and Ali reported that price decreased over their sample period from 0.4 percent in 1991 to 6.1 percent in 1998. At the sample average of 7 percent, the price elasticity of demand was found to be unitary. Given the reduction in price over time, they found that price became less elastic over their sample period, from −1.5 in 1991 to −0.9 in 1998.

Jim Landers (2008) employed a similar methodology to estimate a fixed effects panel model for 50 casinos operating in Illinois, Indiana, Iowa, and Missouri between 1991 and 2005. As with Thalheimer and Ali 2003, handle was measured as per capita slot machine handle, price as win percent. Landers also included data on per capita income, days of operation, number of table games, number of slot machines, loss limits, and cruising requirements.

Win percentage in Landers's sample averaged 6.99 percent, ranging from a low of 4.71 percent to a high of 12. percent. Estimating numerous double log models, Landers found price elasticity of demand ranging from −0.75 to −0.87. Neither estimate, however, is statistically different from unitary elasticity at the 10 percent level. Similarly, by including lags of price, Landers found a long-run elasticity of −1.0. Unlike Thalheimer and Ali, Landers, interacting the price variable with a time trend, did not find that elasticity changes over time.

Few others have estimated the price elasticity of demand. Business and Economic Research Limited (BERL) (1997) derived a price elasticity estimate of −0.85 for electronic gaming machines and casino gambling in New Zealand, while Peter Swan (1993) estimated a price elasticity of −1.7 for poker machines and −1.9 for casino gambling in New South Wales, Australia.

Generally, the above studies suggest that the demand for EGDs is either unitary or slightly inelastic. Thalheimer and Ali found demand to be slightly elastic initially but later falling to slightly inelastic. This is plausible as competition in those states increased over their sample period and casinos reduced their price, as demonstrated by John Navin and Timothy Sullivan (2007), in the St. Louis, Missouri metropolitan area. Moreover, the fact that consumers are not aware of the true price and frequently are loyal to certain products or casinos also supports a slightly inelastic demand (Paton, Siegel, and Vaughan Williams 2003). Indeed, most casinos have loyalty programs that reward gamblers with comps, generally in the form of free entertainment, food, or lodging. In addition to creating loyalty, this also may reduce the perceived price for the gambler.

In addition to the problems due to the unavailability and uncertainty of price and quantity of casino gambling, there are other problems that potentially arise with estimation of price elasticity. Price in equation (3.2) above, as with any demand equation, is endogenous, leading to potentially biased estimates of its coefficient. Moreover, given the definition of price in equation (3.1), if handle is measured inaccurately, equation (3.2) will also suffer from division bias (Borjas 1980).

Another obstacle to obtaining accurate estimates of the price elasticity of demand is the aggregate nature of the data collected. Handle and price are generally averages across multiple casinos in different jurisdictions with different competitive environments. Economic theory predicts that a casino with a linear demand curve operating as a local monopoly will have higher prices and elasticity than will casinos operating in more competitive environments. Aggregation hides this variation. Moreover, handle and price are aggregated across machines of different denominations. The elasticity of demand for a $1 machine may be different from that for a $5 machine. One solution, of course, is to analyze individual markets or players, making more precise estimates possible.

Access to individual data, however, is difficult due to its proprietary nature. Sridhar Narayanan and Puneet Manchanda (2008) conducted the only study of which we are aware that examines individual gambling behavior using data from a U.S. casino operating in the southwestern United States.[3] The data are from the casino operator's player reward program, and Narayanan and Manchanda have data for 198,223 gamblers over the period January 2004 through December 2005. Using a random sample of 2,000 gamblers comprising 15,632 gambling trips, Narayanan and Manchanda have a very rich data set consisting of data on total time spent gambling, a history of all wagers made, cumulative win, as well as comps awarded by the casino. However, while Narayanan and Manchanda provide elasticity estimates with respect to comps (0.14), last bet (−0.28), last loss (0.133), last win (−0.020), cumulative loss (−0.245), and cumulative win (0.030), they do not provide any estimates of price elasticity.

Cross-Price Elasticity Studies

Around the world, governments have turned to various forms of gambling as a means of increasing tax revenue, including lotteries, casino gambling, and to a lesser extent pari-mutuel wagering. However, does the introduction of one form of gambling substitute for the other? Governments need to know if this is the case if they are going to maximize tax revenue.

In the United States, pari-mutuel wagering, both horse and dog racing, has experienced a prolonged and significant decline. One response has been to allow slot machines or video lottery terminals at racetracks. These racinos, as they are commonly known, are intended, in part, to increase demand for racing by using revenue from the casino to subsidize purses, under the assumption that larger purses should attract better horses and more wagers.

Thalheimer (1998, 2008) estimated demand equations for pari-mutuel gambling and found that video lottery terminals are strong substitutes for pari-mutuel wagering. An increase in the number of video lottery terminals is associated with a substantial reduction in pari-mutuel handle. For example, relaxing government restrictions on the number, location, and maximum allowable bets on video lottery terminals was estimated to reduce pari-mutuel handle in the long run by 49 percent at Mountaineer Park in West Virginia. Moreover, both Thalheimer (2008) and Ali and Thalheimer (2002) found low elasticities of handle with respect to purse size, 0.08 and 0.15, respectively. Thus even if EGDs enable larger purses, the impact on total wagering handle is small.[4]

Studies on the degree of substitution between lottery and casino gambling are more numerous and varied but generally find casino gambling and lotteries to be substitute forms of gambling. For example, Donald Elliott and John Navin (2002) found that for every additional dollar of per capita revenue from riverboat gambling, state lottery per capita revenues decline by $1.38. Based on average tax rates, this equates to a loss of $0.83 in net lottery revenue. Stephen Fink and Jonathan Rork (2003) also found a substitution effect between lottery and casino gambling, with each additional dollar of casino gambling tax revenue reducing net lottery proceeds by $0.56. On the other hand, Mehmet Tosun and Mark Skidmore (2004) found a complementarity between video lottery terminals and lottery games in West Virginia border counties. However, in the interior West Virginia counties, video lottery was still a substitute for lottery games.

Donald Siegel and Gary Anders (2001) found that a 10 percent increase in slot machines at Indian casinos is associated with a 3.8 percent decline in lottery revenue and a 4.2 percent decline in lotto revenue, with seasonally adjusted figures equaling 2.8 percent and 3.7 percent, respectively. David Giacopassi, Mark Nichols, and Grant Stitt (2006) found that lottery play in Shelby County (Memphis), Tennessee, which is adjacent to a large number of casinos in Tunica County, Mississippi, is approximately $10 lower per month per eligible gambler (over 18 years of age) relative to other counties in Tennessee. Craig Landry and Michael Price (2007) also found that casino gambling is a substitute for lottery, with states that earmark lottery revenues seeing a 3 percent reduction in per capita lottery expenditures due to casino legalization versus a 17 percent reduction for general fund lottery states.

The substitutability of different forms of gambling has also been done within casinos. Thalheimer and Ali (2008a) found that an increased number of table games on riverboat casinos reduce slot machine revenues but that total casino revenue increases. Ina Levitzky, Djeto Assane, and William Robinson (2000), in contrast, estimated a model of gambling revenue in Las Vegas and found that the introduction of more table games reduces total gambling revenue.

Recently the issues of smoking and alcohol consumption and their complementarity to casino gambling have been explored. Chad Cotti and Douglas Walker (2010), using data from all counties in the United States, found that the introduction of casino gambling increases the number of alcohol-related fatal car crashes, but this increase is restricted to rural counties, with more urban counties experiencing a decrease. Thalheimer and Ali (2008b) found that the smoking ban implemented in Delaware resulted in a 15.9 percent reduction in handle. Similarly, Thomas Garrett and Michael Pakko (2010) showed that the ban on smoking in casinos in Illinois reduced gambling revenue by 20 percent and attendance by 10 percent, suggesting that gamblers made both fewer visits and gambled less.

The above studies demonstrate that casino gambling and lottery as well as casino gambling and pari-mutuel wagering are substitutes. Smoking and alcohol consumption, in contrast, are complements to casino gambling. None of the above studies, however, provide traditional cross-price elasticity estimates. Of the studies reviewed above, only Thalheimer and Ali 2008b used handle as the dependent variable. All others analyzed casino revenue. No studies examined changes in the price of lottery, pari-mutuel wagering, cigarettes, or alcohol that would yield traditional cross-price elasticity estimates.

Income Elasticity Studies

While there is a small literature on the income elasticity of casino gambling, those studies follow a larger literature on the income elasticity of state and local taxes. One of the first such papers, by Harold Groves and Harry Kahn (1952), used double-log OLS specification to estimate long-run income elasticity of various state taxes using annual tax revenue data. Thomas Cargill and William Eadington (1978) and David Babbel and Kim Staking (1983) followed this by examining income elasticity of gambling, although only Cargill and Eadington examined casino gambling specifically using California personal income to show the responsiveness of gambling revenue in Nevada to regional income changes.[5] In addition, Thalheimer and Ali (2003) and Landers (2008) examined how casino demand has responded to changes in market and state income in the riverboat states of Iowa, Illinois, Missouri, and Indiana.

More recent studies of the income elasticity of state taxes distinguish between growth and variability of tax bases by separately estimating long-run and short-run elasticities.[6] Nichols and Tosun (2008) built on this broader literature to estimate the short- and long-run income elasticities of casino gambling but made several improvements and contributions. First, they used quarterly data, as in Fox and Campbell (1984), but expanded the analysis to a number of states (Colorado, Connecticut, Illinois, Indiana, Iowa, Louisiana, Mississippi, Missouri, Nevada, New Jersey, and South Dakota) instead of just one. Second, they used data on the actual tax base for the first time in the literature, thus removing the potential error inherent in previous studies that used proxies. Third, they added a new estimate of the income elasticity of gross casino gambling revenues to the list of past elasticity estimates on such state taxes as the individual income tax, general sales tax, corporate income tax, motor fuel tax, tobacco tax, and alcohol tax. Finally, they also examined the responsiveness of the tax base to changes in regional income in the vicinity of the state and changes in national income. This is important because casino gambling revenues might be quite sensitive to visitors from the state's region or even from the entire nation, as in the case of Nevada.

Methodology and Income Elasticity Estimates

Long-Run Elasticity

The basic model used to estimate the long-run elasticity of demand is given by

$$R_{j,t} = \beta_0 + \beta_1 INC_{j,t} + \varepsilon_{j,t}, \tag{3.3}$$

where $R_{j,t}$ is the natural log of gross casino gambling revenue for country or state j at time t and $INC_{j,t}$ is a measure of income, such as the natural log of personal income, for country or state j at time t. The coefficient on $INC_{j,t}$ provides the income elasticity of demand, thereby predicting the long-run responsiveness of the tax base to income. Nichols and Tosun (2008) further used the dynamic OLS (DOLS) estimator with heteroskedasticity and autocorrelation consistent (HAC) standard errors, and estimated the following equation with Newey-West (1987) standard errors:

$$R_{j,t} = \beta_0 + \beta_1 INC_{j,t} + \beta_2 SLOTS_{j,t-2} + \beta_3 TABLES_{j,t-2} + \beta 4 S_t + \sum_{t=-m}^{n} \Delta INC_{j,t} + \varepsilon_{j,t}, \tag{3.4}$$

where $SLOTS_{j,t-2}$ is the natural log of the number of slot machines in state j at time $t-2$, that is, lagged two quarters; $TABLES_{j,\,t-2}$ is the natural log of the number of tables games in state j at time $t-2$; S_t represents seasonal dummies for spring, summer, and fall to account for potential seasonal variation in gambling revenue; and $\Delta INC_{j,t}$ is the change in the natural log of income with the number of lags and leads determined using the Bayesian Information Criterion (Stock and Watson 2007). $SLOTS_{j,t-2}$ and $TABLES_{j,t-2}$ are included to ensure that the impacts on the tax base from relaxing regulatory constraints, such as a new or expanded casino, or a change in the mix of slots versus tables, are not attributed to a change in income.

Short-Run Elasticity

Nichols and Tosun (2008) followed Bruce, Fox, and Tuttle (2006) to derive short-run elasticity estimates using an Error-Correction Model (ECM) allowing for asymmetric income elasticity and adjustment to equilibrium. In the short-run, changes to the tax base may come from changes in income or an adjustment toward the long-run co-integrating relationship derived from equation (3.4) above, both of which may differ depending on whether the actual tax base is above or below the long-run value. Hence they estimated short-run elasticities estimated using the following model:

$$\begin{aligned}\Delta R_{j,t} = {} & \beta_0 + \beta_1 \Delta INC_{j,t} + \beta_2 \Delta SLOTS_{j,t-2} + \beta_3 \Delta TABLES_{j,t-2} + \beta_4 S_t \\ & + \beta_5 (D_{j,t}{}^{*} \Delta INC_{j,t}) + \beta_6 \varepsilon_{j,t-1} + \beta_7 (D_{j,t-1}{}^{*} \varepsilon_{j,t-1}) + \mu_{j,t}\end{aligned} \tag{3.5}$$

where variables are described as above and $D_{j,t} = 1$ if $\varepsilon_{j,t} > 0$ in equation (3.4). $\varepsilon_{j,t-1}$ is the error correction term and β_6 captures the adjustment in period t to the disequilibrium in period $t-1$, that is, the difference between the last period's actual tax base and the long-run co-integrating relationship predicted by equation (3.4). The inclusion of the interaction term, $D_{j,t-1}{}^{*}\varepsilon_{j,t-1}$, allows for this adjustment to differ depending on whether the actual tax base is above or below its long-run value.

Cargill and Eadington (1978) analyzed seasonally adjusted data for the period 1960–1974 and found that the long-run income elasticity of gross gambling revenue is fairly elastic with significant variation across three regions in Nevada. The highest is in the Las Vegas region (1.75), followed by the Lake Tahoe (1.25) and the Reno-Sparks (1.05)

Table 3.3 Long-Run State, Regional, and National Income Elasticity Estimates for Casino Revenue, Short-Run Regional Income Elasticity and Long-Run State Income Elasticity for Sales and Income Tax

	State Income	Regional Income	National Income	Short-Run Regional Elasticity	Adjustment to Long-Run Equilibrium	Sales Tax[a]	Income Tax[a]
Destination Resorts							
Nevada	0.30***	0.44***	0.50***	0.56	−0.46***	0.78**	1.03***
(1983:q2:2006:q2)	(0.035)	(0.075)	(0.080)	(0.42)	(0.10)		
New Jersey	0.38***	0.46***	0.35***	0.76**	−0.41***	1.05**	2.01**
(1985:q1:2006:q2)	(0.05)	(0.06)	(0.05)	(0.32)	(0.09)		
Mississippi	2.05***	1.15***	1.53***	−0.64	−0.28**	0.48**	1.91**
(1994:q4–2005:q2)	(0.57)	(0.40)	(0.35)	(1.14)	(0.11)		
Riverboat Casinos							
Iowa	1.28***	1.49***	1.05***	2.92***	−0.97***	0.37**	2.35**
(1995:q3–2006:q2)	(0.36)	(0.20)	(0.21)	(0.66)	(0.04)		
Illinois	1.72**	2.10***	1.85***	0.66	−0.15*	0.87**	1.56**
(1995:q3–2006:q2)	(0.83)	(0.70)	(0.66)	(1.11)	(0.07)		
Missouri	2.37***	2.32***	1.85***	1.19	−0.34**	0.64**	2.29**
(1995:q1–2006:q2)	(0.33)	(0.26)	(0.23)	(0.81)	(0.13)		
Louisiana	1.36***	0.66**	0.69**	2.04	−0.54	0.51**	2.27**
(1995:q3–2005:q2)	(0.17)	(0.25)	(0.29)	(2.40)	(0.44)		
Indiana	0.53	0.02	−0.30	1.71**	−0.67***	0.47***	2.43**
(1997:q1–2006:q2)	(0.47)	(0.66)	(0.26)	(0.92)	(0.15)		
Mining Towns							
Colorado	1.27***	1.42***	1.58***	1.01**	−0.33***	0.78**	1.26**
(1993:q2–2006:q2)	(0.16)	(0.22)	(0.32)	(0.57)	(0.09)		
South Dakota	1.25***	1.43***	1.31***	0.22	−0.47***	1.15**	1.03***
(1990:q2–2006:q2)	(0.14)	(0.19)	(0.15)	(1.03)	(0.10)		
Indian Casinos							
Connecticut	1.36***	1.28***	0.96**	1.18	−0.20	1.24**	0.96***
(1995:q3–2006:q2)	(0.25)	(0.24)	(0.43)	(0.87)	(0.12)		
(1) Fixed Effects (AR1, no slots or table games)	1.18*** (0.17)						
(2) Fixed Effects (AR1, with slots & table games)	0.76*** (0.23)						

Note: Newey-West standard errors in parentheses. All time series regressions, except for Louisiana and New Jersey, include number of slot machines and table games as control variables. For Louisiana and New Jersey data on the number of slot machines were not available. For those states significant regulatory changes are controlled with dummy variables. Data on table games were not available for Colorado and Connecticut.

[a] Income and sales tax elasticities are taken from Tuttle, Bruce, Fox, and Tuttle 2006 and Holcombe and Sobel 1997. Elasticities from Holcombe and Sobel 1997 are used for Indiana (sales tax) and Nevada, South Dakota, and Connecticut (income tax).

[b] Regression (1) does not include number of slots or number of table games and is run with data on ten states excluding only Nevada. Regression (2) includes number of slots and number of table games but is run with data on six states, excluding New Jersey, Colorado, Connecticut, and Louisiana, for which there no slots and/or table games data are available. In regression (2) we used number of slots and table games lagged two quarters to be consistent with time series regressions.

*, **, and *** represent significance from zero at the 10, 5, and 1% level, respectively.

Source: Nichols and Tosun (2008).

regions. Thalheimer and Ali (2003) did not explicitly estimate income elasticity but noted that the percentage change in handle decreases up to real per capita income of $16,500 (which is greater than the mean of $15,000), then increases. They also note that a greater proportion of income is wagered at both low and high levels of income relative to the middle. Landers (2008) found elasticity estimates ranging from 1.4 to 1.9, although none were statistically different from unity.

Long-run and short-run income elasticity estimates from Nichols and Tosun 2008 are shown in table 3.3. Results for states are grouped under four types of industry structures for casinos: destination resort casinos (Nevada, New Jersey, and Mississippi), riverboat casinos (Iowa, Illinois, Missouri, Louisiana, and Indiana), mining town casinos (Colorado and South Dakota), and Indian casinos (Connecticut). The table also presents separate state elasticity estimates based on time series regressions and overall long-run elasticity estimates from panel regressions. Finally, the long-run income elasticity estimates for casino gambling are compared to long-run income elasticity estimates for two major state revenue sources, the personal income tax and the sales tax. The overall long-run income elasticity estimates from the fixed effects regressions are 1.18 and 0.76 depending on whether slots and tables are included as control variables in equation (3.4). Both estimates are statistically equal to unity and are consistent with the earlier results of Cargill and Eadington (1978) and Landers (2008). However, estimates do vary substantially across and within market structures. The main results can be summarized as follows: (1) long-run income elasticities for casino gambling are relatively high and more similar to the ones for the state personal income tax; (2) more mature casino markets, such as Nevada and New Jersey, show significantly lower long-run elasticity estimates, which could indicate lower growth potential as casino programs age; (3) Nevada's casino gambling is more sensitive to national income changes, whereas casino gambling in other states responds more to state and regional income changes; (4) short-run income elasticity estimates are generally lower than long-run estimates, but those estimates are significantly higher when they are separated into below long-run equilibrium and above long-run equilibrium of casino gambling revenue; and (5) there appears to be a fairly rapid adjustment to long-run equilibrium of casino gambling revenue.

Conclusions

Casino gambling, much like the lottery, has greatly expanded over the last two decades. This trend has been driven, in large part, by governments seeking new sources of tax revenue. Gambling is somewhat unique in that governments generally tax a portion of revenue rather than profit. Therefore, governments would clearly prefer that operators maximize revenue, a goal that may conflict with profit maximization. Similarly, governments prefer a stable and growing tax base over time. To understand whether these objectives can be achieved, policy makers require accurate estimates of the price and income elasticity of casino gambling.

Despite the important policy implications of these figures, there are few academic studies that provide estimates of the elasticity of casino gambling. The estimates of price elasticity from Thalheimer and Ali (2003) and Landers (2008) reviewed above, in general converge on unity, suggesting that casinos are, on average, maximizing revenue. While excellently done, both studies derive their estimates from riverboat casino states in the United States. Do these estimates generalize to other jurisdictions? Are there significant differences in price elasticity between government owned (e.g., Canada, Holland) and privately owned casinos? How do elasticity estimates vary by type of game, game denomination, competitive environment, and industry structure (e.g., tourist oriented versus locals oriented)? How do they vary by specific income and demographic groups? What are the consequences of continued casino expansion and the expansion of on-line gambling?

Similarly, the income elasticity estimates reviewed here also tend to be unity or higher, except for the mature markets of Nevada (casinos legalized in 1931) and Atlantic City, New Jersey (casinos legalized in 1976 and operating in 1978), where the estimates are significantly lower. Does the growth in casino revenue slow as markets mature? What is the potential for revenue growth in the face of new competition from other jurisdictions and the Internet? How does casino revenue behave over the business cycle, and how has it responded to the most recent recession, the most severe recession since many casinos have been legalized?

These are several important policy questions that academics should explore. Particularly interesting would be micro-oriented studies that analyze data on individual gamblers. This could best be accomplished by acquiring data from player loyalty programs offered by many casinos. While anonymity would have to be preserved, such data would allow differences in elasticity by volume of play, type of game, gender, and ethnicity, to name only a few, to be estimated. Alternatively, researchers may be able to generate data by conducting experiments. This would be a very effective way to overcome the unavailability of handle data for table games and to observe how it varies with price. To date, Walls and Harvey (2005) is the only study of which we are aware that conducts such an experiment, changing price and observing betting patterns in an experimental roulette game.

Finally, there is need for more study on competition among various forms of gambling. Does the introduction of casino gambling impact lottery sales, and does this impact vary over time as, perhaps, the price of casino gambling declines? As Internet gambling grows in popularity how will traditional casinos be impacted, and how will they respond? How does the introduction of casino gambling by one jurisdiction impact another, and how does it impact the potential for revenue generation? While the European Union has allowed each country to retain its own unique gambling policy, similar to individual states in the United States, would a broader regional or national approach to gambling result in a more efficient use of resources and an increase in revenue?

The academic community has only recently begun to undertake the study of casino gambling as it would any industry. The lack of a studies estimating elasticity is evidence of this. Further research is critical if we are to understand gambler behavior and the revenue-generating potential of an activity that, while newly legalized, has been around perhaps as long as humankind. It is an industry that poses many fascinating academic and public policy questions and is deserving of continuing academic inquiry.

Notes

1. In the United States the term *handle* is used for total wagers, whereas in many other parts of the world, for example, Australia and South Africa, the term *turnover* is used.
2. Riverboat casinos were often required to "cruise." In Iowa and Illinois the boats would take a two-hour excursion, after which they would return to the dock in order to allow new customers to board and existing customers to exit if they chose. Existing customers were allowed to remain for another cruise if they chose but had to pay another admission fee (generally $2). In Missouri boats that were on rivers had the same restrictions. Others, denoted as "boats in moats," were on land next to the river. They had water that surrounded them and had to mimic cruising. Customers were given a certain time to enter the casino but during the next two hours were prohibited from entering the casino. Customers could, however, leave whenever they chose. Over time states dropped these restrictions. For a history and analysis of these and other restrictions see Nichols 1998.
3. The name and precise location of the casino are confidential. Narayanan and Manchanda (2008) describe the casino as a local monopoly with the nearest competitor 30 miles away.
4. Thalheimer (2008) notes that the larger purses did allow the racetrack to sell the simulcasts of its live races. The elasticity estimates above refer only to handle from live races at the racetrack.
5. On the other hand, there is a much broader literature on the income elasticity of lottery expenditures. One of the earlier studies by Babbel and Staking (1983) finds that lotteries have close to unitary income elasticity. Other studies find very different elasticity estimates, including negative income elasticity for West Virginia in Garrett and Coughlin 2009. See Babbel and Staking 1983, Mikesell 1989, 1994, Garrett and Coughlin 2009, and Coughlin and Garrett 2009 for examples of those different income elasticity estimates for lotteries.
6. See Fox and Campbell 1984, Dye and McGuire 1991, Sobel and Holcombe 1996, Holcombe and Sobel 1997, Dye 2004, and Bruce, Fox, and Tuttle 2006 for examples of those studies.

References

Ali, Mukhtar M., and Richard Thalheimer. 2002. "Product choice for a firm selling related products: A parimutuel application." *Applied Economics* 34: 1251–1271.

Babbel, David F., and Kim B. Staking. 1983. "An Engel Curve analysis of gambling and insurance in Brazil." *Journal of Risk and Insurance* 50(4):688–696.

Beenstock, Michael, and Yoel Haitovsky. 2001. "Lottomania and other anomalies in the market for lotto." *Journal of Economic Psychology* 22:721–744.

BERL (Business and Economics Research, Ltd.). 1997. Sensitivity analysis of gross win to price elasticities of demand. In Responsible gaming: A Commentary by the New Zealand Lotteries Commission on the Department of Internal Affairs' proposals for gaming and gambling. Contained in *Gaming: A new direction for New Zealand, and its associated reports.* Wellington: New Zealand Lotteries Commission.

Borjas, George. 1980. The relationship between wages and weekly hours of work: The role of division bias. *Journal of Human Resources* 15(3):409–423.

Bruce, Donald, William F. Fox, and M. H. Tuttle. 2006. Tax base elasticities: A multi-state analysis of long-run and short-run dynamics. *Southern Economic Journal* 73(2):315–341.

Cargill, Thomas F., and William R. Eadington. 1978. Nevada's gaming revenues: Time characteristics and forecasting." *Management Science* 24(12):1221–1230.

Cotti, Chad D., and Douglas M. Walker. 2010. The impact of casinos on fatal alcohol-related traffic accidents in the United States. *Journal of Health Economics* 29:788–796.

Coughlin, Cletus C., and Thomas A. Garrett. 2009. Income and lottery sales: Transfers trump income from work and wealth. *Public Finance Review* 37(4):447–469.

Dye, Richard F. 2004. State revenue cyclicality. *National Tax Journal* 57(1):133–145.

Dye, Richard F., and Therese J. McGuire. 1991. Growth and variability of state individual income and general sales taxes. *National Tax Journal* 44(1):55–66.

Eadington, William R. 1999. The economics of casino gambling. *Journal of Economic Perspectives* 13(3):173–192.

Elliott, Donald, and John Navin. 2002. Has riverboat gambling reduced state lottery revenue? *Public Finance Review* 30(3):235–247.

Farrell, Lisa, Edgar Morgenroth, and Ian Walker. 1999. A time series analysis of UK lottery sales: Long and rhort run price elasticities. *Oxford Bulletin of Economics and Statistics* 61(4):513–526.

Fink, Stephen, and Jonathan Rork. 2003. The importance of sself-selection in casino cannibalization of state lotteries. *Economics Bulletin* 8(10):1–8.

Forrest, David, David Gulley, and Robert Simmons. 2000. Elasticity of demand for UK national lottery tickets. *National Tax Journal* 53(4, pt.1):853–864.

Fox, William F., and Charles Campbell. 1984. Stability of the state sales tax income elasticity. *National Tax Journal* 37(2):201–212.

Garrett, Thomas A., and Cletus C. Coughlin. 2009. Inter-temporal differences in the income elasticity of demand for lottery tickets." *National Tax Journal* 62(1):77–99.

Garrett, Thomas A., and Michael R. Pakko. 2010. The revenue performance of casinos after a smoking ban: The case of Illinois. Working Paper No. 2009–027B. St. Louis, Mo.: Federal Reserve Bank of St. Louis.

Giacopassi, David, Mark W. Nichols, and B. Grant Stitt. 2006. Voting for a lottery. *Public Finance Review* 34(1):80–100.

Groves, Harold M., and C. Harry Kahn. 1952. The stability of state and local tax yields. *American Economic Review* 42(1):87–102.

Gulley, David, and Frank Scott. 1993. The demand for wagering on state-operated lotto games." *National Tax Journal* 46(1):13–22.

Holcombe, Randall G., and Russell S. Sobel. 1997. *Growth and variability in state tax revenue.* Westport, Conn.: Greenwood.

Landers, Jim. 2008. What's the potential impact of casino tax increases on wagering handle: Estimates of the price elasticity of demand for casino gaming. *Economics Bulletin* 8(6):1–15.

Landry, Craig E., and Michael K. Price. 2007. Earmarking lottery proceeds for public goods: Empirical evidence from U.S. lotto expenditures. *Economics Letters* 95:451–455.

Levitzky, Ina, Djeto Assane, and William Robinson. 2000. Determinants of gaming revenue: Extent of changing attitudes in the gaming industry. *Applied Economics Letters* 7:155–158.

Lin, Chin-Tsai, and Chien-Hua Lai. 2006. Substitute effects between lotto and big lotto in Taiwan. *Applied Economics Letters* 13:655–658.

Mikesell, John L. 1989. "A note on the changing incidence of state lottery finance." *Social Science Quarterly* 70(2): 513–521.

Mikesell, John L. 1994. "State Lottery Sales and Economic Activity." *National Tax Journal* 47(1): 165–171.

Morgan, Douglas W., and Jon David Vasche. 1982. A note on the elasticity of demand for wagering. *Applied Economics* 14:469–474.

Narayanan, Sridhar, and Puneet Manchanda. 2008. An empirical analysis of individual level casino gambling behavior." Research Paper No. 2003. Stanford, Calif.: Stanford University.

Navin, John C., and Timothy S. Sullivan. 2007. Do riverboat casinos act as competitors? A look at the St. Louis market. *Economic Development Quarterly* 21(1):49–59.

Newey, Whitney K., and Kenneth D. West. 1987. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* 55(3):703–708.

Nichols, Mark W. 1998. Deregulation and cross-border substitution in Iowa's riverboat gaming industry. *Journal of Gambling Studies* 14:151–172.

Nichols, Mark W., and Mehmet S. Tosun. 2008. The income elasticity of gross casino revenues: Short-run and long-run estimates. *National Tax Journal* 61(4):635–652.

Paton, David, Donald S. Siegel, , and Leighton Vaughan Williams. 2003. The demand for gambling: A review. In *The economics of gambling*, edited by Leighton Vaughan Williams. Routledge: London, pp. 247–264.

———. 2004. Taxation and the demand for gambling: New evidence from the United Kingdom. *National Tax Journal* 57(4):847–861.

Siegel, Donald, and Gary Anders. 2001. The impact of casinos on state lotteries: A case study of Arizona. *Public Finance Review* 29(2):139–147.

Sobel, Russell S., and Randall G. Holcombe. 1996. Measuring the growth and variability of tax bases over the business cycle. *National Tax Journal* 49(4):535–552.

Stock, James H. and Mark W. Watson. 2007. *Introduction to Econometrics*. Pearson: Boston.

Suits, Daniel. 1979. The elasticity of demand for gambling. *Quarterly Journal of Economics* 93:155–162.

Swan, Peter L. 1993. Report on the likely effects of slot machines in a casino on the operations and viability of the registered club and hotel industries. . Sydney: Chief Secretary of New South Wales.

Thalheimer, Richard. 1998. Parimutuel wagering and video gaming: A racetrack portfolio. *Applied Economics* 30:531–543.

———. 2008. Government restrictions and the demand for casino and parimutuel wagering. *Applied Economics* 40:773–791.

Thalheimer, Richard, and Mukhtar M. Ali. 1995. The demand for parimutuel horse race wagering and attendance. *Management Science* 41(1):129–143.

———. 2003. The demand for casino gaming. *Applied Economics* 35:907–918.

———. 2008a. Table games, slot machines, and casino revenue. *Applied Economics* 40:2395–2404.

———. 2008b. The demand for casino gaming with special reference to a smoking ban. *Economic Inquiry* 46(2):273–282.

Tosun, Mehmet S., and Mark Skidmore. 2004. Interstate competition and lottery revenues. *National Tax Journal* 57(2):163–178.

Walls, W. David, and Patrick J. Harvey. 2005. Modelling gambling demand in a laboratory casino: Discovering the importance of individual-specific effects. In *Information Efficiency in Financial and Betting Markets*, edited by Leighton Vaughan Williams. Cambridge: Cambridge University Press, 262–276.

Yu, Kam. 2008. *Measuring the output and prices of the lottery sector: An application of implicit expected utility theory*. NBER Working Paper No. 14020. Cambridge, Mass.: National Bureau of Economic Research.

CHAPTER 4

THE ECONOMICS OF ASIAN CASINO GAMING AND GAMBLING

RICARDO CHI SEN SIU

> The chief factor in the gambling habit is the belief in luck; and this belief is apparently traceable, at least in its elements, to a stage in human evolution antedating the predatory culture.
>
> (Veblen 1899)

> . . . the Chinese are a nation of gamblers. . . they hope for a miracle, a streak of luck, a big win.
>
> (Blanchard2002)

Introduction and Approach

Rapid growth in the Asian casino gaming markets since the last quarter of the twentieth century has evidently drawn the interest of academics in the field. Despite a number of socially and politically controversial issues, the evolution and organization of the casino industry with regard to the contextual settings in Asia deserve extensive exploration.

Introduction

While gambling is a common activity for humans, the demand for various types of games and related gaming services is not identical across different regions. Given the differences in social, political, and economic factors, such as culture, education, public administrative structure, regulatory system, income, and so on among different countries in the world, the organization and market performance of casino gaming and gambling in Asia are unique. For example, while the markets in America are largely dominated by gaming machines, it is well-known that Asia is a table-game market (see, e.g., Siu and Eadington 2009). In addition, while most casino firms in the world are competing for patrons from the mass gaming market whereby lower operating risk and higher profit margins are expected, Asian casinos are more inclined to pursue VIP (or premium) players with a considerable volume of turnover but higher liquidity risk and lower profit margins (see, e.g., Siu 2009a).

In Asia, the composition of the casino gaming markets is largely driven by the particular desires of players for various games. Nevertheless, given the fixed odds of casino games, the role of non-price factors in exploring the demand-side issues presents an interesting topic for study. Besides that, given the unique features of this industry in that its practice may easily be correlated with socially and ethically controversial underground economic activities (e.g., crime, drugs, commercial sex, money laundering, etc.), a free market approach may be inappropriate. Indeed, institutions and institutional changes have exerted significant influences in the recent progress and organization of this industry, hence leading to differences in performance across various Asian casino jurisdictions.

In the early decades of the twenty-first century, it is commonly anticipated that increasing income and wealth effects will accelerate the growth and development of casino gaming in Asia. For example, as cited in "Gambling on Asia's middle classes" (GBGC 2011), the World Bank estimated that "the global middle class will increase from $430 million in 2000 to 1.2 billion in 2030. China and India will account for two-third of this expansion." As the two most-populated countries in Asia (also in the world), both China and India would undoubtedly expand the potential demand for casino gaming should they experience a rapid increase in the numbers of people reaching the middle classes. Accordingly, direct and indirect economic benefits are appealing to the host economies and related world investors. Nevertheless, taking into account the underdeveloped social and public administrative systems of most Asian countries, social costs and impacts associated with the potential expansion of casino gaming are also imperative issues for the related parties.

This chapter explores the aforementioned topics. In particular, the determinants of the Asian demand for casino gaming and factors responsible for the evolution and organization of the casino industries will be examined. In addition to the market fundamentals, the significance of the unique features of the Asian culture and related institutional structure of the gaming industrial performances will be uncovered. Lastly, controversial debates over the social benefits and costs of casino gaming in Asia will also be evaluated.

Approach of This Study

Only a limited amount of literature is available as reference for a study of the economics of casino gaming and gambling. That is because this is a relatively new area of study. Research on gambling in Asia is particularly scarce. In congruence with the modern development of casino gaming in Las Vegas since the 1980s and the subsequent expansion of this industry within North America, a few representational studies have been devoted to explore the practice of casino gaming and its economic impacts, such as taxation, social benefits and costs, and economic growth (see, e.g., Walker and Barnett 1999; Eadington 1984, 1999; Garrett and Nichols 2005; Nichols and Serkan 2008; and Walker 2007). Besides that, aware of the rapid expansion of casino gaming in East Asia since the dawn of the twenty-first century (e.g., the success of Macau (Macao) in surpassing the reported gross gaming revenue (GGR) of Las Vegas since 2006 and the approval of two casino licenses by the Singapore government in 2006), Ricardo Siu (2006a and 2007a) and Guoqiang Li, Xinhua Gu, and Siu (2010) also embarked on studies related to the organization of casino gaming and its impacts on economic growth and development. To enhance our understanding in this area with a specific focus on Asia, this chapter will first develop adequate evidence sourced from various contexts in Asia to support the pragmatic research work in the related topics.

In the process of consolidating and analyzing the evidence, related reasoning from institutional economics will follow, especially in uncovering the evolutionary aspects of casino gaming in Asia as well as the forces that drive the development of various major casino jurisdictions. For example, the argument of a ceremonial-instrumental dichotomy that was raised by Thorstein Veblen (1899), with emphasis placed on uncovering the opposing forces between tradition and technological progress in economic changes. Besides that, the situation, structure, and performance (SSP) institutional impact theory (Schmid 1987, 39–43) in revealing the interrelationship between the rights structure in an economic society and its impacts on economic performance, and the nature of path dependence in economic changes (North 1997), are found to

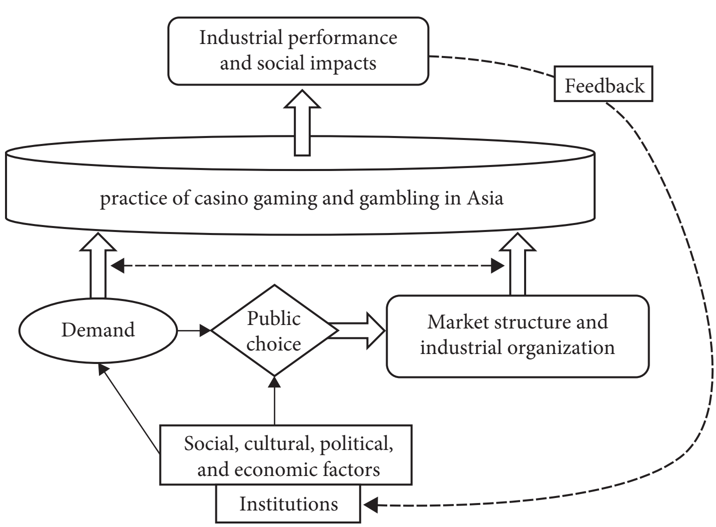


FIGURE 4.1 A pragmatic approach in exploring Asian casino gaming and gambling topics

provide the pragmatic grounds for this study. In principle, by following the institutional economics approach, this study "does not attempt to build an all-embracing general theory" (Hodgson 1998, 168) for casino gaming and gambling. This approach is instead consistent with the argument of Robert Prasch (2008, 6–7), who indicated that "there is no single 'theory of the market'." Instead, the influence exerted on industrial changes and performance by particular "institutions, habits, rules, and their evolution" (Hodgson 1998, 168) will be emphasized. As a recap of the aforementioned, figure 4.1 depicts the organization and flow of this study.

Current Layouts of the Asian Casino Gaming Markets

In comparison with Europe, America, Oceania, and Africa, the evolution of land-based casino gaming in Asia has its own process and hence holds unique features and structure. For example, according to the level of modern development and organization of businesses, the practice of this industry is evidently associated with duality features. Given the unique situation present in various Asian societies, the structure employed by the respective authorities and power groups, alongside the layouts and performance of the related markets, are found to be quite different.

Composition of Land-Based Casino Jurisdictions in Asia

According to the United Nations, Asia is geographically composed of 30 legally recognized independent countries, 1 de facto country (i.e., Taiwan) and 2 Special Administrative Regions (SARs) (i.e., Hong Kong SAR and Macau SAR) of China[1] (see figure 4.2). At the end of 2011, casino gaming has been legalized in 19 of the 33 Asian countries/regions (as labeled by "♠" and "♥" in figure 4.2). Of the other 14, 12 countries/regions have outlawed casino gaming and the governments also enforce related laws to ban the business. For the remaining 2 countries (Indonesia and Uzbekistan, which are labeled with "♦"), casino businesses are operated in the open, even though casino gaming is banned by law. In Indonesia, around 3 to 4 casinos are currently in operation; one is located in Surabaya, the country's second largest city. In Uzbekistan, although casinos have been formally banned since the beginning of the 1990s, at this writing a casino is operating in Tashkent, the capital city.

During the process of reviewing and compiling related information to construct figure 4.2, several particular features associated with the existing composition and layouts of the Asian casino gaming markets were identified. First, religion was found to be a major factor responsible for the lack of casino gaming in a country. For example, casino gaming is outlawed in most Islamic countries in Central and West Asia, such as

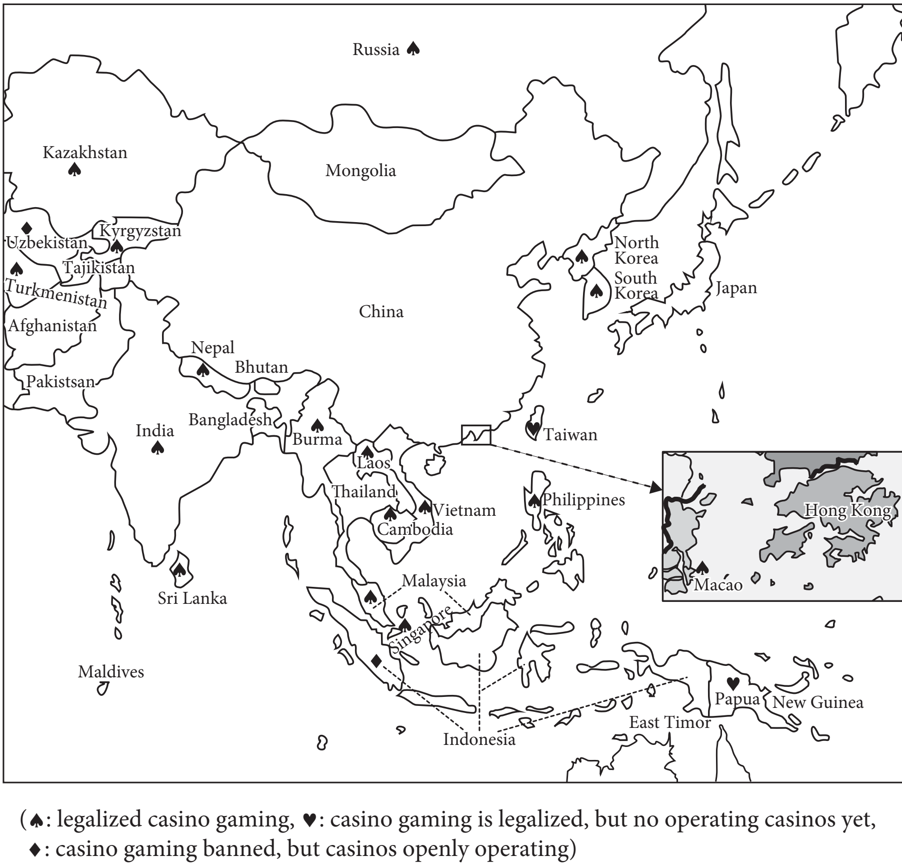


FIGURE 4.2 Land-based casino jurisdictions in Asia (as of Dec. 31, 2011)
Legend: ♠ = legalized casino gaming; ♥ = casino gaming is legalized but no operating casinos yet; ♦ = casino gaming banned but casinos openly operating

Afghanistan, Pakistan, Tajikistan, and the Maldives. The same also applies to Indonesia. In East Timor, the reasons for outlawing casino gaming are unclear even to foreigners, but inasmuch as the majority in the country are dedicated Roman Catholics, religion could possibly be a factor. In addition, moral and ethical concerns of other major Asian religions, such as Hinduism, Buddhism, and Confucianism, also play essential roles in influencing political interests and, hence, the public choice for the physical layouts and magnitude of casino gaming in various countries. While social factors play a part in determining the institutional context, scope, and scale of the related markets at the outset, economic factors lead to the observed performance of the markets.

In the particular social context of Asia, it is recognized that the physical scale of casino gaming in most of the jurisdictions in Central and South Asia are more traditional in organization and small in scale. While casino industries in these countries are generally represented by fewer than 10 firms, the largest casinos commonly located in hotels or hotel resorts are in the scale of around 20 to 40 tables and a few dozen to a few hundred gaming machines. In contrast, the industries are much larger in scale and modern in organization in East and Southeast Asia. Macau, as the world's casino capital, has a geographical area of less than 30 square kilometers, yet the industry comprises more than 30 casinos (in 2011) and approximately 5,000 gaming tables and 14,000 slot machines. The recent development of certain individual mega casino resort properties in Malaysia, Singapore, South Korea, and the Philippines have also reached a scale and organization comparable to those in Las Vegas.

Last but not least, another unique feature associated with the practice in Asian casino gaming markets is the relatively high proportion of table games to gaming machines. While the ratio of gaming tables to gaming machines in a typical North America casino is around 30, it is less than 10 in an Asian casino (see, e.g., Siu and Eadington 2009, 45). In one extreme case, for example, there were only 16 tables and no equipped gaming machines, as in the largest casino in Sri Lanka, the Ritz Club (International Land Casinos Directory—Sri Lanka). This unique composition of gaming devices is likely to persist in the Asian markets for the foreseeable future.

Duality of Markets

Broadly speaking, the term *duality of markets* refers to the existence of two non-parallel segments (e.g., existing traditional sector versus newly developed modern sector) or two opposing forces (e.g., local versus international business practices), which interact with each other and produce the dynamics of market evolution. Indeed, this concept is identical to the ceremonial-instrumental dichotomy described by Veblen (1899), which provides the pragmatic grounds to examine the differences in the evolutionary paths between various economies (or an industry in different economies).

Gambling activities have long been a part of the history and culture in Asia, especially in China. However, related businesses are negatively perceived by the general public and have a low social status. This is due to the fact that most Asian societies are less developed. Thus, they lack an appropriate regulatory system to police the socially undesirable activities associated with gambling businesses. Consequently, stereotypical perceptions are derived from the concept of *gambling dens* instead of casinos or casino resorts. The former is a common term used in the past in Asia to describe a casino-like gaming venue. Indeed, gambling dens operate in the open and are suspected by many as having been triad related or controlled throughout Asian history. Together with underdeveloped regulatory systems in most of the countries, these gambling dens have generated an adequate force to encapsulate the modern development of this industry.

On the other hand, following the rapid integration of Asia (especially East Asia) into the global economy beginning in the last quarter of the twentieth century, improvement of regulatory systems (including transparency and enforcement mechanisms) and political reforms carried out by various Asian governments have opened a new chapter in the development of casino gaming. In addition, triggered by the decision of the

Macau SAR government to eliminate the casino monopoly in 2002, and the decision of the Singapore government to legalize casino gaming in 2005, the casino jurisdictions in East Asia have evidently entered into a phase that looks geared to improving the industrial structure and organization. It is anticipated that the impacts of the related instrumental changes may also be diffused to other countries in Asia. Nevertheless, while the entertaining features of modern casino gaming are incorporated into casinos, traditional components that are associated with less than desired activities continue to take place. In addition, constrained by underdeveloped legal systems in many Asian countries, casino gaming is to a large extent still being carried out in a gray area somewhere between law and custom.

Structural Issues Associated with the Related Markets

With reference to the current layouts of the Asian casino gaming markets, it is possible to pin down a number of structural issues that are associated with/confronted by the industries and related policy makers. Among these issues, the role played by the public sector definitely warrants careful examination. Since a large number of Asian countries have underdeveloped political and administrative systems, the current gaming laws may be either insufficiently constructed or weakly enforced, or subject to ineffective changes due to social and political interests. The operations of this 24/7 type of business, which involves a considerable amount of cash turnover, are difficult to regulate, and consequently it is difficult to safeguard the interests of modern corporate investors.

Given the particular concern around the public sector structure, the next prominent issue is the corporate governance structure of the casino firms. Aside from Singapore and Malaysia, South Korea, and probably the newly developed segment in Macau, which have put forth the required layouts to balance business objectives and community interests, most other casinos in Asia do not seem to have a well-defined or efficient internal governance structure. Pursuant to the current social, regulatory, and economic environments in many Asian countries, the corporate governance structure of many casinos is either largely shaped (or restricted) by public interests or loosely drawn due to underdeveloped regulatory frameworks and enforcement mechanisms.

For example, it is well known that casino gaming is prohibited in Thailand. Consequently, small casinos are made available to Thai patrons in easily accessible neighboring countries, such as Cambodia, Laos, and Burma. These casinos would not be overly keen to have in place a corporate governance structure that would balance business turnovers and social impacts. Accordingly, crime, drugs, commercial sex, and other socially and ethically controversial activities are found to be associated with the related casino gaming businesses. Similarly, as casino gaming is prohibited in Mainland China, legalized casinos that operate in the neighboring countries (e.g., North Korea, Vietnam, Burma, Kyrgyzstan, and Kazakhstan) widely publicize their services to Chinese patrons. Consequently, it has been the ongoing endeavor of the Chinese government to detect cross-border gambling activities and to pressure these neighboring governments to shut down casinos in locations that are easily accessible by the Chinese (e.g., Rank 2005; BBC News 2007).

On the basis of this evidence, it is obvious that the issues associated with revenue (GGR) maximization and harm (social costs and impacts) minimization are of particular concern due to the unique social and administrative settings. Besides that, it is clear also that excess demand on the existing industry in its current scale is a common phenomenon. In order to adequately present the perspective that a free market approach may not be a desirable means of closing the gap in demand, the unique socioeconomic factors that contribute to the Asian demand for casino gambling will be discussed in the next section.

Unique Contributing Factors to the Asian Demand for Casino Gambling

Despite the conceptual robustness of the law of demand (Marshall [1890] 1997, ch. 3; Samuelson and Nordhaus 1995, 24), it is evident that this concept is not necessarily sufficient to reveal the quantity demanded nor applicable enough to directly uncover the demand behavior of casino gambling in practice (Siu 2007a and 2011). A major issue associated with economic studies is that the unit price paid by individual patrons to enjoy a casino game is unobservable. Indeed, once the odds of casino games are fixed, influences from a number of non-price factors (e.g., traits of a society in terms of its culture, religion, political structure, education, etc., as well as income and wealth) may contribute to the decision processes of individuals and endogenously determine their choices (including various types of games and total amount of funds to be spent).

A Socioeconomic Approach to Demand for Casino Gambling

To a large extent it cannot be argued that the contextual settings of a society shape the ideology and value system of its members, hence influencing the representation of their demand behavior for gambling. For example, it is of little dispute that the unique traits of Chinese society contribute to a number of key attributes (especially influences from Buddhism and Confucianism; see, e.g., Nepstad 2000) that induce a relatively higher propensity of the population to gamble in comparison to people in other societies. As a matter of fact, "gambling followed the development of the Chinese society every step of the way and by 1000 B.C. it became as inseparable from Chinese culture" (Progress Publishing Co. 2006).

To account for social influence on the demand for casino gambling, the consumers' utility function in a social environment as formulated by Gary Becker and Kevin Murphy (2000, 9) provides relevant grounds for this effect.

$$U = U(x, y; S) \tag{4.1}$$

In equation (4.1), x and y denote the quantity of all kinds of economic goods and services, and S represents the "social influences" on the utility function through the stocks of "social capital" (e.g., social norms, religion, education, etc.). In principle, "changes in social capital do not shift the utility function, but raise or lower the level of utility with the stable function, U." Thus, if x denotes the quantity of casino gambling and y the quantity of all other economic commodities, one's utility derived from casino gaming (x) depends on whether one's "friends and neighbors" choose to gamble in casinos. In Backer and Murphy's general argument, "x and S are *complements* so that an increase in S raises the marginal utility from x, even when the increase in social capital itself lowers utility." It follows that changes in social capital stocks will alter one's marginal utility for casino gambling in the same direction, that is,

$$\frac{dx}{dS} > 0. \tag{4.2}$$

It is worth noting that, unlike in the traditional approach to demand theory, the price of a casino game may not be directly expressed in the discussion of casino gambling. Accordingly, the demand function for casino gambling of an individual in a social environment is subject to the budget constraint of

$$G + p_y y = I, \tag{4.3}$$

where gross expenditure spent on casino gambling is $G = f(x)$. While x is determined by different variables, such as the number of hands played, G is jointly determined by x and other variables, such as the fixed odds of a game and the average amount of bets.

With reference to the socioeconomic approach to demand for casino gambling, it is evident that as stocks of social capital (S) are heterogeneous across societies and nations, so too is the specific representation of behavioral demand for casino gambling. In the case of Asia, the stocks of social capital are clearly inseparable from the unique composition and traits of culture, religion, education, and so on across various societies/nations.

Influence of Culture and Religion on the Demand for Casino Gambling

As an essential component of social capital, culture incorporates a set of socially prescribed behaviors (e.g., social norms and values, habits, etc.) and social cognition for people to interact with one another. In various societies, although the cultures are not identical, they are determined by a number of factors that have resulted from their respective paths of evolution. For example, geographical location, language, and administrative structure traditionally adopted by different Asian societies may have had substantial impacts on cultural traits. Among these factors, religion is probably the most influential. According to Philip Wilkinson (2008, 14), "virtually every culture that we know of has some kind of religion." In societies, "religions provide a sense of moral purpose.... Religion often occupies the centre of the world stage in politics, diplomacy, and even war.... Despite the changing nature of our times, religion is still at the heart of our lives" (10). Indeed, this social capital stock shapes beliefs about and the desire for casino gambling.

In East and South Asia, the demand behavior for gambling is deeply linked to the region's unique cultures and religions (e.g., Buddhism and its factions, and Hinduism). In addition to the commonly known Chinese culture of gambling (see, e.g., Nepstad 2000; Chien and Hsu 2006), it has also been recorded that "Indian culture adopted gambling from the beginning of Indian civilization, which started 4000 years ago" (Progress Publishing Co. 2006). Consequently, socially prescribed behavior for gambling allows or even promotes the desire for casino gambling. In contrast to Western culture and religion, superstition is the primary explanation for the cognitive preference (and illusion of control) of the Chinese for games carried out by humans instead of machines. This explains the phenomenon in which traditional Chinese patrons and a large proportion of patrons from East Asia are more inclined to choose table games rather than slot machines.

In contrast to East and South Asia, gambling is widely perceived as socially unacceptable behavior in the Muslim world and hence is prohibited (Binde 2005, 5; Wilkinson 2008, 137). Given that there is no endowment of related social capital, casino gambling is obviously not a choice for religious Muslims. Besides that, social influences may generate a negative utility (e.g., anticipated punishments) to individuals who find ways to gamble despite the legal ban. In other words, the utility function as stated in equation (4.2) could be written as:

$$\frac{dx}{dS_R} \leq 0 \text{ for the Muslim world,} \tag{4.4}$$

where S_R denotes the religion of Islam.

Impacts from Education and Societal Modernization

From an evolutionary perspective, "culture is subject to a process of cumulative change. Culture has ceremonial and technological aspects" (Hamilton 2004: 111). That is, throughout the process of cultural change part of a society's culture may be sustained or reinforced while others may be altered or replaced. To a large extent, while religion is an aspect of culture that has ceremonial meaning, education and modernization of a society over time generate technological changes that influence the value system, patterns of life, and market behaviors. For example, in China and a number of Asian countries public policies (including education) have been formulated to drive the community away from smoking. By the same token, education and societal modernization also influence the traditional view on casino gambling but in a direction that differs from that of smoking. While the majority of the views in various societies tend to oppose smoking, casino gambling is increasingly perceived as a form of entertainment rather than a sinful activity. This change well explains the phenomenon of rising Asian demand for casino gaming.

Yet there is research that validates the impacts of changes in education on the demand for casino gambling, with plenty of related studies that look at the demographical features of problem gambling in North America and Australia (see, e.g., State Government of Queensland 2005, figure 6; and Shinogle et al. 2011: table 4.20). The results suggest a positive correlation between the education level of a community and the proportion of its residents who have participated in gambling activities (including casino gambling). Nevertheless, there is no consistent pattern across different countries/regions in terms of whether education has an impact on the frequency of casino visits and amounts spent.

According to a 2005 study by the State Government of Queensland on a group of people with normal behavior who gambled (i.e., classified as "recreational gambling"), 46 percent had an education level of 10 years or less; 32 percent, up to 12 years; and 19 percent, above 12 years. When the same study was applied to people with a higher frequency of gambling (i.e., at "low risk" or "moderate risk" for gambling), the ratio to education level of more than 12 years remarkably fell from the 19 percent to 9 percent and 7 percent, respectively. Interesting in terms of the gambling group with the highest risk (i.e., "problem gambling"), it was found that the ratio to education level of more than 12 years increased all the way up to 22 percent.

On the other hand, with reference to an empirical study conducted by Daniel Lai (2011) on "four ethno-cultural minority groups (i.e., Chinese, Filipinos, South Asians, and Vietnamese) in Edmonton, Calgary, Toronto and Vancouver," it was identified that higher educated Asian groups in Canada are relatively less likely to become addicted to gambling. In addition, other parameters used in this study also suggested that "cultural differences in gambling (behavior) do exist." Despite the fact that the findings reported in most of the studies are not sourced from research intended to elucidate casino gambling in Asia, the results are related enough to fairly verify the possible influence of education level to demand by Asians for casino gambling. It is just a matter of the degree of demand that varies across different Asian societies.

The Role of Income and Wealth Effects

Apart from cultural and social factors, income and wealth serve as the two most important economic factors that directly enter into the decision-making process of individuals for casino gambling demand. Similar to the demand for such luxurious commodities as tourism and recreation, the demand for casino gambling is regarded as income elastic. For example, as measured by the GGRs, double-digit compound annual growth rates (CAGRs) of spending in casino gambling were reported by major casino industries in Asia (e.g., Macau, South Korea, the Philippines, Malaysia) in the first decade of the twenty-first century, with the CAGRs generally being higher than GDP growth in the region. Associated as part of economic development, income growth in Asia is widely correlated with increases in the income of the working class and expansion of the middle class, and this is especially true for China, India, and other emerging East and Southeast Asian countries/regions.

In looking at the various major economic approaches recently employed to examine gambling behavior, the Friedman-Savage utility function (Binde 2009, 26–28) provides a reasonable framework for running analyses. In principle, the Friedman-Savage utility function argues that as wealth of an individual increases "marginal utility first decreases, then increases, and finally decreases again" (27). As such, the "convex middle segment of the (expected utility) curve represents the choices, ... of the upper stratum of individuals belonging to a social group, for whom an increase in wealth will move them up to a higher social group"; therefore, taking the risk of gambling may "allow a working class man or woman to fulfill social aspirations of becoming middle class" (27). Alternatively, while money lost on gambling "does not qualitatively lower a person's social status... the chance of winning substantial sums offers an opportunity for a qualitative social advancement" (27), and therefore, a force is formed which promotes demand for gambling.

Following the success of China's economic reform and the rapid migration of labor-intensive production sectors from East and Southeast Asia to South and Central Asia since the end of the twentieth century, the income of the Asian working class and the size of the middle income group have evidently expanded. For example, Benjamin Chiang and Sherry Lee (2010), estimated that 100 million Chinese people in 2010 had an annual income of around USD20,000, while the annual disposable income of another half a billion people was between USD1,600 and USD4,900.[2] In addition, it was perceived that "during periods of tremendous economic growth in India and China, (the) household saving rate... increased nicely" (Hunkar 2009). As a proportion of household disposal income, China's household saving rate had increased from around 14 percent in 1991 to 28 percent in 2008 while the same rate increased from 23 percent to 32 percent in India. Even though the household saving rate in some of the Asian countries, such as South Korea, largely fell from over 20 percent to around 7 percent in 2008, it was still higher than that in the United States, which had a rate below 4 percent. As instituted by particular Asian cultures and related social factors, the increasing wealth of low-income to middle-income groups moving to the "convex middle segment of the (expected utility) curve" has unarguably been expanding over time.

Overall, non-price factors encapsulate the social choice of gambling in Asia, hence providing indispensable and pragmatic grounds for individuals to consummate their desires for casino gambling. These factors have different degrees of influence across Asian countries/regions, and these factors result in different patterns and degrees of demand for casino gambling. In the foreseeable future, related traits from the demand side will continue to differentiate Asian casino gambling behaviors from those in other parts of the world.

Industrial Organization of Casino Gaming in Asia: Same Principles but Different Practices

Since the dawn of the twenty-first century, casino gaming as a modern industrial sector has been noticeably expanding in the countries/regions of East and Southeast Asia. Following the decision of the Macau government in 2002 to replace its casino monopoly with an oligopolistic structure, and the decision of the Singapore government in 2005 to legalize casino gaming, other jurisdictions, such as Malaysia, South Korea, the Philippines, and Cambodia, also have taken measures to review and promote the organization of their businesses. In addition, such small-scale casino jurisdictions as Vietnam, India, and Sri Lanka, among others, have demonstrated interest in addressing the expanding and excess demand for casino gambling in the markets. In view of the differences in social, institutional, and economic settings, disparities exist even though there are shared similarities in the industrial organization of casino gaming across various Asian countries/regions.

Existing Scale and Structure of the Markets

Based on the number of casinos and reported GGRs, the scale of various Asian casino jurisdictions is shown in figure 4.3.[3] In 2010, the annual outputs (measured by GGR) of most Asian casino jurisdictions are around or below USD1 billion, with the exception of South Korea and the two outliners of Singapore and Macau. In addition, except for Cambodia, Kazakhstan, the Philippines, South Korea, and Macau, other Asian casino jurisdictions are built with around 10 or fewer casinos. Similar to most jurisdictions in the world, entry into the market is highly restricted by the limited number of licenses predetermined by the respective governments, which in turn, are influenced by the interests of various social and political groups rather than pure market forces. For example, while a monopoly structure is approved by the Malaysian government, duopoly is definitely the choice of the Singapore government.

To a certain extent, the differences in the GGRs across various jurisdictions could partially be explained by the variations in their respective physical and absolute

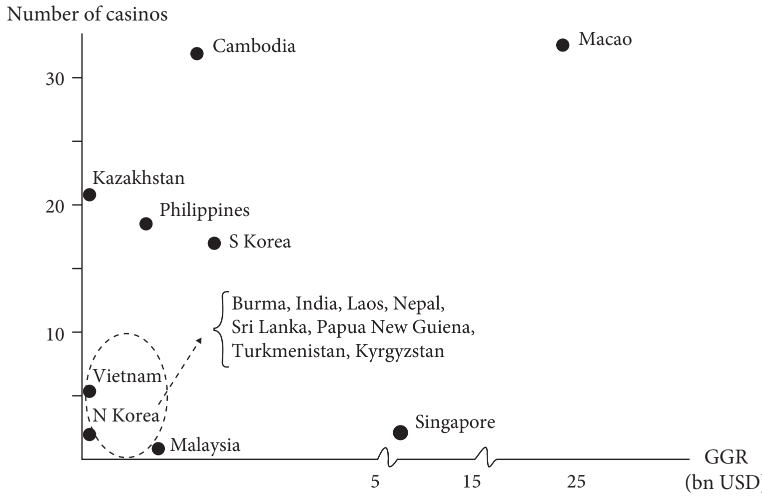


FIGURE 4.3 The scale of various Asian casino jurisdictions (around 2009 and 2010)

configurations (i.e., the scale of gaming devices, associated non-gaming hospitality facilities, etc.).[4] According to some related evidence as presented in table 4.1, it is reasonable to infer that the business turnover of some of the Asian casino industries is somewhat determined by their physical capacities. For example, it is interesting that the physical scale of the industry in Singapore is around one-fifth of that in Macau in 2011,

Table 4.1 Approximate Numbers of Gaming Tables and Machines in Some of the East and Southeast Casino Jurisdictions (around 2009 and 2010)

Casino Jurisdiction	Number of Gaming Tables	Number of Gaming Machines	Source
Macau	~4,900	~14,000	http://www.dicj.gov.mo/web/en/information/DadosEstat/2010/content.html#n4
Singapore	~1,000	~3,000	Curtis, Cheung, and Dobson (2011), 29 See also http://casinocity.sg/
South Korea	~700	~1,800	Kim and Matthew (2011), 15.
Malaysia	~400	~3,100	http://www.ildado.com/land_casinos_malaysia.html
Philippines	~800	~5,200	http://www.ildado.com/land_casinos_philippines.html; http://www.casinocity.com/ph/manila/allistar/

Note: In the same period of time, Las Vegas was equipped with around 3,000 gaming tables and 200,000 gaming machines.

and this is close to the ratio of their reported GGRs. Nevertheless, the relationship between physical configuration and business turnover may not be general enough to explain the differences in business performance across various jurisdictions (e.g., between Singapore and Philippines, South Korea and Malaysia, etc.). In terms of the jurisdictions in Central and West Asia (e.g., India, Nepal, Kazakhstan, and Kyrgyzstan), even though their absolute scales are generally smaller than those in East and Southeast Asia, a general relationship between their physical scale and business turnover is difficult to see.[5]

As an industry that is a socially sensitive and exposed to public desire for extensive government regulation and control, the casino industry is significantly curbed by various social and political inclinations instead of the interest of market participants (risk-seeking casino patrons and profit-seeking casino firms). This simple and unarguable principle for explaining the scale and structure of the casino industry is indeed in line with the phenomenon seen in the Asian markets. Nevertheless, owing to the normative nature of social costs and impacts in public policy discussions, the determinants of the scale of the industry and structure in various jurisdictions vary accordingly.

Changing Dynamics to the Organization of the Industry

Given that the modern progress of casino gaming in Las Vegas can be traced back only to the 1970s (Schwartz 2003), many Asian jurisdictions were seen to be in their developing stages or sporadic in development until the end of the twentieth century. In the new millennium, we have witnessed a change in the organization of traditional gambling-based casino businesses toward the concept of casino resorts. Led by Macau and Singapore, the modern development process of casino gaming in Asia has accelerated.

Because casino gaming is a policy-based industry, the practice and business routines of the industry are largely shaped by public interest and changes at the outset. The dramatic success of Macau, for example, is highly accredited to the full support granted by the Chinese government under its one country, two system policy. Indeed, not only is Macau able to retain its existing casino industry despite the ban in Mainland China, but it also is blessed with continuous support from the Chinese government. Since the elimination of the casino monopoly in 2002, and the approval of new licenses to world-class casino operators, the number of casinos almost tripled from 2002 to 2012. In addition, the scale and organization of the industry as a whole have also undergone significant changes from their traditional layout.

Parallel to the increase in the fixed capacity of the industry and the service quality driven by competition, the number of visitor arrivals from both Mainland China and outside the Mainland to Macau has significantly increased. Under a special visa policy approved by the Chinese government in the summer of 2003 allowing more Chinese residents from wealthier cities in Mainland China to travel to Hong Kong and Macau, visitor arrivals from the Mainland to Macau dramatically increased (from around 2.3 million in 2000 to around 16.2 million in 2011).

Congruent with changing public interests and the expanding accessibility of the market, the organization of the casino firms in Macau (both local and foreign) has evolved over time. On the one hand, it is evident that foreign casino firms from North America and Australia have been adjusting their business models to accommodate this Chinese-based market (especially the composition of gaming devices as well as the practice of competition, etc.). On the other hand, local casino firms have been improving the content of their traditional and monotonic scope of gambling to ensure their competitiveness. Consequently, as compared with other overseas jurisdictions, even though casino gaming in Macau is still largely a table game market (which accounted for over 94 percent of the 2011 GGR in Macau), new aspects of its organization have come into place and led to changes in the performance of this industry. As revealed in figures 4.4 and 4.5, while the number of slot machines has grown at a faster rate than that of gaming tables since 2005, the profitability per slot machine has risen much higher than that of the gaming tables. Indeed, this evidence further verifies the changing trend as identified by Siu and Eadington (2009, 47–54) in that the casino industry in Macau is in a transition stage and is moving toward becoming a high value–added service sector.

Regardless of the differences in business scale, industrial structure, and organization, some of the major Asian jurisdictions have also undergone changes that are similar to those in Macau. In the Philippines, for example, casinos have long been regulated and managed by a government-owned corporation, the Philippine Amusement and Gaming Corporation (PAGCOR). To promote the efficiency and competitiveness of

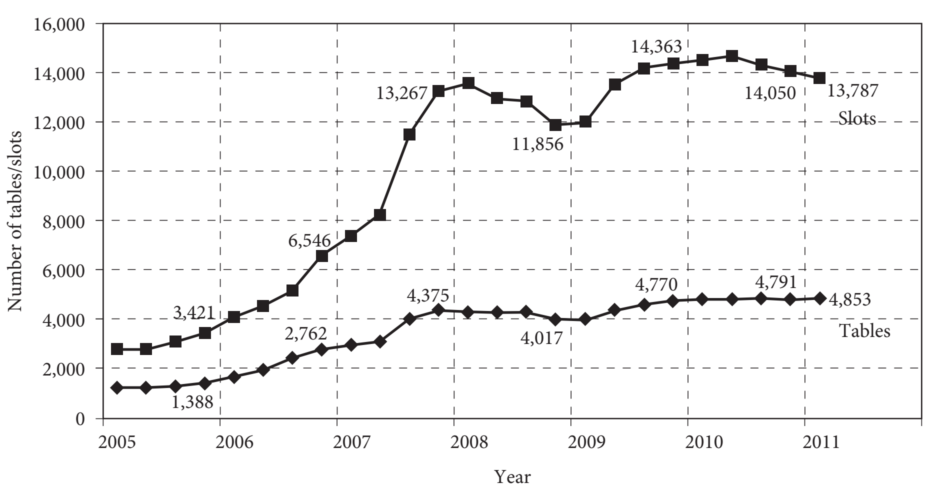


FIGURE 4.4 Changes in the number of tables and slot machines in the casino industry of Macau, SAR, China (2005.1q–2011.1q)

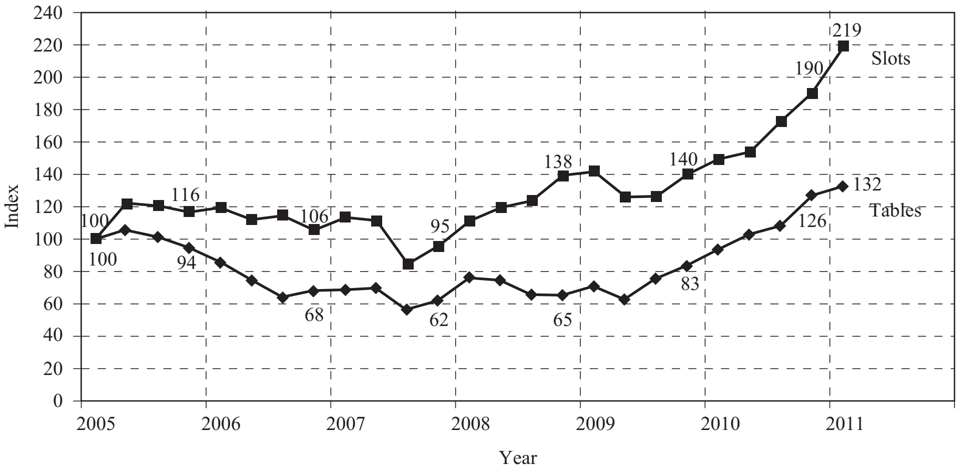


FIGURE 4.5 Changes in the performance of table games and slot machines in the casino industry of Macau, SAR, China (GGR index per table and slot machine: 2005.1q = 100)

this industry in the region, the Philippines congress amended the PAGCOR charter in 2004 to allow the private sector (including local and foreign firms) to take part in casino gaming. Consequently, the Malaysia-based Genting Club, which is well-known in Asia, opened its world-class casino resort (Resorts World Manila) in the capital of the Philippines in 2009. This single property added over 20 percent more to the total number of gaming machines in the industry and more than 50 percent to the number of gaming tables. Together with the enhancement of non-gaming facilities, casino gaming in the Philippines has entered a new stage of development.

In addition, Kangwon Land, which opened in 2000 as South Korea's largest casino resort, is the only casino in the country allowed to receive local customers, and it accounts for over 60 percent of casino revenues in the country. To further diversify, and hence strengthen its competitiveness, Kangwon Land even took progressive steps to improve its organization so that it would be similar to the Genting Highlands in Malaysia and the new casino resorts in Macau. This included a major proposal to further expand its property with luxury condominiums and a world-class water park (Lee 2010b, 22). Indeed, led by Kangwon Land, other existing small- and medium-scale casinos in the country are also improving the organization of their businesses.

Apparently the dynamics generated through the process of changes in organization in major jurisdictions like Macau, the Philippines, and South Korea, as well as the new force formed in Singapore, will further diffuse to other existing and emerging jurisdictions in Asia. Indeed, changing trends in the organization of casino gaming as a particular economic sector that provides entertainment services have emerged. The related changes will continue to be extended to other Asian jurisdictions but at different paces and in different forms subject to the social, political, and economic contexts of the respective countries/regions.

The Practice of Market Competition

That the theoretical prices of casino games are neither observable by patrons nor directly responsive to changes in market conditions (see, e.g., Siu 2011), indicates that the nature of casinos is to interact with their customers and compete in the markets mainly through a set of non-price elements. Under the change in the organization of casino gaming toward that of a modern economic sector which provides a package of gaming and non-gaming entertainment experiences to customers, the differentiation of services has become a common principle in competition. Nevertheless, owing to the aforementioned differences in social, cultural, and political settings across countries/regions, the practice of product differentiation between casino firms as a form of competition varies, and the degree of variation is largely influenced by public interests. For example, the promotion of the gaming business is totally or partially prohibited in some countries. Alternatively, non-gaming attractions (e.g., unique accommodation features such as those featured in the interior design of buildings and the service quality, renowned restaurants and shops, particular types of entertainment and performances/shows, etc.) are the focus of advertisements designed to bring customers to these casinos.

With regard to industrial development stages, competition in the Asian casino markets could be broadly categorized into four different forms. First, there is oligopolistic competition between a few large-scale casino firms within a jurisdiction (e.g., six licensed world-class casino firms in Macau and two in Singapore). Second, one or two major casinos enjoy a major share or a high reputation in the market, accompanied by a number of small- and medium-scale casinos that run their businesses in different locations/segments within the jurisdiction (e.g., South Korea, the Philippines). Third, a number of small- and medium-scale casinos offer gaming opportunities in particular locations and compete for potential patrons at a more modest level (e.g., Cambodia, India, Nepal, Kazakhstan). Fourth and last, subject to the location of a jurisdiction, casinos within the industry as a whole will compete for potential patrons on a geographical basis (e.g., competition between Malaysia and Singapore, Burma and Laos, as well as Macau and other jurisdictions for Chinese patrons).

Regardless of the different forms of competition, their practices share some common features. In the most modernized gaming corporations that are operating in the major jurisdictions, such as Macau, Singapore, Malaysia, South Korea, and so on, casino gaming has been packaged as part of a set of comprehensive entertainment and leisure services supplied on individual properties as unique destinations. Under the guise of the label "casino resorts," however, differences in the facilities installed in the various properties represent important measures to differentiate services from those of competitors. For example, a casino gaming package that is bundled with particular features and services provided by the Wynn Resorts are different from those provided by the Venetian and MGM in Macau. In addition, in order for casinos to remain competitive in various jurisdictions, non-gaming features adopted by the same casino firm also are distinct. This could be verified by comparing the non-gaming features of the Venetian in Macau with those of Marina Bay Sands in Singapore, as both properties are owned and operated by Las Vegas Sands. It is also true when comparing the Genting Highlands in Malaysia with the Resorts World Sentosa in Singapore and the Resorts World Manila in the Philippines, all of the Genting Group.

In the Asian market it is especially interesting that a community network constitutes the most important factor in promoting the casino business. Led by Macau, casinos compete with one another by collaborating with a special type of junket operator who provides a specific type of agency service for the casinos by underwriting significant amounts of nonnegotiable chips and directly or indirectly reselling them to patrons for the purpose of earning a commission or sharing a certain percentage of the total wins and losses from the businesses. Under this form of organization, casino firms could secure "basic" turnovers by directing marketing efforts to the junket operators (for more details about this business practice see Siu 2006a and 2007b). Due to the dramatic success of Macau in terms of reported GGR, some Asian jurisdictions have shown interest in embracing this business model. Nevertheless, concerns over low transparency and legitimacy of certain business activities under legal frameworks in different jurisdictions are obstacles to replicating this model. For example, related business is highly controlled in Singapore.

Following the modern development and the dynamics of organizational changes of Asian casino gaming industries, competition through effective amalgamation between gaming and non-gaming facilities would evidently be an indispensable measure for casinos to ensure their competitive edge. Nevertheless, as the regulatory framework and the legal enforcement mechanism in many countries are yet to have sufficient capacity to monitor the related market activities, irregular forms of competitions are likely to exist and persist in some markets, at least for the first half of the twenty-first century.

The Role of Institutions and Their Changes in the Process of Asian Market Progress

In comparison with practices in other markets, the demand for casino gambling is greatly influenced by social and cultural phenomena. Given the particular attributes of the demand and supply of casino gaming, and public concerns over possible adverse effects brought about by related activities, the structure (and organization) of this industry in any one jurisdiction generally are chosen and outlined through the legislative power of the government. In other words, the customs, legal system, and their interactions and changes over time in a society contribute to the development path and performance of a market. To explore the issues associated with the observed performance and progress of the Asian markets, the SSP institutional impact theory by A. Allan Schmid (1987, 39–41) will be followed.

Choice of Industrial Structure and the Impacts on Market Performance

According to the SSP institutional impact theory, while situational variables (i.e., "attributes of individuals, the community, and goods" [39] present in a market are inherent in any given period of time, its rights structure (e.g., entry and exit conditions, allowable actions, distribution of decision power, rules that govern/guide interactions between market participants) "is a matter of human choice" (40). Consequently, market performance (represented by the distribution of income among various competing groups) "is a function of alternative rights given the situation" (ibid).\

In Asia, given the unique situational variables (i.e., the aforementioned demand- and supply-side attributes), the changes in casino gaming is partially the consequence[6] of the changing rights structure as chosen by the related power groups. Besides that, the differences in the industrial structure and organization across various Asian jurisdictions also explain the differences in development and performance. For example, the impressive performance attained by Macau's casino industry since the turn of the twenty-first century was triggered after a decision made by the government to replace the industry monopoly with an oligopoly. In addition, the unexpected remarkable growth in the industry's GGR from the middle of the 2000s was palpable due to a crucial (but controversial) decision by the Macau government (i.e., double the number of legal entitlements shortly after three new licenses were granted in the beginning of 2002). Subsequently, the fixed scale (both gaming and non-gaming hospitality capacities) of the industry expanded by more than double the proposed scale of the three original newly licensed casino firms. Moreover, the hands-off approach adopted by the Macau government to cope with the practices of the casino gaming business has tolerated increasing numbers of new casinos to attain increasingly higher volumes of customer flow and business turnovers. In parallel with the dramatic increase in the industry's GGR, and the surge in government taxation as well as income of the related groups, questionable activities undertaken by casinos to win over their competition (especially in the utilization of a business model that incorporates independent third-party operated gambling rooms) have, however, concealed the real performance of the market (see, e.g., Siu 2006b, 2009b, and 2010). After a decade of progress, even though Macau has been crowned the world's largest casino gaming jurisdiction in terms of reported GGR, its ambition to become the world's casino resort destination may still be hindered by the irregular structure of the industry itself.

In contrast to Macau, the Singapore government perceives casino gaming as a new segment to be integrated into its existing tourism and meetings, incentives, conferences and exhibitions (MICE) industries (tagged "integrated" resort instead of "casino" resort by the government). The primary purpose of legalizing casino gaming is to enhance its competitiveness in the regional tourism market. In debates on granting approval to casino gaming and formulating the structure of the business, the government has clearly delivered a message to the community and potential operators in that casino gaming would be "a very small component of a much bigger whole" (Seneviratne 2005) of the country's economy. Besides that, casino gaming would be subjected to strict regulations and highly controlled by the government. Thus, a transparent and effective corporate governance structure to ensure the integrity of the industry, instead of simply pursuing a high volume of GGR, had been formed at the outset. With proactive plans enforced by the government to interact with the industry, Singapore's casino gaming is highly reputable and efficient in a business sense. Consequently, in the first full year of its operation in 2011, Singapore's casino GGR was larger than that of Las Vegas and was crowned as the second large casino jurisdiction in the world.

Despite the impressive performances reported in Macau and Singapore, it is clear that the underlying forces driving their achievements are quite different. To a large extent, public choices, which fashion the structure and organization of industries, have established the direction of development for casino firms, hence instituting their performances. Indeed, differences in the performances across various Asian jurisdictions (e.g., some related features are presented in figure 4.3 and table 4.1) could also be explained with the same set of logic. For example, the relative outstanding performance of Kangwon Land in South Korea was largely determined by public interests to develop a single large-scale casino resort property as a measure to revitalize the declining economy in the Gangwondo Province in the late 1990s. Consequently, the provincial government took a dominant portion of the property's shareholding, thereby supporting and influencing its business organization. As outlined and supported by the related policies, Kangwon Land outperformed other casinos in the country.

In contrast, most jurisdictions in Central and West Asia are composed of smaller-scale casinos and casino hotels (there are few casino resorts) where traditional and low value-added gaming services are provided. Despite the fact that the market size of casino gaming in Central and West Asia is smaller, the structure and organization of the individual industries as outlined by public interests influence their performances. For example, legal requirements for the structure and organization of casino gaming in India have long been a controversial topic among political and religious parties, which inhibit progress of the industry. In an attempt to enhance tourism development in Goa, a number of offshore and onshore casinos recently were approved. Nevertheless, the Goa Anti-Gambling Act allows protesters to restrict the industry by imposing additional licensing requirements or banning local residents from entering into the casinos, and so on. Furthermore, some licensed casinos may be intermittently forced to stop business, and business turnover therefore becomes very difficult to project for investment purposes (see,e.g., Oberai 2011; Hand 2011; Yogonet Group 2011; Casinocitytimes.com 2008).

In other small West Asian jurisdictions adjacent to Muslim countries, structure and organization as well as performance of the casino gaming markets are also restricted by related social and political considerations. Alternatively speaking, in addition to the attributes of demand and supply, differences in the institutional frameworks as a determinant of the ways that the industries are structured and organized contribute as indispensable factors to explain the differences in market performance across the Asian jurisdictions.

Influence of Institutions and Changes to Their Business Routines

In institutional economics "a routine is regarded as automatic behavior, in contrast to designed and implemented strategic plans," which involves "the notion of procedural memory, or recurrent interaction patterns and involving change driven by individuals" (Lazaric 2011, 147). While it is commonly agreed that "management and authority appear to be good candidates for the formulation of routines" (148), their actions in practice are institutionalized by the legal and cultural environments of society. Owing to this relationship, established routines are altered which would correspond to the changes in the related institutions.

In terms of casino gaming in Asia, as long as a government defines an acceptable middle ground for the interaction of casino firms and patrons, the market evolves from thereon. Subsequently, decisions made by the management of casino firms would then generate business routines. In a dynamic environment where institutions may change due to discrete changes in public interest (hence altering the existing structure of the industry), or changes in the situational variables after their interactions with an approved structure, or both, business routines do evolve.

Indeed, the business model that incorporates independent third-party operated gambling rooms as evolved in Macau from the beginning of the 1980s, and their changes since the alteration of the market structure in 2002, constitute a typical case that reveals a business routine and its evolution (Siu 2006 and 2007b; Chang 2007). Given the unique social and political contexts of Macau in the 1980s and 1990s (Siu 2006a: 971–974; Eadington and Siu 2007, 9–13), various evidence shows that while individual governors were chasing short-term monetary benefits before Macau's handover back to China in 1999, the primary interests of the family-run gaming monopoly was to retain its monopoly position by means of increasing tax payments. Consequently, outsourcing the casino business to certain independent third parties who could ensure contact with big players and hence an ever increasing GGR became the "optimal choice" for this monopoly. Following the initial success of this business arrangement, this evolved into the norm in the Macau casino industry. Under this norm, all of the related parties are bounded by a set of "informally" established rules.

As the ultimate beneficiary, the Macau government has traditionally taken a hands off (or "silent") approach toward the practice of this business model irrespective of the presence of socially and administratively controversial activities in the market. In the casino monopoly, this low amount of transparency was a business routine to attract high GGRs from certain socially and ethically controversial business activities. Indeed, evidence shows that VIP plays (the practice of gambling rooms was the main contributor) had long represented over 60 or 70 percent of the industry's reported GGRs in Macau. To sustain a high turnover in this business, informal contracts between casinos and a particular group of gaming agents had been established as the solution. Since the termination of the gaming monopoly, changes in the business routines of casino firms have evidently taken place. Nevertheless, the hands-off nature of the government continues, as it is bounded by its reliance on high gaming tax revenues, which had hindered a number of its initial attempts to revise the industry's business routines toward those of a modern economic sector. Instead, evidence (see, e.g., Chang 2007) shows that although business routines do evolve, they are adjusted such that the reliance of the industry's GGR on the gambling room business model sustains or is even reinforced.

Since the decision made by the Singapore government to legalize casino gaming, a new model of business organization and routine for the practices in this industry in Asia has evolved. In comparison with Macau, the social structure and legal system (including its enforcement mechanism) in Singapore are much better developed. It is commonly anticipated that the aforementioned business routines employed in Macau will not be replicated in Singapore. Rather Singapore's casinos have formulated a set of their own business practices that accommodate its specific social and political contexts. Indeed, the two contrasting and noticeable cases presented by Macau and Singapore are undoubtedly "valuable reference points for other Asian casino jurisdictions" but "should never be considered directly replicable" (Siu 2008, 21).

Dynamic Relationship between Institutional Changes and Market Progress

Since the beginning of the twenty-first century, changes in social perception in favor of casino gaming as a modern entertainment sector, accompanied by reduction of regulatory restrictions, have generated positive dynamics to further push the progress of this industry. In other words, the recent changes in the formal rules and informal constraints in the related Asian jurisdictions have expanded the boundary of their markets, allowing the industry as a whole to enjoy the benefits of increasing returns to scale. In turn, the direct and indirect economic incomes generated during the process of market growth may provide additional justification for casino gaming to be legalized or expanded further.

In addition to the cases of Macau and Singapore, the dynamic relationship between institutional changes and market progress can be identified also in various major jurisdictions as well as in the emerging markets in Central and West Asia. For example, after its independence from the former Soviet Union bloc in 1991, the loosely formulated and enforced legal system in Kazakhstan might unintentionally foster local enterprises to chase after short-term monetary returns. However, it has been reported that "casinos mushroomed across" the major cities of the country (Yogonet Group 2007), and shady activities, including money laundering, were widely evident. As a measure to counterreact to social pressure, the government instituted new policies to close down most of the country's casinos in 2007 and permitted only a few of them to continue business in "two small resort towns" (ibid). As discipline of the market was restored to a socially acceptable level, and potential economic benefits derived from the casino industry once again caught the attention of the general public and the government, the number of approvals granted to casinos was increased to 28 in 7 cities in 2009 (Angel 2009). During this process, the progress of casino gaming in Kazakhstan from the 1990s was obviously influenced by changes in related regulations and public interests, given the market demand.

In considering the unique attributes of casino gambling and gaming, as well as the interrelated changes between social, political, and economic institutions in the progress of the Asian jurisdictions, their ambitions to develop in a modern sense as high value-added entertainment businesses have evidently accelerated. Other than any significant social and/or political conflicts as discussed, the existing development path will continue, and the efficiency of their business routines will be evaluated.

Social Benefits and Costs Associated with the Growth of the Markets

On account of the rapid growth of the Asian economies, especially the dramatic income and wealth effects that have arisen in China and the anticipated progress of India, the growth in casino gaming would unarguably bring about business opportunities to other related sectors as well as influence the social and economic welfares of local communities. However, owing to the less developed social and political systems in most Asian countries, drawbacks such as social moral values and regulatory systems counteract the monetary benefits derived (and potentially derived) from the growth in casino gaming. By analyzing such from the interest of sustainable growth, a reasonable (but not necessarily optimal) balance between the growth of the industry and its social costs and impacts must be attained.

Direct and Indirect Social Benefits Associated with Growth in Casino Gaming

In line with the experiences of other nations throughout the world (see, e.g., Hall and Hamon 1996; Nichols, Giacopassi, and Stitt 2002; Meister 2005; etc.), the recent development of casino gaming in various Asian jurisdictions re-validates the argument that legalization and growth of this industry provide a considerable amount of economic income, hence resources and opportunities for promoting tourism, and urban development and re-development. In the process, investments made by related firms further stimulate employment and income in the jurisdiction as well as business opportunities for other hospitality-related businesses, such as hotels and retail (i.e., the accelerator and multiplier effects[7]). To sustain the growth of this particular entertainment-based business, the provision of “fresh experiences” to complement routine gambling activities has been a common and necessary tactic undertaken by the industry. Thus periodical renovations and reinvestment in fixtures not only maintain the vigor of the industry; they also contribute to economic growth.

Returning to the example of Kangwon Land in South Korea, it is well documented that “casino development as a strategy for economic rejuvenation of the county is well recognized” (Cho 2002, 185). On top of the casino revenue generated by this property, local employment is safeguarded. In addition, the success of Kangwon Land also contributes a set of fixtures that enhance the development of the county’s tourism and social welfare. As pointed out by the CEO of Kangwon Land, 10 percent of the property’s annual profits since its opening in 2000 “have been spent on various ventures to support local communities” (Lee 2010a).

Similarly, monetary contributions to local communities made by the respective Asian casino industries (e.g., Macau, the Philippines, Malaysia, etc.) in one form or another also are unquestionable. Indeed, contributions are made either directly by the casino firms to support various social and charity programs or indirectly through tax payments, which strengthen the fiscal position of the public sector to sponsor social welfare and investment programs. For example, even though in the case of Malaysia where casino gaming is highly restricted and only approved in a remote area of the country, Genting Highlands organizes and supports various forms of charity concerts and sport events as a measure to win social recognition.

Casino Gaming as a Threat to Social Harmony and Economic Efficiency

Despite the explicit social benefits, the social costs of casino gaming have long been the most crucial aspect arguing against the legitimacy and progress of the industry. Nevertheless, differentiation between social impacts and social costs, and measurability of social costs in gambling studies are always controversial topics in economics (Walker and Barnett [1999] conducted comprehensive review discussions on these topics). It is pointed out that in gambling studies “most (non-economists) authors have adopted an ad hoc approach—asserting that some activities constitute costs to society and then quantifying the impact of those activities” (ibid, 183). Besides, from an economic perspective, not only have researchers inappropriately classified numerous consequences of pathological gambling as “social costs”; they have also omitted several legitimate social costs from their studies. Some of these costs are associated with government restrictions and the legalization process” (204). Nevertheless, the arguments by Walker and Barnett are still subject to continuous debate (see, e.g., Collins and Lapsley 2003).

Consistent with the arguments of Walker and Barnett (1999), and in view of the particular social and political contexts in Asia, it is arguable that to a large extent negative social impacts (e.g., adverse effects to social harmony) rather than social costs may pose major threats to the legalization and growth of casino gaming. Parallel to the recent growth of casino gaming in Asia, and driven by the less developed social structure in many of the jurisdictions, extra (instead of gross) amounts of economic resources consumed by the non-casino sectors (including the public sector) to cope with gambling problems could take place at the cost of other social and economic sectors, in the form of forgoing better alternative uses of the same resources. Inasmuch as most Asian economies have been in stages of development or transition since the last quarter of the twentieth century, economic efficiency from alternative uses of certain public resources to cope with negative spillover effects onto society could be much higher. For example, economic returns derived from funding a job training program (or an infrastructure project, etc.) could be much higher than a responsible gambling program (or a pathological gambling center). Nevertheless, it is probably more equitable to categorize the direct expenditures spent by casino firms on various internal and external programs to grapple with or prevent negative spillover effects as business costs rather than social costs.

Owing to high propensity of Asians to demand casino gambling, and the related drawbacks in the regulatory systems of many jurisdictions, concerns over the negative effects on social harmony are indeed much more profound than those over social costs. For example, even though the Singapore government is widely accredited as being highly transparent as well as efficient and proactive, there have been widespread debates and concerns about casino legalization, in particular, that the business of casino gambling may significantly hinder social harmony in the country. Indeed, protesters of casino gaming, such as Singapore's National Council of Churches, have stressed that a "country which may pride itself on having the best entertainment resort with gambling facilities is unlikely to be a wholesome family-friendly society, which our government seeks to advance" (Seneviratne 2005). In addition, surveys conducted by local organizations have pointed out that "Singaporeans are more concerned about the proposed casino eroding their value system rather than its social cost" (ibid).

In other socially and/or politically less developed Asian countries, widely reported social problems linked to the practice of casino gaming have posed significant threats to social harmony. For example, crime, commercial sex, drugs, money laundering, and so on have triggered many debates over the social impacts of casino gaming. It is however necessary to carefully examine and gauge whether or not the gross amount of expenditures spent to deal with related problems presumably brought about by legalization and practice of casino gaming are straightforward social costs, given that a large proportion of these monetary payments do not necessarily lower the net economic efficiency.

Balance between Scale of Business and Social Costs/Impacts

Given the evidence illustrated so far in relation to the issues of Asian casino gambling and gaming, it is clear that the development and growth of casino gaming in Asia are by no means cost-free. As a socially sensitive economic sector where such non-market factors as culture and religion play essential roles in evaluating the overall social value, negative social impacts may drive the focus of the community toward only the social costs associated with casino gaming. Given that social acceptance, hence continuous public approval of casino gaming, are preconditions for the industry to remain in business, it becomes evident that a reasonable balance between the scale of business in this industry and its social costs and impacts is a key challenge deemed to be addressed by all related parties.

As summarized in figure 4.6, to sustain the growth of Asian casino gaming, each jurisdiction should aspire to attaining growth at a reasonable scale by taking into consideration the social benefits and costs. In this process, the feasibility may be adjusted in accordance with the social and political perceptions on the social impacts of this industry. Owing to the nature of such normative variables as culture, religion, political interests, and the like, related institutions should recognize that a balance between the scale of casino gaming and its social impacts could hardly be optimal, as are the eventual social benefits and costs observed by the general public.

In considering the variations in social values, political systems, economic structures, and stages of economic development, the necessary course of actions to realize the principles as outlined in figure 4.6 to balance the scale of casino gaming and its social costs and social impacts would never be identical across various Asian countries/regions. Indeed, it is a matter of the degree of public interest in a jurisdiction. While culture and religion in some jurisdictions may play a dominant role, political parties in other jurisdictions may be able to exert more influence in determining the scale of this industry and the changes involved. Furthermore, in some other jurisdictions, economic interests may take the lead in driving related decisions of the community and government.

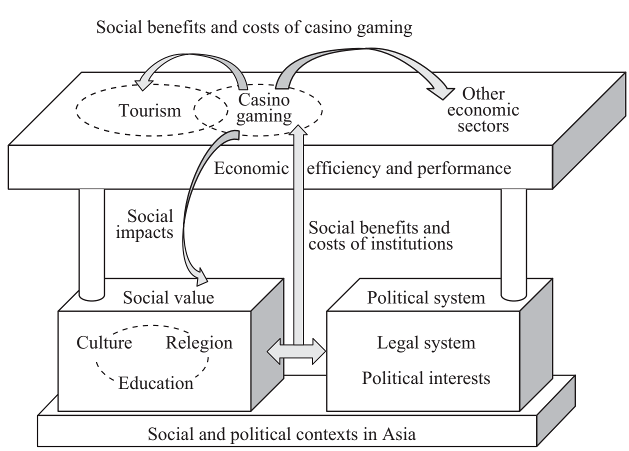


FIGURE 4.6 Social benefits, costs and impacts of casino gaming in the Asian jurisdictions

Conclusions and Remarks

In a nutshell, the practices of casino gaming and gambling in Asia share many common features with those in world markets. Nevertheless, the unique attributes of the demand in Asia for casino gambling, and the structure/organization of the individual industries as outlined by public interests at the outset, differentiates its performance from other jurisdictions in the world. Besides that, as excess demand is a common phenomenon associated with the world casino gaming markets, and the scale of the casino gaming industry is highly determined by public choice, they differ across various Asian jurisdictions. These differences are largely explained by dissimilarities in (1) culture, religion, level of education, stage of economic development, and hence the income and wealth effects on the demand side; (2) industrial structure (including the forms of competition), business organization and routines on the supply side; and (3) legislative and regulatory systems, enforcement mechanisms, and dynamic interactions between the power groups and the community on the institutional side. Despite dissimilarities in scale and performance across various Asian casino markets, the gaming industries share a common progression toward becoming a more modern and high value-added entertainment economic sector instead of retaining a traditional business form of gambling.

In considering the potential growth of casino gaming in Asia, which is supported by the increasing income and wealth of the Asian economies and acceleration of interregional and international flows of population that have occurred since the end of the twentieth century, proactive public policy measures undertaken by the related governments are necessary to sustain the growth of this industry and its real contributions to the host jurisdictions. Indeed, evidence shows that for the sake of formulating an effective institutional framework to balance the social, political, and economic interests of the related parties, a free market approach or merely the replication of success stories from one or a few jurisdictions may not serve the purpose. Instead, each jurisdiction should cautiously examine both the positive and negative experiences of other jurisdictions and refer to its own social, political, and economic contexts when approving or altering the scale of the casino gaming industry and the related rules. In contrast to the traditional economic sectors, other than economic interests, social and political perceptions on the role of casino gaming exert significant influence on the practice of this industry.

Notes

1. According to the United Nations, Taiwan is classified as a de facto country instead of a legally recognized independent country. Furthermore, in consideration of the particular historical and political backgrounds of Hong Kong and Macau as the two SARs of China, and for the purposes of this study, the term *China* in this study refers to Mainland China,

where casino gaming is outlawed, but it has been legalized in both the Macau SAR and Taiwan.
2. With reference to the living standards in China, this group is regarded as its middle income class.
3. Given that the transparency of many casino jurisdictions in Asia is quite low, official data which indicate the number of licensed casinos in related industries and their respective GGRs are unavailable. Also, statistics displayed in various sources are occasionally found to be inconsistent. Accordingly, only an approximation of the related scale instead of exact numbers of the markets are indicated in the scatter chart (Figure 4.3). Nevertheless, the data consolidated and presented in this chart fairly depict the existing scale of the related industries.
4. Parallel to the number of licenses as predetermined by a jurisdiction's government, the physical configuration of a casino industry is neither solely nor explicitly determined by the markets. Instead, it is a matter of public choice largely influenced by the institutions in a society and their dynamic changes over time.
5. Statistical verification is constrained by the lack of data with regard to the many Asian casino jurisdictions.
6. Emphasized in the role of institutions in this section, influences from both demand- and supply-side factors and their attributes on the performance of casino gaming markets have the same importance.
7. Due to the lack of related data, empirical testing on these two effects has never been conducted. Nevertheless, the positive spillover effects from the growth of casino gaming are widely acknowledged by related business sectors and the general public.

References

Angel. 2009. Casinos in Kazakhstan. Palacelive.com, Dec. 2; http://palacelive.com/2009/12/02/casinos-in-kazakhstan.

BBC News. April 17, 2007. China border casino crackdown; http://news.bbc.co.uk/2/hi/asia-pacific/6562401.stm.

Becker, Gary S., and Kevin M. Murphy. 2000. *Social economics: Market behavior in a social environment*. Cambridge, Mass.: Harvard University Press.

Binde, Per. 2005. Gambling across cultures: Mapping worldwide occurrence and learning from ethnographic comparison. *International Gambling Studies* 5(1):1–27.

——. 2009. *Gambling motivation and involvement: A review of social science research*. Östersund: Swedish National Institute of Public Health; http://www.fhi.se/PageFiles/9161/R2009-20-gambling-motivation-a-review.pdf.

Blanchard, Maurice. 2002. Chinese game: gambling history and tradition in China (Part 1). *Gambler's Tribune* , June 12; http://www.gambletribune.org/article148.html (no longer on website).

Casinocitytimes.com. May 20, 2008. Goa allows four more casinos; http://www.casinocitytimes.com/news/article/goa-allows-four-more-casinos-173174.

Chang, Octo. 2007. Revenue sharing revealed. *Inside Asian Gaming*, March 1; http://www.asgam.com/article.php?id_article=1211&page=1.

Chiang, Benjamin, and Sherry Lee. 2010. China's middle-class catches spending fever. *CommonWealth Magazine*, May 27; http://english.cw.com.tw/article.do?action= show&id= 11985&offset= 0.

Chien, Grace C. L., and Cathy H. C. Hsu. 2006. Gambling and Chinese Culture. In *Casino industry in Asia Pacific: Development, operation, and impact*, edited by Cathy H. C. Hsu. Binghamton, N.Y.: Haworth Hospitality Press, 201–224.

Cho, M. H. 2002. Tourism redevelopment strategy: The case of the Kangwon Land Resort Casino. *Anatolia: An International Journal of Tourism & Hospitality Research* 13(2): 185–197.

Collins, David, and Helen Lapsley. 2003. The social costs and benefits of gambling: An introduction to economic issues. *Journal of Gambling Studies* 19(2):123–148.

Curtis, Harry C., Louise Cheung, and Brian H. Dobson. 2011. Macau and Singapore are undersupplied. Nomura Equity Research Americas, May 10

Eadington, William R. 1984. The casino gaming industry: A study of political economy. *Annals of the American Academy of Political and Social Science* 474(1):23–35.

——. 1999. The economics of casino gambling. *Journal of Economic Perspective* 13(3):173–192.

Eadington, William R., and Ricardo C. S. Siu. 2007. Between law and custom—Examining the interaction between legislative change and the evolution of Macao's casino industry." *International Gambling Studies* 7(1):1–28.

Garrett, Thomas A., and Mark W. Nichols 2005. Do casinos export bankruptcy? Working Paper No. 2005-019A. St. Louis, Mo.: Federal Reserve Bank of St. Louis..

GBGC (Global Betting and Gaming Consultants). 2011. Gambling on Asia's middle classes. GBGC.com, April 22; http://www.gbgc.com/2011/04/gambling-on-asia's-middle-classes.

Hall, C. Michael, and Christopher Hamon. 1996. Casinos and urban redevelopment in Australia. *Journal of Travel Research* 34(3):30–36.

Hamilton, David. [1970] 2004. *Evolutionary economics: A study of change in economic thought.* New Brunswick, N.J.: Transaction.

Hand, S. 2011. Casino may soon be out of reach for Goans. CasinoExecutiveNetwork.com, March 28; http://www.casinoexecutivenetwork.com/news/casinos-may-soon-be-out-of-reach-for-goans.

Hodgson, Geoffrey M. (1998). The approach of institutional economics. *Journal of Economic Literature* 36(1):166–192.

Hunkar, David. 2009. Household savings in China, India, and South Korea. SeekingAlpha.com, Dec. 13; http://seekingalpha.com/article/177932-household-savings-in-china-india-and-south-korea.

International Land Casinos Directory. n.d.; http://www.ildado.com/land_casinos_directory.html.

Kim, I., and Matthew O. 2011. Korean tourism. CLSA Asia-Pacific Markets, Feb. 21.

Lai, Daniel. 2011. Gambling and experience of ethno-cultural minorities Faculty of Social Work, University of Calgary; http://dspace.ucalgary.ca/bitstream/1880/48427/1/Lai_Gambling_Ethnocultural_Minorities_Feb_2011.pdf.

Lazaric, Nathalie. 2011. "Organization routines and cognition: An introduction to empirical and analytical contributions. *Journal of Institutional Economics* 7(2):147–156.

Lee, Yunekyong. 2010a. Gangwon's biggest booster. *Korea Joongang Daily*, Oct. 6; http://joongangdaily.joins.com/article/view.asp?aid=2926785.

——. 2010b. Estimating the future growth potential of Kangwon Land casino. UNLV Theses/Dissertations/Professional Papers/Capstones Paper 764; http://digitalcommons.library.unlv.edu/cgi/viewcontent.cgi?article=1764&context=thesesdissertations&sei-redir=1#sear ch=%22Kangwon%20Land%202003%22.

Li, Guoqiang, Xinhua Gu, and Ricardo Chi Sen Siu 2010. The impacts of gaming expansion on economic growth: A theoretical reconsideration. *Journal of Gambling Studies* 26(2): 269–285.

Marshall, Alfred. [1890] 1997. *Principles of economics.* Reprint, Amherst, N.Y.: Prometheus.

Meister, Alan. 2005. *A survey of urban gaming in America.* Los Angeles: Analysis Group, Inc.; http://www.gabrielinotribe.org/EmpiricalStudies/Urban_Gaming_Study_full_final.pdf.

Nepstad, Peter. 2000. The gods of gambling. *The Illuminated Lantern*, April–May; http://www.illuminatedlantern.com/cinema/archives/the_gods_of_gambling.php.

Nichols, Mark, and Mehmet S. Tosun. 2008. The income elasticity of gross casino revenues: Short-run and long-run estimates. *National Tax Journal* 61(4):635–652.

Nichols, Mark, David Giacopassi, and B. Grant Stitt. 2002. Casino gambling as a catalyst of economic development: perceptions of residents in new casino jurisdictions. *Tourism Economics* 8(1):59–75.

North, Douglass C. 1997. The contribution of the new institutional economics to an understanding of the transition problem. WIDER Annual Lecture, vol. 1. Helsinki: United Nations University–World Institute for Development Economics Research, March; http://www.wider.unu.edu/publications/publications.htm.

Oberai, Sanjit. 2011. Forget the beaches, Goa is about gambling now. Firstpost.com, May 31; http://www.firstpost.com/business/legal-gambling-in-goa-to-get-bigger-by-2013-17765.html.

Prasch, Robert E. 2008. *How markets work: Supply, demand and the "real world."* Northampton, Mass.: Edward Elgar.

Progress Publishing. 2006. A history of gambling;. http://www.crapsdicecontrol.com/gambling_history.htm.

Rank, Michael. 2005. China cashes in its N Korean casino chips. *Asia Times Online*, Feb.24; http://www.atimes.com/atimes/China/GB24Ad06.html.

Samuelson, Paul A., and William B. Nordhaus. 1995. *Economics.* 15th ed. New York: McGraw-Hill/Irwin.

Schmid, A. Allan. (1987). *Property, power, and public choice: An inquiry into law and economics.* 2nd ed. New York: Praeger.

Schwartz, David G. (2003). *Suburban Xanadu: The casino resort on the Las Vegas strip and beyond.* New York: Routledge.

Seneviratne, Kalinga. 2005. Not ready for bet, just yet. *Asia Times Online*, Feb. 1; http://www.atimes.com/atimes/Southeast_Asia/GB01Ae02.html.

Shinogle, Judith, Donald F. Norris, DoHwan Park, Rachel Volberg, Donald Haynes, and Eric Stokan. 2011. Gambling prevalence in Maryland: A baseline analysis. Maryland Institute for Policy Analysis and Research; http://dhmh.maryland.gov/mha/Documents/GamblingPrevalenceinMD.pdf.

Siu, Ricardo C. S. 2006a. Evolution of Macao's casino industry from monopoly to oligopoly—Social and economic reconsideration. *Journal of Economic Issues* 40(4):967–990.

———. 2006b. The puzzles behind the mania: Macao's unfinished projects. *Casino & Gaming International* 2006(4):59–63.

———. 2007a. Is casino gaming a productivity sector? A conceptual and cross-jurisdiction analysis. *Journal of Gambling Business and Economics* 1(2):129–146.

———. 2007b. Formal rules, informal constraints, and industrial evolution—The case of the Junket Operator Regulation and the transition of Macao's casino business. *UNLV Gaming Research & Review Journal* 11(2):49–62.

——. 2008. After the games in Macau and Singapore, what next? *Casino & Gaming International* 2008(4):17–21.

——. 2009a. "A financial analysis of the integration of the Vegas casino resort model and the Macao gambling room model." Paper presented in the 14th International Conference on Gambling and Risk-Taking, Lake Tahoe, Nevada, May 25–29.

——. 2009b. 2008: Macau's mood; Facing challenges and embracing opportunities. *Casino & Gaming International* 2009(2):47–51.

——. 2010. The marvel and dilemma in the progress of Macau's casino industry. *Macau Business*, December, 37.

——. 2011. A conceptual reconsideration of price issues with casino gaming. *Gaming Law Review and Economics* 15(5):267–277.

Siu, Ricardo C. S., and William R. Eadington. 2009. Table games or slots? Competition, evolution, and game preference in Macao's casino industry. *Journal of Gambling Business and Economics* 3(1):41–63.

State Government of Queensland, Department of Corrective Services. 2005. Games people play: Problem gambling among offenders in Queensland community corrections. November; http://www.olgr.qld.gov.au/resources/responsibleGamblingDocuments/gamesPeoplePlay.pdf.

Veblen, Thorstein. [1899] 1959. *The Theory of the Leisure Class: An Economic Study of Institutions.* Reprint, New York: Mentor Books.

Walker, Douglas M. 2007. *The Economics of Casino Gambling.* New York: Springer.

Walker, Dougas M., and A. H. Barnett. 1999. The social costs of gambling: An economic perspective. *Journal of Gambling Studies* 15(3):181–212.

Wilkinson, Philip. 2008. *Religions.* New York: Dorling Kindersley Ltd.

Yogonet Group. 2007. Kazakhstan closes down casino in big cities. Yogonet.com, April 4; http://www.yogonet.com/english/history/7681.

——. 2011. Goa government orders closure of two off shore casinos. Yogonet.com, Feb. 1; http://www.yogonet.com/english/2011/02/ 01/goa-government-orders-closure-of-two-off-shore-casinos.

CHAPTER 5

HOW DOES IMPLEMENTATION OF A SMOKING BAN AFFECT GAMING?

JOHN C. NAVIN, TIMOTHY S. SULLIVAN, AND WARREN D. RICHARDS

1 Introduction

Basic microeconomic theory tells us that when consumer choice is restricted consumers are made worse off. When faced with two alternatives, one that restricts choice and one that does not, consumers prefer the unrestricted choice. Many state and local governments have found that restricting consumer choice, as it relates to gaming, has had such an impact. These restrictions have taken the form of restrictions on hours, location, loss limits, and cruising requirements, for example. In the early days of riverboat gaming, a number of restrictions were placed on access to the riverboats. When Illinois opened its first riverboat casino in 1991, a prohibition against any land-based gaming was instituted. Boats would literally cruise the river while gaming took place and then return to the dock. While Illinois was prohibiting land-based gaming, Missouri found that it could set up “boats in moats.” These were small pools, or moats, dug out next to the river and filled with river water. Barges were then floated in these moats and casinos were built on top of them. These new casinos no longer made patrons subject to the boarding, disembarking, and cruising requirements. According to Bernard Goldstein, chairman and CEO of Isle of Capri Casinos, “bettors didn’t like cruising. When they wanted to get on, they wanted to get on, and when they wanted to get off, they wanted to get off.[1]” Similarly, both Missouri and Iowa placed loss limits on their gamblers. The casinos and the state gaming commissions quickly realized that restrictions on the amount people could wager and lose were costing them money. A study by John McGowan and Muhammad Islam (2003) estimated that removal of the loss limits would increase state and local tax revenue in Missouri by $50 million. In addition they estimated that $27 million in state and local tax revenue was being lost to Illinois by patrons being able to simply drive into Illinois and not face the constraint of a loss limit. Over time, restrictions on entry and spending have been removed, and the casinos have seen an increase in both admissions and total handle.

As those restrictions on gaming have disappeared, new restrictions have appeared—not limited to just riverboats, but to all establishments. These restrictions are manifested in smoking restrictions, including total smoking bans.

Smoking bans have proliferated across the United States and the globe in recent years. As shown in figure 5.1, as of 2011 roughly half of all U.S. states had banned smoking in all public places. A handful of other states have banned smoking in most workplaces and some hotels and restaurants. A number of cities and counties have implemented local smoking bans where no statewide ban exists.

More recent bans have been implemented in response to growing health concerns regarding second-hand smoke. Earlier bans focused mostly on the discomfort suffered by nonsmokers and those (such as asthma patients) who are especially sensitive to cigarette smoke. As such, most early bans required businesses to set aside smoking areas in government buildings or required restaurants to designate a portion of their seating as a nonsmoking area. However, as concerns have increased regarding the possible health effects of second-hand smoke, state legislatures have focused on smoking as a potential workplace safety issue. The policy arguments that call for safety equipment and ergonomic workspaces have been extended to the protection of workers from cigarette smoke. Generally there are few exemptions to the smoking bans. In many areas the only exemptions are hotels, long-term care facilities (which may, under certain restrictions, designate smoking rooms), private residences, certain private clubs, and tobacco shops.

Smoking restrictions, such as these, raise objections from smokers' groups and the business community. Smokers, naturally, resent the inconvenience of being unable to smoke or of being forced outside to smoke. Businesses, especially bars and restaurants that rely on discretionary dollars, fear they will lose revenue if patrons are unable to smoke (or are heavily inconvenienced). Patrons suffering such inconvenience may not stay as long or may simply stay at home, where they can smoke more conveniently.

The gaming industry, and casinos especially, are quite unhappy with the smoking restrictions. Anecdotal evidence suggests that a disproportionate number of casino patrons smoke. Patrons forced to leave the casino to smoke may be less likely to return to the table or may forgo visiting the casino altogether. This is a major concern for establishments within a short distance of gaming establishments not subject to such restrictions. For example, the Illinois riverboat casinos in Alton and East St. Louis, as shown in figure 5.2, are within easy driving distance of four casinos in Missouri. (All four are less than 30 miles away.) The Casino Queen in East St. Louis is directly across the Mississippi River (approximately two miles) from both Lumiere Place and the newly opened River City casino in south St. Louis and is a short (roughly 25-mile) drive from the two casinos west of St. Louis. While the Illinois casinos are subject to the smoking ban, gamblers wishing to smoke and gamble need only drive across the river to Missouri, where no such restrictions exist.

This chapter reviews the current literature on the impact of restrictions on patron smoking on gaming behavior. It also provides a brief case study of the riverboat gaming market in Illinois and Missouri to highlight the issues associated with the imposition of localized smoking restrictions. The case study uses thirteen years' worth of data—two and a half of which follow the smoking ban's implementation—to evaluate whether the smoking ban appears to have reduced the revenues of the casinos located in the Illinois portion of the St. Louis Metropolitan area.[2] We use monthly data on the amount of money wagered at table games and slot machines at the casinos in the St. Louis Metropolitan Area. These data are reported by the casinos to the respective state gaming commissions and are published each month. Controlling for other factors, we will attempt to isolate the impact of the smoking ban on the Illinois casinos.

The remainder of this chapter is organized as follows. Section 2 summarizes previous studies regarding the impact of smoking bans. Section 3 summarizes the data used in the study. Section 4 uses change-point analysis to examine the effect of the Illinois smoking ban. Section 5 uses the Missouri casinos as a control group to examine the effect of the ban on Illinois casinos. Section 6 uses regression methods to isolate the effect of the ban. Section 7 provides a summary and discusses implications of the findings.

2 Research on Smoking and Gaming

While the press is filled with speculation regarding the impact of smoking bans on gaming, relatively few careful empirical studies have been conducted. Most press pieces report year-to-year changes following a ban, with no attempt to control for other factors that may be at work. Over the last few years, however, there has been an increase in interest as to the impact of smoking bans—in part because they have a large impact on tax revenues. This section examines the impact of smoking bans on various types of gaming establishments.

There is a growing literature that examines the relationship between smoking and gaming. Todd Harper (2003) found a number of interesting facts as they relate to gambling and smoking. Based on data from Victoria, Australia, he found that smokers spend, on average, more than twice as much as nonsmokers at electronic gaming devices. D. S. McGrath and S. P Barrett (2009) also found that smoking was correlated with a higher rate of gaming. N. M. Petry and C. Oncken (2002) demonstrated that problem gamblers who smoke both gamble more often and spend more money than nonsmokers. Smoking appears to be correlated also with problem gaming (see Grant et al. 2008 and Rodda, Brown, and Phillips 2004 for example). This correlation has only worked to reinforce the will of those looking to ban smoking in the presence of any type of gaming.

Delaware was the first state to implement a statewide smoking ban that impacted gaming establishments. Delaware legalized racinos in 1994. (A racino is a horseracing track that also provides electronic gaming.) By the end of 1996, all three of Delaware's racetracks permitted electronic gaming. Three studies have focused on the effect of Delaware's statewide ban. L. L. Mandel, B. C. Alamar and S. A. Glantz (2005) used data from three racinos (Delaware Park, Dover Downs, and Harrington) to examine the impact of the November 2002 statewide smoking ban on total gaming revenue at each location as well as average revenue per machine at each location. Using ordinary least squares and controlling for time, income, and seasonal effects, the authors found no impact on either total revenue or average revenue per machine. Michael Pakko (2006a and 2006b) also examined Delaware's data. Using the same data as that used by Mandel, Alamar, and Glantz and incorporating a slightly different estimation methodology, Pakko found a clear adverse effect on revenue from gaming as a result of the smoking ban. Richard Thalheimer and Mukhtar Ali (2008) used a different approach, estimating separate demand functions for each of the three racinos. Using seemingly unrelated regressions, and total handle, as opposed to total revenue, they found that the impact of the smoking ban was significant—accounting for almost a 16 percent decrease in gaming demand. In later article Pakko (2008) found that those racinos which exhibited the greatest loss in revenue were those that faced direct competition from gaming establishments that did not restrict the ability of patrons to smoke.

As the proliferation of smoking bans, both local and national, has grown, the impact on alternative forms of gaming has been investigated. M. K. Pyles and E. J. Hahn (2009) examined the impact of local smoking bans on charitable gaming revenues. Using a fixed-effects model at the county level they found no measurable impact of the smoking ban on charitable gaming. A recent report by the Minnesota Gambling Control Board (2008) found a significant negative impact on charitable gaming receipts after the statewide smoking restriction was put in place. They found a reduction of between 7.5 and 8 percent of statewide gross receipts. Further they found that sites within 10 miles of the state border reported a loss in gross receipts of 17.7 percent post smoking ban. In addition to simply being located near a border, the presence of tribal gaming establishments (bingo halls, casinos, etc) also appears to impact venues that become subject to smoking bans. The state of Washington implemented its smoking ban in 2005, and while it covered the state's non-tribal casinos, Native-American establishments were exempted. John Douglas (2008) reported a reduction in revenue of 13 percent at Washington's non-tribal establishments. S. A. Glantz and R. Wilson-Loots (2003) examined the effect of smoking restrictions on profits from bingo and charitable gaming in Massachusetts. Using a liner regression model and establishing dummy variables to indicate the presence of a complete ban on smoking in public areas, they found that none of the smoking ban dummy variables were significant. This study did find a year-to-year decline in those profits over time but was unable to decipher the cause of the year-to-year decline in profits

A. Lal and M. Siahpush (2008) examined the effect of smoking bans on electronic gaming expenditures in Victoria, Australia, as that effect compared to the rest of Australia. Using data on monthly gaming expenditures from 1998 through 2005, they found that the smoking ban resulted in a 14 percent reduction in expenditures on electronic gaming devices.

Jonathan Macy and Erika Hernandez (2011) examined the impact of a local smoking ban on an off-track betting (OTB) facility in Indiana. Using regression analysis to compare per capita wagering at the OTB facility that banned smoking against two facilities that did not, the authors found no difference in the longitudinal trend in per capita wagering.

An additional area of focus has been the impact of a statewide smoking ban on riverboat casinos. Riverboat casinos operate in a number of states (Indiana, Iowa, Missouri, Illinois, Mississippi, and Louisiana), but only Illinois has banned smoking on riverboat casinos. The focus of the remainder of this chapter is an examination of the impact of the Illinois smoking ban. To date, three previous studies have examined this ban. John Navin, Timothy Sullivan, and Warren Richards (2009) examined the impact of the smoking ban on the two Illinois casinos (Casino Queen in East St. Louis and the Argosy Casino in Alton) located in the St. Louis metropolitan area. Using monthly data from 2000 through 2009, they found that Illinois casino revenue dropped between 8 and 24 percent depending on the type of gamine device (tables or electronic) and the method of estimation used. More recently Thomas Garrett and Michael Pakko (2010) also examined the impact of the Illinois ban on both admissions as well as revenue. Garrett and Pakko divided the Illinois gaming market into four areas: Chicago, Southern (Metropolis, Ill.), Quad Cities, and St. Louis. Using a log-linear model and data from 1997 through 2008, they found a decline in both admissions and adjusted gross revenue (total handle less payouts). Their revenue estimates ranged from a 10 percent to 30 percent decline for the Illinois casinos. Finally, Jenine Harris et al. (2011) examined monthly casino admissions in Illinois and the surrounding states that have non-tribal casino gaming (Missouri, Indiana, and Iowa) to test for the impact of the smoking ban. Using a model that incorporates, time, location, and various measures of economic activity, Harris et al. found that the decline in Illinois' casino admissions was not significantly different from those of surrounding states. They further concluded that the decline in Illinois' casino revenue is not therefore due to the smoking ban but rather is a result of slower economic activity.

As with much research surrounding smoking and gaming, both sides of the debate are quick to dismiss studies by the proponents of the other side. The studies described above that found no impact due to a smoking ban were all published in *Tobacco Control*, a journal that is seen as a critic of the tobacco industry. Such studies are roundly criticized by smokers' rights groups, which offer up opposing studies arguing that smoking bans hurt business.

While no single study can hope to end the debate, the data set we use for this chapter's study is somewhat unique, and we hope it will work to highlight some of the issues discussed above as they relate to smoking bans. While this study looks specifically at the riverboat gaming issue, the results obtained here are generalizable to other forms of gaming where local substitutes not subject to the same form of consumer restriction exist. We used data for a single metropolitan area that (at the time) hosted five casinos within a 30-mile radius. We found that two of the casinos were affected by a smoking ban implemented on one date; three were not. This data set thus provides an excellent opportunity to isolate the impact of the smoking ban from that of other factors. In addition we have expanded on our earlier work by adding an additional 18 months of data following the implementation of the Illinois ban.

3 Data and Methodology

The Illinois smoking ban went into effect on January 1, 2008, and, in theory, examining the impact of the smoking ban should be as simple as examining the Illinois casinos for signs of a structural break in activity. If the smoking ban had an important negative effect on the Illinois casinos near the Missouri border, we should expect to see a sizable drop in two data series collected by the Illinois gaming commission—table drop (the amount collected at table games, such as blackjack) and slot handle (the amount deposited into slot machines).

As shown in figures 5.3 and 5.4, there was a substantial drop in activity at the Alton (Ill.) casino in January 2008. Table drop fell by nearly a quarter—from just over \$3 million per month throughout most of 2006–2007 to levels near \$2.5 million each month during 2008. Slot handle meanwhile fell by about a third—from levels near \$150 million to levels close to \$100 million.

Figures 5.5 and 5.6 show a similar drop in activity at the East St. Louis (Ill.) casino. Compared with the corresponding month of 2007, table drop at the East St. Louis casino plunged by about one-fourth in the second half of 2008. Similarly monthly slot handle totals fell by about one fourth in the second half of 2008.

In isolation, while interesting, this evidence of drops can't necessarily be attributed to the smoking ban. The drops could reflect the economic downturn, new entertainment alternatives in the St. Louis area, or changes in gamblers' tastes.

A further important complication in the analysis is the opening of Lumiere Place, a major St. Louis (Mo.) casino, on December 19, 2007. Lumiere Place immediately took a position near the top of the St. Louis area casinos in terms of revenue. Furthermore, as might be expected, its opening corresponded with an aggressive marketing campaign. Because Lumiere Place opened less than two full weeks before the Illinois smoking ban went into effect, it is extremely difficult to disentangle the two effects on Illinois casinos. In fact, because it opened in mid-December, we do not have even one complete month's worth of data with Lumiere Place in the region prior to the smoking ban's implementation.

We pursued three strategies to disentangle the results. First we used a change-point test to see if a change point was detected and, if so, whether a December or January change point dominated. Second, we treated the three Missouri casinos (other than Lumiere Place), which were impacted by Lumiere Place but not the smoking ban, as a control group for isolating the effect of Lumiere Place. Finally, we used a time series regression analysis.

4 Change-Point Analysis

Change-point analysis is a collection of methods used to look for structural change in a time series. The simplest case considers whether the population mean differs before and after some point in time (the change point). Other analyses may be concerned with changes in other parameters, such as the median or variance. Non-parametric tests consider whether the underlying distribution has changed (without specifying a parameter).

Not surprisingly, given the patterns found in figures 5.3 through 5.6, a simple test of difference in means indicates that both the Alton and the East St. Louis casinos have statistically different means for the periods before and after the smoking ban. As shown in table 5.1, the Alton casino's monthly average table drop fell from over $4 million to just over $2 million. Its slot handle fell from a monthly average of over $140 million to just under $100 million. Similarly, as shown in table 5.2, the East St. Louis casino's table drop fell from a monthly mean of about $12 million to an average of about $10 million. Its monthly average slot handle fell from about $212 million to just over $190 million. All of these decreases are statistically significant, using a standard difference in means test.

An important shortcoming of these difference-in-means tests is that they cannot identify a unique change point. As previously discussed, the lower average during this 30-month period might reflect the opening of Lumiere Place in December 2007. As shown in tables 5.1 and 5.2, similar drops are found when testing for a difference in means beginning in December 2007. The standard difference-in-means test has no method for determining which bifurcation dominates.

The tests described by A. N. Pettit (1979), however, can test the relative importance of the breakpoints. Pettit provides a framework for using the Mann-Whitney-Wilcox *U* test for a difference in distributions to test for a change point. In addition, Pettit suggests a method for using the magnitude of the *U* statistic to select from different potential change points.

Pettit's method provides mixed results. *U* test statistics for the Alton casino indicate that the opening of Lumiere was slightly more important than the smoking ban for table drop, while for slot handle the two events are equal in terms of providing a break point. For the East St. Louis casino, the implementation of the smoking ban appears to be a stronger break point than the opening of Lumiere. However, as a reminder, Lumiere Place was open for less than half of December 2007, meaning that the full impact of the opening may not be reflected in the December 2007 data. This would have the effect of biasing the results *against* a December 2007 change point.

While the change point results indicate that the January 2008 change point (implementation of the smoking ban) is important, it leaves open the possibility that two separate change points occurred. It is entirely possible that the Illinois casinos were hit with a double punch—first the opening of Lumiere Place and then the implementation of the smoking ban. The struggling economy during the past year is likely a third factor. The next section attempts to control for these other effects by using the Missouri casinos as a control group.

5 Comparison with the Missouri Casinos

While the evidence presented in previous sections of this chapter indicates that the Illinois casinos in the St. Louis Metropolitan Area saw sizable drops in gaming revenue following the implementation of the smoking ban, they are far from definitive. They leave open the possibility that the entrance of a new competitor drove the results (and in fact one piece of evidence in section 4 indicates that the new competitor was a more important factor). In addition the general downturn in the U.S. economy coincided (roughly) with the implementation of the smoking ban.

Fortunately, a ready control group is available for the analysis—the three casinos in the Missouri portion of the St. Louis Metropolitan Area that existed prior to December 2007. As shown in figure 5.2, the President Casino in St. Louis is less than two miles away from the new Lumiere Place (much like the Casino Queen, located in East St. Louis). Similarly, the two casinos west of St. Louis—Harrah's in Maryland Heights and Ameristar in St. Charles—are both about 20 miles from Lumiere Place (much like the Argosy in Alton). The effect of the new casino on these three casinos should be similar to the effects on the Illinois casinos. Furthermore, as all of the casinos are located in the same metropolitan area (within a 30-mile diameter) prevailing economic conditions should be very similar.

As shown in table 5.3, from 2007 to 2008 the two Illinois casinos saw a drop in their combined slot handle that was higher in both absolute and percentage terms than that of the three preexisting casinos in Missouri. The two Illinois casinos saw a combined decrease of about 24 percent ($1.2 billion), compared with a decrease of about 11 percent ($0.8 billion) at the three Missouri casinos. Lumiere Place's slot handle during 2008 was about $1.6 billion, indicating that it could not have single-handedly accounted for the combined decrease at the five older casinos.

As shown in table 5.4, from 2007 to 2008 the two Illinois casinos saw a decrease in table drop that was lower in dollar terms but higher in percentage terms than that of the three preexisting casinos in Missouri. The two Illinois casinos saw a combined decrease of about 16 percent ($31 million), compared with a decrease of about 13 percent ($40 million) at the three Missouri casinos. Lumiere Place's table drop during 2008 was about $108 million, indicating that it could have accounted for the combined decrease at the five older casinos.

These results suggest that the smoking ban may have had an important effect on the casinos in the Illinois portion of the St. Louis Metropolitan Area. If we are willing to accept that the year-to-year decreases at the three preexisting Missouri casinos represent the combined effects of Lumiere Place and the poor economy, the excess decrease at the Illinois casinos would represent the effect of the smoking ban.

We formalized this analysis by examining the percent of revenue at the five preexisting casinos that can be accounted for by the two Illinois casinos. The logic of this analysis is straightforward: changes in the economy and the opening of Lumiere Place would affect the size of the St. Louis market to be distributed across these five casinos but (to the extent that the five casinos operate within the same market) would not affect the distribution between Illinois and Missouri.

As seen in table 5.5, among the five preexisting casinos, there was a shift in gaming revenue from the Illinois side to the Missouri side following the smoking ban. Between July 1997 and the implementation of the smoking ban, the two Illinois casinos accounted for (on average) 40 percent of the table drop and 41 percent of the slot handle. Following the smoking ban, these percentages dropped to 34 percent and 36 percent, respectively. Not only did the gaming pie shrink (due to the economy and the opening of Lumiere Place), but the percent of the pie going to Illinois got smaller as well.

6 Regression Analysis

Finally we estimated two regression models. The first regression examines the Illinois casinos individually. The second regression combines the two Illinois casinos.

The first method utilized was to run individual regression models for the two Illinois casinos including variables for the St. Louis area's unemployment rate, the opening of Lumiere Place, and the implementation of the smoking ban. This is essentially a multivariate version of the difference-in-means tests examined in section 4, and it allowed us to simultaneously examine the three effects. In particular, we estimated the model as

$$REV_{it} = \alpha_i + \pi_i UE_t + \lambda_i LUMIERE_t + \beta_i SMOKEBAN_t + \varepsilon_{it},$$

where REV_{it} is the revenue (either slot handle or table drop, either in levels or log form) at casino-i (Alton or East St. Louis) during month-t; UE_t is the St. Louis Metropolitan Statistical Area unemployment rate during the month; $LUMIERE_t$ is a dummy variable, equal to one-half in December 2007 and one starting in January 2008 (to represent the fact that Lumiere Place was only open for part of the December 2007); $SMOKEBAN_t$ is a dummy variable, equal to one starting in January 2008; and ε_{it} is a zero-mean, homoskedastic error term assumed to follow a first-degree autoregressive process. Also included (but not shown in the above equation) is a set of month dummies to account for seasonality.

Table 5.6 indicates that the Alton (Ill.) casino's revenues were sizably lower after the smoking ban. The estimated coefficients indicate that (controlling for other factors) the casino's monthly slot handle dropped by about $23 million due to the implementation of the smoking ban. The semi-log model estimates a 20 percent decrease in the slot handle due to the smoking ban. The effect on monthly table drop is statistically insignificant in the linear model, but the semi-log model estimates a 23 percent negative impact.

Table 5.7 indicates that the East St. Louis casino's slot handle was lower following the smoking ban but finds little evidence that the table drop was affected. The linear model estimates indicate that (controlling for other factors) monthly slot handle revenues decreased almost $50 million due to the smoking ban. The semi-log model estimates a 23 percent decrease. Neither model finds a statistically significant impact on table drop.

The second method utilized was to sum the two Illinois casinos' revenues and estimate that

$$REV_t = \alpha + \pi\, UE_t + \lambda LUMIERE_t + \beta SMOKEBAN_t + \varepsilon_t,$$

where REV_{it} (either in levels or its log) is the combined revenue (either slot handle or table drop) at the Illinois casinos (Alton and East St. Louis) during month-t. UE_t, $LUMIERE_t$, and $SMOKEBAN_t$ are as defined earlier.

Table 5.8 indicates that most of the drop in Illinois casino revenues can be attributed to the smoking ban. All else being constant, the implementation of the smoking ban corresponds to a drop of 26 percent (or $86 million per month) in slot handle and a decrease of 22 percent (or $3 million) in table drop. Both effects are statistically significant.

7 Conclusion

The proliferation of legalized gambling in the United States over the past half century has been dramatic. In a fairly short period of time state lotteries have become legal and ubiquitous. Casinos in their various forms—standalone, riverboats, and those attached to other enterprises—have left their previous confines of Las Vegas, Atlantic City, and Native-American reservations. They now dot the landscape, in some cases as sizable tourist attractions.

As legalized gaming expanded, it has become an important source of revenue for state and local governments. For example, many states earmark lottery money for education. In addition, casinos often pay sizable boarding fees and taxes to states, counties, and cities. In many cases tax revenues from the gaming industry are the difference between balanced budgets and sizable cuts to public services.

Public officials have struggled to balance the economic and financial importance of the gaming industry with general efforts to limit access to gaming. In the United States these limits start with restrictions on a casino's location. These range from typical zoning regulations that affect all businesses to gaming-specific rules that limit gaming to boats and Native-American reservations. Additional regulations include loss limits (that place an upper bound on the amount of money that a gambler may lose in one day) and limits on times that gamers may enter the casino. Prior research (discussed in the introduction of this chapter) indicates that such limits reduce the amount of wagering that takes place at casinos.

This chapter examined the effect of indoor smoking bans in the state of Illinois. While the bans are not directed specifically at casinos, the bans have a disproportionate effect on casinos (as well as bars and some restaurants). Using a variety of methodologies, the results indicate that, as expected, the Illinois smoking ban appears to have encouraged a sizable number of gamers to travel across the border to Missouri.

This chapter does not attempt to provide a societal cost-benefit analysis to smoking bans. Aside from the inconvenience to the smokers, the ban appears to reduce casino revenues, which indirectly reduces state and local tax revenues. These must be balanced against the benefits of the ban, which might include the public health benefits of potentially lower smoking rates and the reduction in second-hand smoke exposure for the casino employees. Inconvenience caused by the ban that discourages problem gambling and its potential side effects would be an additional potential benefit. Since most of these issues are nearly impossible to quantify, we do not attempt to pass judgment on the ban.

Appendix

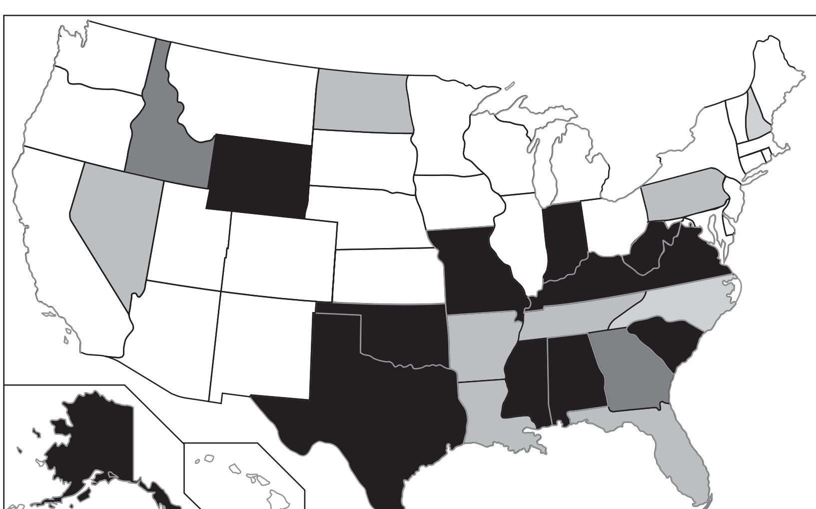


FIGURE 5.1 Statewide smoking bans in the United States

Source: U.S. States Smoking Bans. Wikipedia; http://en.wikipedia.org/wiki/File:US_states_smoking_bans.svg.

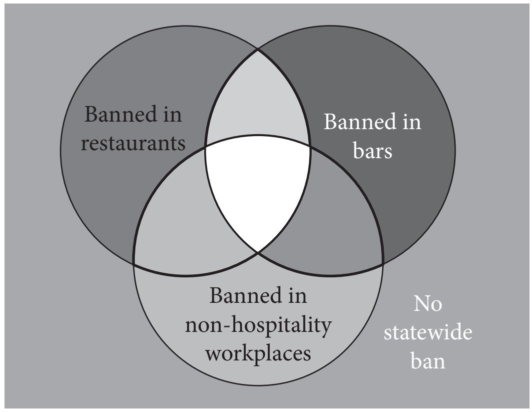


FIGURE 5.1 (*Continued*)

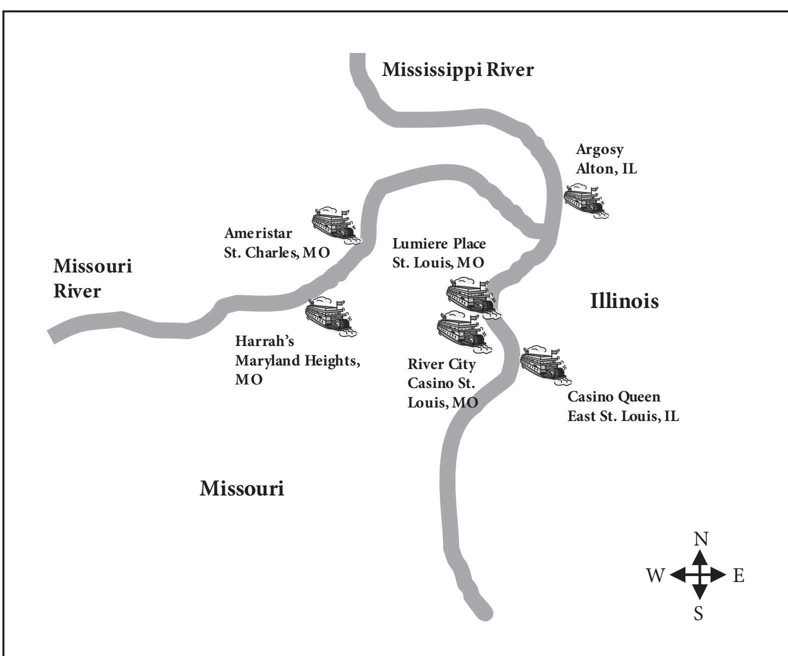


FIGURE 5.2 Casinos in the St. Louis metropolitan area

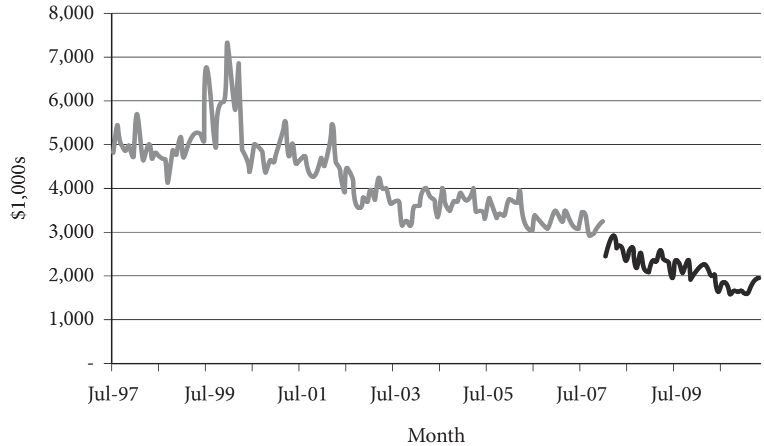


FIGURE 5.3 Table drop, Alton Casino

Source: Illinois Gaming Commission

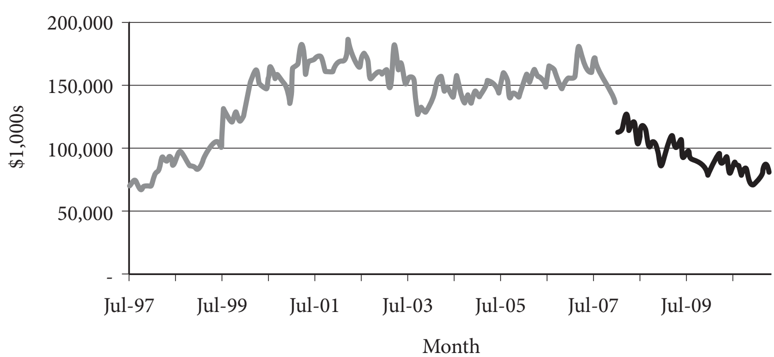


FIGURE 5.4 Slot handle, Alton Casino

Source: Illinois Gaming Commission

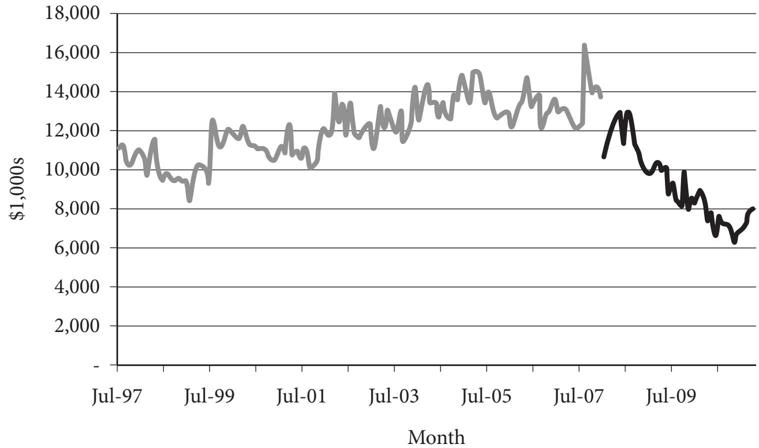


FIGURE 5.5 Table drop, East St. Louis

Source: Illinois Gaming Commission

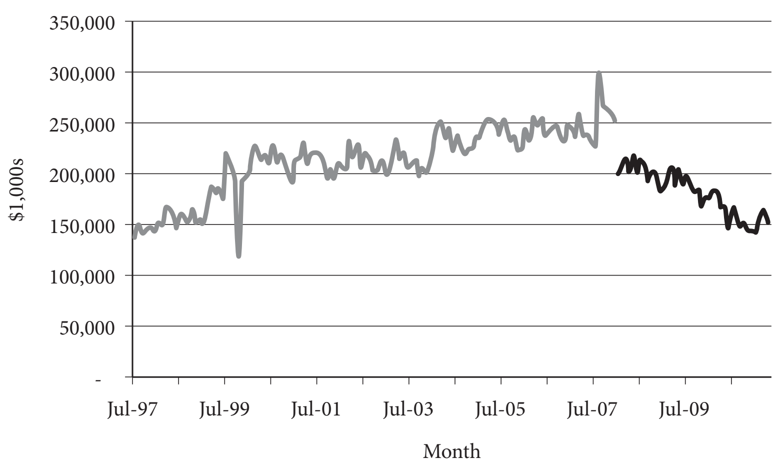


FIGURE 5.6 Slot handle, East St. Louis

Source: Illinois Gaming Commission

Table 5.1 Change-Point Tests for Alton Casino (Ill.)

Time Period	Table Drop	Slot Handle
July 1997–Dec 2007		
Mean	4,244	140,640
n	126	126
Jan 2008–June 2010		
Mean	2,313	99,102
n	30	30
z—statistic	20.3***	12.0***
U—statistic	72***	36***
July 1997–Nov 2007		
Mean	4,252	140,678
n	125	125
Dec 2007–June 2010		
Mean	2,342	100,290
n	31	31
z—statistic	19.3***	11.1***
U—statistic	73***	36***

Z-statistic is from a test of the null hypothesis that the means are equal, versus the alternative that the means are different. Test allows for unequal variances.
U-statistic is from the Wilcoxon Rank-sum test of equivalent distributions.
Means are reported in $1,000s.
*** indicates statistical significance at the 0.01 level.

Table 5.2 Change-Point Tests for East St. Louis Casino (Ill.)

Time Period	Table Drop	Slot Handle
July 1997–Dec 2007		
Mean	12,075	211,873
N	126	126
Jan 2008–June 2010		
Mean	9,959	190,169
N	30	30
z—statistic	6.2***	5.2***
U—statistic	28***	20***
July 1997–Nov 2007		
Mean	12,062	211,545
N	125	125

(*Continued*)

Table 5.2 (*Continued*)

Dec 2007–June 2010		
Mean	10,081	192,192
N	31	31
z—statistic	5.6**	4.2***
U—statistic	25***	17***

Z-statistic is from a test of the null hypothesis that the means are equal, versus the alternative that the means are different. Test allows for unequal variances.
U-statistic is from the Wilcoxon Rank-sum test of equivalent distributions.
Means are reported in $1,000s.
*** indicates statistical significance at the 0.01 level.
** indicates statistical significance at the 0.05 level.

Table 5.3 Annual Slot Handle Sums ($1,000s)

Years	Illinois Casinos (2)	Missouri Casinos (3)	Lumiere
1997	1,282,420*	1,712,296*	
1998	2,937,803	3,791,442	
1999	3,516,918	4,260,818	
2000	4,381,272	5,525,907	
2001	4,367,535	6,507,868	
2002	4,561,257	6,808,298	
2003	4,314,689	7,142,920	
2004	4,546,957	7,659,987	
2005	4,652,003	7,655,842	
2006	4,783,087	7,514,346	
2007	4,894,994	7,044,033	58,019*
2008	3,735,450	6,245,286	1,589,262
2009	3,394,179	6,132,851	1,844,438
2010*	1,548,519*	2,878,366*	840,416*

Sources: Missouri and Illinois gaming commissions. Lumiere opened December 19, 2007. For all casinos, 1997 data include July through December and 2010 data include January through June.

Table 5.4 Annual Table Drop Sums ($1,000s)

Years	Illinois Casinos (2)	Missouri Casinos (3)	Lumiere
1997	94,244*	125,868*	
1998	177,774	242,532	
1999	195,454	254,161	
2000	196,092	288,825	
2001	189,121	302,829	
2002	199,281	306,637	
2003	193,239	305,695	
2004	205,976	322,645	
2005	206,394	308,410	
2006	198,233	315,975	
2007	200,453	304,219	4,794*
2008	169,306	263,750	108,826
2009	138,576	312,626	169,600
2010*	60,275*	153,357*	75,216*

Sources: Missouri and Illinois gaming commissions. Lumiere opened December 19, 2007. For all casinos, 1997 data include July through December and 2010 data include January through June.

Table 5.5 Percentage in Illinois

Time Period	Table Drop	Slot Handle
July 1997–Dec 2007		
Mean	40.1	40.8
N	126	126
Jan 2008–June 2010		
Mean	33.6	36.2
N	30	30
z—statistic	6.5***	12.1***
U—statistic	24***	53***

The variable measures the percentage of revenue at the five preexisting casinos that is accounted for by the two Illinois casinos.
Z-statistic is from a test of the null hypothesis that the means are equal, versus the alternative that the means are different. Test allows for unequal variances.
U-statistic is from the Wilcoxon Rank-sum test of equivalent distributions.
*** indicates statistical significance at the 0.01 level.

Table 5.6 Coefficient Estimates for Alton (Ill.) Casino

	Slot Handle	*Ln* (Slot Handle)	Table Drop	*Ln* (Table Drop)
UE	−1,136	−0.004	−227***	−0.05***
	(1,887)	(0.01)	(51)	(0.02)
LUMIERE	−13,632	−0.09	−517	−0.17
	(13,387)	(0.10)	(670)	(0.14)
SMOKEBAN	−22,863**	−0.20***	−559	−0.23**
	(9,954)	(0.08)	(549)	(0.12)
R^2	0.96	0.97	0.92	0.95
N	156	156	156	156

Sample is July 1997 through June 2010.
Standard errors are given in parentheses.
Slot handle and table drop are measured in $1,000s. *UE* is measured in percentage points.
The model also includes month dummies. Estimated correcting for *AR*(1) errors.
*** indicates statistical significance at the 0.01 level.
** indicates statistical significance at the 0.05 level.
** indicates statistical significance at the 0.10 level.

Table 5.7 Coefficient Estimates for East St. Louis (Ill.) Casino

	Slot Handle	*Ln* (Slot Handle)	Table Drop	*Ln* (Table Drop)
UE	7,502**	0.04**	214	0.02
	(3,386)	(0.02)	(183)	(0.02)
LUMIERE	−9,313	−0.06	−1,829	−0.16
	(27,269)	(0.15)	(1,467)	(0.13)
SMOKEBAN	−49,918**	−0.23*	−1,538	−0.13
	(22,005)	(0.12)	(1,178)	(0.10)
R^2	0.83	0.81	0.84	0.85
N	156	156	156	156

Sample is July 1997 through June 2010.
Standard errors are given in parentheses.
Slot handle and table drop are measured in $1,000s. *UE* is measured in percentage points.
The model also includes month dummies. Estimated correcting for *AR*(1) errors.
*** indicates statistical significance at the 0.01 level.
** indicates statistical significance at the 0.05 level.
** indicates statistical significance at the 0.10 level.

Table 5.8 Coefficient Estimates for Combined Illinois Casinos

	Slot Handle	*Ln* (Slot Handle)	Table Drop	*Ln* (Table Drop)
UE	6,407	0.02	−87	−0.01
	(4,733)	(0.01)	(177)	(0.01)
LUMIERE	−14,045	−0.04	−920	−0.06
	(17,483)	(0.05)	(841)	(0.06)
SMOKEBAN	−86,053***	−0.26***	−3,114***	−0.22***
	(17,400)	(0.05)	(845)	(0.06)
R^2	0.93	0.93	0.85	0.87
N	156	156	156	156

Sample is July 1997 through June 2010.
Standard errors are given in parentheses.
Slot handle and table drop are measured in $1,000s. *UE* is measured in percentage points.
The model also includes month dummies. Estimated correcting for *AR*(1) errors.
*** indicates statistical significance at the 0.01 level.
** indicates statistical significance at the 0.05 level.
** indicates statistical significance at the 0.10 level.

Table 5.9 Summary of Results

Strategies	Slot Handle	Table Drop
Change Point		
Alton	no impact	no impact
East St. Louis	negative	negative
Missouri Control Group		
Combined	negative	negative
Regression		
Alton	negative	negative
East St. Louis	negative	no impact
Combined	negative	negative

Notes

1. Brunker, Mike. "Riverboat Casinos Going Nowhere Fast," msnbc.com, Aug. 9, 2004; http://www.msnbc.msn.com/id/5539243/ns/news-the_mighty_miss/t/riverboat-casinos-going-nowhere-fast/#.TonA_mUfeEw.
2. This chapter is based on the paper "Will Riverboat Profits go up in Smoke? The Effect of the Illinois State Smoking Ban on Casino Revenue," presented at the Midwest Economics Association Annual Meetings, Cleveland, March 2009.

References

Douglas, John H. 2008. Smoking bans in tribal casinos: Health issue or labor's latest smokescreen assault on tribal sovereignty? *Indian Gaming*, May, 22–25.

Garrett, Thomas A., and Michael R. Pakko. 2010. *The revenue performance of casinos after a smoking ban: The case of Illinois.* Working Paper No. 2009–027B. St. Louis, Mo.: Federal Research Bank of St. Louis.

Glantz, S. A., and R. Wilson-Loots. 2003. No association of smoke-free ordinances with profits from bingo and charitable games in Massachusetts. *Tobacco Control* 12:411–413.

Grant, Jon E., Suck Won Kim, Brian L. Odlaug, and Marc N. Potenza. 2008. Daily tobacco smoking in treatment-seeking pathological gamblers: Clinical corellates and co-occurring psychiatric disorders. *Journal of Addiction Medicine* 2(4):178–184.

Harper, Todd. 2003. Smoking and gambling: A trance induced ritual? *Tobacco Control* 12(2):231–233.

Harris, Jenine K., Bobbi J. Carothers, Douglas A. Luke, Hiie Silmere, Timothy D. McBride, and Martin Pion. 2011. Exempting casinos from the smoke-free Illinois act will not bring patrons back: They never left. *Tobacco Control* 12(3):373–376.

Lal, A., and M. Siahpush. 2008. The effect of smoke-free policies on electronic gaming machine expenditure in Victoria Australia. *Journal of Epidemiology and Community Health* 62:11–15.

Macy, Jonathan T., and Erika L. Hernandez. 2011. The impact of a local smoke-free law on wagering at an off-track betting facility in Indiana. *Tobacco Control* 20(5):383–384.

Mandel, L. L., B. C. Alamar, and S. A. Glantz. 2005. Smoke-free law did not affect revenue from gaming in Delaware. *Tobacco Control* 14(1):10–12.

McGowan, John R., and Muhammad Islam. 2003. Economic effects of proposed tax increases and removal of loss limits on the Missouri gaming industry. White paper. St. Louis: Taxpayers Research Institute of Missouri and Associated Industries of Missouri.

McGrath, D. S., and S. P. Barrett 2009. The comorbidity of tobacco smoking and gambling: A review of the literature. *Drug and Alcohol Review* 28(6):676–681.

Minnesota Gambling Control Board. 2008. Charitable gambling impact study: A brief review of the fiscal impact of a statewide smoking ban on lawful gambling. Roseville: Minnesota Gambling Control Board, Department of Revenue.

Navin, John C., Timothy S. Sullivan, and Warren D. Richards. 2009. *Will riverboat profits go up in smoke? The effect of the Illinois' state smoking ban on casino revenue.* Edwardsville: Department of Economics and Finance, Southern Illinois University.

Pakko, Michael R. 2006a. Smoke-free law did affect revenue from gaming in Delaware. *Tobacco Control* 15(1):68–69.

———. 2006b. *No smoking at the slot machines: The effect of a smoke-free law on Delaware gaming revenues.* Working Paper No. 2005-054C. St. Louis, Mo.: Federal Reserve Bank of St. Louis.

———. 2008. Clearing the haze? New evidence on the economic impact of smoking bans. *Regional Economist*, Jan., 10–11.

Petry, N. M., and C. Oncken. 2002. Cigarette smoking is associated with increased severity of gambling problems in treatment seeking pathological gamblers. *Addiction* 97(6): 745–753.

Pettit, A. N. 1979. A non-parametric approach to the change-point problem. *Applied Statistics* 28(2):126–135.

Pyles, M. K., and E. J. Hahn. 2009. Smoke-free legislation and charitable gaming in Kentucky. *Tobacco Control* 18(1):60–62.

Rodda, Simone, Stephen L. Brown, and James G. Phillips. 2004. The relationship between anxiety, smoking, and gambling in electronic gaming machine players. *Journal of Gambling Studies* 20(1):71–81.

Thalheimer, Richard, and Mukhtar M. Ali. 2008. The demand for casino gaming with special reference to a smoking ban. *Economic Inquiry* 46(2):273–282.

CHAPTER 6

OVERVIEW OF THE ECONOMIC AND SOCIAL IMPACTS OF GAMBLING IN THE UNITED STATES

DOUGLAS M. WALKER

Over the past two decades academic interest in gambling behavior and the economic and social impacts of legal gambling has increased significantly. Researchers in many nations have devoted more attention to gambling issues as the industry has grown. The majority of published research has focused on problematic gambling behaviors and the diagnosis and treatment of them. The research on gambling *behavior* has grown significantly. However, a relatively small portion of gambling research has been performed in the business and economics disciplines. This is curious, given that the major reasons casinos and other forms of legal gambling exist are economic in nature. Indeed, despite the negative impact that the 2007–2009 recession has had on the casino industry worldwide, governments continue to look toward legal casinos as way to alleviate fiscal stress. Nowhere is this more evident than in the United States, as numerous states are in the process of or are currently considering legalizing casinos. A similar pattern can also be seen in countries across the globe.

The purported economic benefits from casino gambling include tax revenues, increased employment, higher wages and payments to capital, and enhanced economic growth. These benefits, should they occur, are not necessarily without costs. For example, the casino industry may partially or entirely cannibalize other industries. In addition, a small percentage of gamblers may exhibit problem gambling behavior. Such people are believed to cause significant social costs. The analyses of these benefits and costs of gambling—the economics of gambling—is a young field of research, with only a handful of researchers actively researching the various issues.

This chapter describes some of the critical economic and social issues, mainly related to casino gambling, and my research on them. Although the empirical analysis tends to focus on the United States, the various issues are relevant to all countries that have, or are considering adopting, commercial casino gambling. This chapter is organized into six sections, by topic. Section 1 examines the explanatory factors in the adoption of commercial casinos. Section 2 is a discussion of the economic growth effects of casino gambling, including how gambling can affect growth after a natural disaster. Section 3 discusses the relationships among different gambling industries and the implications of these relationships on government tax revenues. Section 4 is an introduction to the social costs of gambling, including the relationship between casinos and crime. In Section 5, I describe some the hurdles in cost-benefit analysis as it applies to gambling. Section 6 concludes.

1 Determinants of Casino Adoption

The legalization of casino gambling is typically a controversial issue. Disagreement over casinos arises from moral and religious objections to gambling, concerns over potentially negative social impacts, as well as uncertainty as to the economic benefits from legalization. Clearly, recent history has shown that casinos have been legalized in a variety of jurisdictions, for a number of different reasons. Perhaps the greatest motivation for introducing casinos is to raise tax revenues. Casinos are generally taxed at a relatively high rate and therefore offer politicians a relatively easy source of revenue. Yet, casinos are not always welcome, as demonstrated by the ongoing debate over casino adoption in the Penghu Islands, Taiwan. A recent newspaper article suggests that Kyrgyzstan may ban all casinos and online casinos (Pumper 2011). U.S. states that already welcome legalized gambling in the form of lotteries are sometimes unsympathetic to the prospects of casino gambling. What factors help to explain why casinos are—or are not—legalized? This issue has not been analyzed for casinos, until recently. Peter Calcagno, John Jackson, and I recently examined this issue for the United States (Calcagno, Walker, and Jackson 2010). The economics literature has a number of papers that have examined the adoption of lotteries. Our study followed this work and applied a similar model to analyze what factors seem to explain the adoption of casinos in the United States.

Prior to 1989, commercial casinos were legal only in Las Vegas, Nevada, and Atlantic City, New Jersey. After an important legal decision (*California v. Cabazon Band of Mission Indians* 1987) and subsequent legislation (Indian Gaming Regulatory Act 1988) the stage was set for commercial casino legalization. By 1995 commercial casinos had been legalized in eight states; thirteen states had them by 2010. Tribal casinos now operate in around 30 states. If casinos represent an easy source of tax revenues, and if a state's population can easily travel to out-of-state casinos, why not just legalize casinos in your state? The interesting question may be why more U.S. states have not introduced commercial casinos. Calcagno, Walker, and Jackson (2010) examined state governments' adoption of commercial casinos; it did not examine tribal casino decisions.

Our analysis followed the earlier analysis by Jackson, David Saurman, and William Shughart (1994) in explaining lottery adoption in the United States We posited a Tobit model to explain the probability and timing of casino adoption using state-level annual data from 1985–2000, a period covering most of the casino expansion outside of Las Vegas and Atlantic City. Among the variables included in the model are "fiscal" variables, such as long- and short-term debt; whether the state has tax/expenditure limits, such as balanced budget provisions; the level of state government revenue; and the amount of federal government transfers to the state government. The model also includes variables measuring the influence of the different political parties in the states as well as the existing gambling opportunities within the state and in nearby states. Finally, a variety of demographic variables were included.

Calcagno, Walker, and Jackson (2010) found mixed evidence to support the proposition that "fiscal stress" explains casino adoption in U.S. states. They found more clear evidence that interstate competition helps to explain casino adoption. States seem to legalize casinos in order to attract tourism and to keep their own gamblers in the state ("defensive legalization"). States are more likely to introduce casinos—and to do it sooner—if neighboring states have casinos. But there is little evidence that intrastate competition among gambling industries is relevant to the decision to adopt casinos. Hence, the evidence is consistent with common sense, that legislators look to casinos mainly as a way to increase tax revenues.

Although Calcagno, Walker, and Jackson (2010) examined data for the United States, the same framework can be utilized at an international level to explain why some countries have introduced casinos and others have not. The analysis could also be applied to a more local level to explain why some communities welcome casinos and others do not. Clearly the decision to adopt casinos depends, at least in part, on how legislators and voters believe casinos will affect the local or state economy. Several related issues are examined below.

2 Casino Gambling and Economic Growth

Although the casino industry argues that it spurs local and regional economic growth by providing high-paying jobs and paying taxes and fees to local and state governments, there is little empirical research on the issue. Studies such as Arthur Andersen (1996), commissioned by the casino industry, are biased and amount to little more than static comparisons or listings of taxes paid and employees hired by the casino industry.

The lack of empirical studies on the U.S. casino industry is not surprising given the relatively recent expansion of casinos. Even so, there have been few studies on the economic effects of the casino industry. The studies that have been published tend to focus on the relationships among casinos, other gambling industries, and tax revenues. Other studies have examined the negative consequences of casino gambling and pathological gambling behavior, such as crime and bankruptcy. Few studies have examined whether casinos stimulate economic growth or supplement state government revenues.

Walker and Jackson (1998) were among the first to study the effects of the new casino industry in the United States. We analyzed the relationship between state-level casino revenues and per capita income. To do this, we developed a process for adapting Granger causality testing to panel data. The reason we utilized a panel was because at the time of the analysis relatively few states had casino gambling. The states that did have casinos, other than Nevada and New Jersey, had them for at most six years. Hence it was impossible at the time to analyze the states individually. In a more recent paper (Walker and Jackson 2007) we repeated the earlier study using annual data through 2005. The process we used to adapt Granger causality testing for use with panel data is described below.

2.1 Granger Causality Applied to Panel Data

Granger causality is said to exist between two variables, say x and y, when past values of one variable (x) significantly enhance the ability to predict future values of the other variable (y). The implication is that the first variable is affecting or "causing" the second. Admittedly, Granger causality does not *prove* the two variables are related, and it does not imply that the one variable is the only, or even most important, factor affecting the other variable. What it does do is allow us to assess the relative likelihood of the following four possibilities: (1) x and y are not related, (2) x Granger causes y, (3) y Granger causes x, or (4) x and y Granger cause each other.

In order to adapt Granger causality analysis to panel data, Walker and Jackson (1998, 52–55) proposed a three-step process: (1) detrending the data, (2) selecting the appropriate time series process that generates each variable, and (3) conducting the Granger causality tests based on the results of the two previous steps. Our goal was to analyze whether there was a Granger causal relationship between casino revenue and economic growth (per capita income) at the state level.

The first step involved detrending the casino revenue and per capita income data. The basic goal was to extract from the data any systematic information associated with state-specific factors (laws, institutions, etc.), time-trend factors, and any idiosyncrasies of the data or data collection. The detrended variables, that is, the residuals from these filtering equations, should be stationary series. This is tested using a unit root test, such as Phillips-Perron. Once the detrended series were confirmed to be stationary we moved to the next step.

Step (2) involved determining the time series (autoregressive or ARMA) process that generated each variable. In other words, we tried to determine how many lagged periods of each variable had a significant predictive power for current observations of the filtered data. The goal was to use the shortest possible lag length for each series such that no systematic relationship remained among the residuals of the estimated process. Once the proper lag length has been determined the Granger causality test is set up. This was step (3) in the testing procedure. It involved estimating a two-equation vector autoregressive (VAR) system in which the current value of each filtered variable was regressed on the appropriate number of past values for both variables. Then a set of F-tests was performed to test whether the filtered residuals have a Granger causal relationship.

2.2 Results

The results of the two studies (Walker and Jackson 1998, 2007) suggest that there is a short-term positive impact of casino gambling on economic growth but that the effect dies out in the longer term. The 1998 study used quarterly data from the period 1978–1996. Our results showed that there was a Granger causal relationship for casino revenue on per capita income. These results are perhaps not surprising. The effects of capital expansion and increased demand on the labor market that would come with the new casino industry in a state could be expected to have a positive impact on income. Joseph Schumpeter (1934) indicated that one possible source of economic growth is the introduction of a new good/service to an economy. Our results are suggestive of such an effect.

When we repeated the analysis more recently (2007), we used annual data from the period 1991–2005. Annual data are preferred to quarterly data, especially since we had to interpolate the quarterly per capita income data from annual observations. The more recent analysis indicated no Granger causality relationship between casino revenues and per capita income. Hence we argued that the introduction of casino gambling has a short-run stimulus effect but that it eventually dies out. Perhaps this can be explained by competition for the gambling dollar with other legal gambling industries within the state itself or through direct competition with gambling opportunities online or in neighboring states. Or perhaps casinos simply replace or cannibalize other non-gambling industries within the state. These issues are discussed in more detail in Section 3.

2.3 An Extension: Hurricane Katrina

As a further application of the theory that casinos cause economic growth, in two recent studies Walker and Jackson (2008a, 2009) examined the effect the casino industry had on the economic recoveries of Mississippi and Louisiana following Hurricane Katrina in 2005. The hurricane completely devastated the casino industry in both states, but shortly after the hurricane the industry began to rebuild. Using quarterly personal income and casino revenue data, we tested the impact that the casinos had on personal income in Katrina-affected states. Our model included variables to account for the hurricane and casino activity after the hurricane. The results suggest that the commercial casino industry has had a significantly positive impact on state-level personal income and that after the hurricane the effect was larger than the “normal” casino effect. Consistent with our earlier papers, the Katrina study suggests that casinos can indeed have a positive impact on state-level economic growth, at least in the short-term. Presumably these effects come about from an amalgamation of capital and labor effects and the attraction of tourism.

The available empirical evidence suggests that casinos do indeed have a positive economic growth effect, though it may be short-lived. Obviously the effect will vary depending on specifics of the jurisdiction and market.

3 Relationships among Gambling Industries

Whether an economic growth impact exists may be of less concern to politicians than the amount of tax revenues that casino or other gambling activities create for the government. Of course the various gambling industries often point to the taxes they pay as a measure of the tax relief provided by the industry to the state's (or local) citizens. For example, Missouri taxpayers and politicians may assume that the $742 million in lottery sales or the $403 million in taxes paid by the casino industry represent net increases in tax revenues (or reductions to the citizens' tax burdens). This may be the case, but it is more likely that government spending has increased or that the taxes raised by gambling are offset by tax losses from other types of consumer expenditures. The tax revenue effect is not as simple as it might at first seem.

A number of studies have examined the impact of one gambling industry on other gambling or non-gambling industries. Other papers have examined the impact of gambling on state tax revenues.[1] Overall, these studies find that one industry either harms another industry or does not affect it. No study has found that different gambling industries help each other. In addition, the effect of gambling on state tax revenues is mixed. Some of the published findings are summarized in table 6.1.

These studies have several limitations in terms of understanding general relationships among U.S. gambling industries and their overall impact on state government revenues. First, for the most part these studies examine a single industry's effect on another industry but not vice versa. Second, most of the studies examine a single state or county. Third, there are some limitations to the types of models used. For example, many of the studies account for the gambling industry only through a dummy variable. Accounting for the mere existence of a casino, for example, is much less enlightening than accounting for the size of the industry.

Jackson and I have provided a general analysis of the relationships among the different gambling industries within the United States (Walker and Jackson 2008b). We also have examined the overall impact of legalized gambling on state government revenues (Walker and Jackson 2011). Our findings from these studies are described below.

Table 6.1 Studies on the Relationships among Gambling Industries

Paper	Years	States/Counties	Findings
Anders, Siegel, and Yacoub (1998)	1990–1996	1 county (Ariz.)	Indian casinos cause a reduction in tax revenue.
Borg, Mason, and Shapiro (1993)	1953–1987	10 states	Lotteries cause a decline in some other tax revenues, but total tax revenues increase.
Elliot and Navin (2002)	1989–1995	All states	Casinos and pari-mutuels harm lotteries.
Fink, Marco, and Rork (2004)	1967–1999	All states	Net increase in lottery revenue causes a decrease in state aggregate tax revenues.
Kearney (2005)	1982–1998	All states	Lotteries do not harm other forms of gambling.
Popp and Stehwien (2002)	1990–1997	33 counties (N.M.)	Indian casinos reduce county tax revenues.
Siegel and Anders (1999)	1994–1996	1 state (Mo.)	A 10% increase in gambling tax revenue leads to a 4% decline in other tax revenues.
Siegel and Anders (2001)	1993–1998	1 state (Ariz.)	Slots harm lottery; horse and dog racing do not affect lottery.
Thalheimer and Ali (1995)	1960–1987	3 tracks (Ohio, Ky.)	Having both lottery and horse racing increases tax revenues.

Source: Walker and Jackson (2008b).

Since no paper had analyzed the general interindustry relationships for gambling in the United States, we attempted to model the revenue for each type of gambling industry in each state as a function of other in-state gambling industries, adjacent state gambling industries, and a variety of demographic criteria. We collected annual data on the volume of each type of gambling for the period 1985–2000: lottery, horse racing, greyhound racing, commercial casino, and Indian casinos.[2] Using the data that were available, we have 816 observations.

We are attempting to explain gambling revenue. Of course, states elect whether to offer particular types of gambling or not. Therefore, there is a self-selection issue that must be dealt with. The dependent variables (industry volume) in our model are left-censored, especially in the cases of casino gambling and horse racing, as fewer states offer these forms of gambling. To deal with the left-censoring, we followed James Heckman (1979) and obtained the inverse Mills ratio from a probit and included it in the model of gambling revenue as an additional explanatory variable.

Table 6.2 Summary of Intrastate Industry Relationships

Model Variable	Casino	Dog Racing	Horse Racing	Lottery
Casino		−	+	−
Dog racing	(−)		−	+
Horse racing	+	−		+
Lottery	−	+	+	
Indian square foot	+	(+)	+	−

( ) indicates statistically insignificant at normal levels.

Source: Walker and Jackson (2008b).

We wanted to model state-level volume of casino gambling, lottery revenue, and dog and horse racing handle. (We did not attempt to model Indian casino square footage.) Because of the nature of the data, we elected to use a seemingly unrelated regression (SUR) analysis. This procedure allowed us to estimate our four-equation model jointly as a system of equations rather than apply OLS to each equation independently. A summary of the results for the interindustry relationships is provided in table 6.2. A positive sign indicates a positive and statistically significant coefficient; a minus sign indicates a negative and significant coefficient; and signs shown in parentheses are statistically insignificant. As shown, the results are mixed. Some industries appear to help each other (complements), such as casinos and horse racing, lotteries and dog racing, and horse racing and lotteries. Others, such as lotteries and casinos, and dog and horse racing, are apparently substitutes.

These results supplement the existing literature by providing a more general analysis on the interindustry relationships. Obviously the relationships among industries may vary by state, as each state has unique laws, demographics, and so on. But the evidence provided by Walker and Jackson (2008b) may be helpful to policy makers and voters currently considering the legalization of new types of gambling. One thing that our analysis suggests quite clearly is that relationships among gambling industries are not straightforward, obvious, or consistent.

3.1 Tax Revenues

The interindustry relationships among gambling industries will obviously have an impact on the effect of legalized gambling on overall state tax revenues. For example, if a state that currently has horse racing wishes to increase tax revenues and is considering legalizing a new type of gambling, it may wish to consider a lottery but not greyhound racing, as Walker and Jackson (2008b) found that lotteries and racing tend to be complements whereas horse and greyhound racing act as substitutes. Just knowing the relationships among the industries is not enough, however. Whether a new type of gambling will increase or decrease overall state tax revenues depends on the interindustry gambling relationships, the relationships between gambling and non-gambling industries, the taxes applied to gambling and non-gambling industry expenditures by consumers, and possibly other factors.

If one considers the level of taxes typically levied on gambling industries, it would seem obvious that legalized gambling will tend to increase state government revenues. Consider, for example, that the average state sales tax rate is somewhere around 5 percent. This tax applies to most consumer goods and often to services.[3] The state lottery typically represents about a 30 percent tax. That is, for each $1 ticket, approximately 50¢ is returned as prizes, 20¢ goes toward administrative costs, and roughly 30¢ are kept as government revenue—effectively tax revenue. This breakdown for each lottery ticket is moderately consistent across different states with lotteries. Casinos, on the other hand, are taxed at rates that vary by state. Typically the gross casino revenues are taxed, and then the casino also pays standard income taxes on any remaining profit, as required in most states. The gross gaming taxes range from a low of around 6 percent in Nevada to a high of 55 percent in Pennsylvania.[4] Whatever the state, it is safe to assume that taxes applied to lotteries or casinos are higher than the regular sales tax. Then it would seem that, even if 100 percent of casino and lottery revenues in a state come at the expense of other non-gambling expenditures, casino and lotteries should increase net state revenue.

We tested this proposition using the same 1985–2000 state-level data as in the interindustry study (2008b) and performed an econometric analysis to determine whether there are some general relationships between gambling industries and tax receipts across states (Walker and Jackson 2011). After controlling for a variety of gambling industry metrics and demographic variables, we found mixed results. In particular, we found that lotteries and horse racing have statistically significant positive effects on state government revenues, but casinos and greyhound racing seem to reduce net government revenues. While the lottery finding was expected, the negative casino result was surprising. Our gambling industry variables included a dummy variable for the existence of each type of gambling in the state as well as a "marginal impact" variable to measure the effect on state revenue from each additional dollar of handle in each industry. The results are summarized in table 6.3.

Table 6.3 Summary of Gambling Industry Effects on Net State Government Revenue

Industry Variable	Casino	Dog Racing	Horse Racing	Lottery
Presence of Industry	−$90m.	−$157m.	$671m.	$315m.
Marginal Impact of $1 Handle	−$0.07	−$7.61	−$1.46	−$0.30*

* indicates statistical *insignificance*. "m." represents millions.

Source: Walker and Jackson (2011).

As table 6.3 shows, the existence of the casino industry in a state corresponds to a reduction in state government revenues, for the average casino state, of $90 million. Each additional dollar of handle has a relatively small negative impact on state revenue of only 7¢.[5] The lottery results show a large positive "presence" effect, with a relatively small decline in state revenue from the marginal dollar from ticket sales. Oddly, horse and dog racing seem to have different effects—horse racing has, on net, a positive impact on state revenue, while greyhound racing appears to have a significantly negative impact.[6]

Overall, our results suggest that lotteries and horse racing have a positive impact on state government revenues but that casinos and greyhound racing actually have a negative impact. The negative result on casinos may indicate that casino expenditures come at the expense of non-casino expenditures to such a large extent that, despite the high tax rates applied to casino revenues, the reductions in non-casino spending lead to declines in sales tax revenues that are even larger. This result surprises us and should be considered by states considering the expansion of existing casinos or the legalization of new ones.

The Walker and Jackson (2011) paper provides a more general analysis than previous studies on the tax impacts of legalized gambling.[7] While state-specific studies often show a positive impact from casinos, our results suggest that, on average, casino gambling probably does not have a positive effect on state revenues. Obviously there will be exceptions. Las Vegas is a prime example of a city in which the casino industry obviously has a positive impact on state government revenues. But the casino industry in other states may not attract as many tourists and may therefore not have a positive impact on state revenues. More study on the tax issue at the market or state level is needed. But our evidence suggests that states should at least be aware that the casino effect is not always necessarily positive with respect to net tax revenues.

4 The Social Costs of Gambling

Whatever economic benefits casinos provide, whether in terms of economic growth, additional tax revenues, or simply an additional choice of entertainment for consumers, there is a potential downside of legalized gambling. In particular, about 1 percent of the general population is believed to comprise pathological gamblers. These individuals are believed to cause an enormous amount of social costs which at least partially offset any economic benefits from gambling.[8]

The gambling literature is fascinating in part because it is the product of researchers from very different academic perspectives. For example, published papers on the social costs of gambling have come from researchers with backgrounds in psychology, sociology, law, political science, public administration, and even landscape architecture. One consequence of having researchers with different areas of expertise discussing a particular issue, such as social costs, is that they often come to very different conclusions—even more so than economists do.

When economists discuss social cost, they usually have something very particular in mind. As a result, a traditional economic analysis of social cost leads to startlingly different conclusions than are otherwise found in the gambling literature. For example, one early social cost analysis performed by William Thompson, Ricardo Gazel, and Dan Rickman (1997) estimated the annual social costs of gambling per pathological gambler at $9,469. A more recent study by Thompson and Keith Schwer (2005) estimated the social costs of pathological gambling in Las Vegas at $19,711. Earl Grinols (2004) averaged a variety of (mostly unpublished and flawed) studies to arrive at a cost estimate of $10,330. Such diversity in cost estimates indicates that the different studies have not measured social costs in the same way.

Questions about research quality/legitimacy have been raised in comprehensive analyses (Australian Government Productivity Commission 1999; National Gambling Impact Study Commission 1999; National Research Council 1999) as well as in more narrow critiques. For example, the National Research Council (1999, 186) explained that "most [cost studies] have appeared as reports, chapters in books, or proceedings at conferences, and those few that have been subject to peer review have . . . been descriptive pieces." The result has been questionable, if not counterproductive, research: "In most impact analyses . . . the methods used are so inadequate as to invalidate the conclusions. Researchers . . . have struggled with the absence of systematic data that could inform their analysis and consequently have substituted assumptions for their missing data" (185).

To illustrate the problems with social cost estimates, consider the work by Thompson and Schwer (2005), summarized in table 6.4. The cost estimate is based on a survey of 99 Gamblers Anonymous members in Las Vegas. The fundamental problem with this study, and others in the literature, is that the authors have failed to define "social cost."[9] Instead, they attempt to estimate monetary values for any negative effects that they can identify, measure, and somehow attribute to gambling. To be sure, there is a lot of jargon in the gambling literature that may be confusing to researchers. All of the following terms describing "costs" have been used in recent papers: private, social; internal, external; direct, indirect; harms, costs; intangible, tangible; external costs, externalities; and pecuniary externalities, technological externalities.

Walker and A. H. Barnett (1999) and Walker (2003) explained why the definition of social cost is important and why researchers should be skeptical of cost estimates that do not explicitly define what they are trying to measure. We give a detailed explanation of the welfare economics (utilitarian) perspective on social costs. Essentially we argue that a social cost requires that an action reduces the total "wealth" in society. This implies that wealth transfers (gambling losses, bad debts, etc.) cannot be considered as social costs.[10] This follows from Gordon Tullock's (1967) classic discussion of theft. In addition, internalized costs would not qualify as social in nature. Many of the so-called social costs estimated by Thompson, Gazel, and Rickman (1997), Thompson and Schwer (2005), and Grinols (2004) turn out to be wealth transfers or internalized costs.

Table 6.4 A Typical Estimate of the Social Costs of Gambling (in USD)

Employment		$ 5,125
missed work	2,364	
productivity losses (quit jobs)	1,092	
fired from work (productivity lost)	1,582	
unemployment compensation	87	
Bad Debts and Civil Court		$10,271
bankruptcy debt loss	9,494	
civil court costs (bankruptcy/debt/divorce)	777	
Criminal Justice System		$ 3,809
theft	3,379	
arrests	95	
trials	85	
incarceration	80	
probation	170	
Treatment and Social Services		$ 506
treatment costs	372	
welfare	84	
food stamps	50	
Total estimated annual social cost per pathological gambler		**$19,711**

Source: Thompson and Schwer (2005, 83)

As a result, we argue that the social cost estimates in the literature seriously overstate the actual social costs of gambling.

If the Thompson and Schwer (2005) estimate is reviewed from an economic perspective, many of the presumed costs drop out from the social cost estimate. For example, all of the employment costs of $5,125 would either be internalized by the employer or employee or they represent transfers of wealth. The bankruptcy debt losses, monetary value of theft, and welfare/food stamps would be transfers. The actual social costs, as defined by Walker and Barnett (1999), would include civil court costs, criminal justice costs, and treatment of pathological gambling. These costs are estimated at about $1,600 (Walker 2008a), significantly different from the $19,711 estimate of Thompson and Schwer.[11] This "economic perspective" is upsetting to many social scientists because it seems to ignore some potentially significant harms. This exercise demonstrates that the definition of social cost *does* matter, especially considering that social cost studies have had a real impact on government policy toward gambling.

4.1 Casinos and Crime

One specific social cost issue that has received recent attention in the literature is the potential relationship between casinos and crime. In probably each jurisdiction in which casinos are being considered there is debate over whether casinos will create or attract crime. For example, the fact that casinos will attract tourists carrying cash might be a catalyst for criminals to flock to casinos—customers may represent easy prey. Alternatively, people who develop into problem gamblers may turn to crime to get money for gambling. Any number of situations might suggest a relationship between casinos and crime.

The study by Grinols and David Mustard (2006) purported to show through a county-level analysis that casinos have, in fact, contributed to crime in the United States. Grinols and Mustard used an exhaustive data set in an econometric analysis of how casinos tend, with some time lag, to contribute to increases in crime rates. Yet, as explained more recently (Walker 2008b), the empirical analysis by Grinols and Mustard (2006), as well as that in other studies, is suspect because they often mismeasure the crime rate.[12]

To explain, consider a community in which there is relatively little tourism, so that crimes are committed by residents and the victims also are residents of the community. Then the crime rate is represented as the number of crimes committed divided by the population:

$$\text{Crime rate} = \frac{\text{\# of crimes committed}}{\text{population at risk}}$$

This rate is often expressed per 100,000 people. Let's use C to represent criminal acts and P to represent population at risk of being victimized. Then the crime rate above is C/P. Now in order to apply the crime rate to a case in which casinos are introduced, we must recognize that likely there will be an inflow of tourists into the jurisdiction that introduces casinos. Simply because the number of people in the area has increased, we normally would expect an increase in the raw number of crimes committed. But the actual risk of being a crime victim may rise *or fall*; it depends on whether the amount of crime increases by a larger or smaller amount than the increase in the population at risk.

More formally, if we now distinguish between criminal acts perpetrated by residents (C_R) and those perpetrated by visitors (C_V); and separate the population at risk into the resident population (P_R) and the visiting population (P_V), then we can see how the crime rate would be represented in a case where there is a substantial amount of tourism. Such is the case with many casino jurisdictions. The crime rate in this case can be written as $(C_R + C_V)/(P_R + P_V)$. As I emphasized in my critique of Grinols and Mustard's paper (Walker 2008b), it is important that the crime rate accurately reflect the risk of being victimized. When a jurisdiction experiences a large number of tourists, both residents and visitors may commit and/or be victimized by crime.

In the analysis by Grinols and Mustard (2006), the authors counted the crimes committed by visitors (C_V) in the numerator but omitted the visitors from their measure of the population at risk, in the denominator. So the crime rate measure they use is $(C_R + C_V)/(P_R)$. This rate will necessarily be larger than $(C_R + C_V)/(P_R + P_V)$. For this reason, I argued that the Grinols and Mustard study may significantly overstate the casino effect on crime. There are several other problems with the Grinols and Mustard analysis. For example, their empirical work does not enable them to differentiate between tourism-related crime, in general, and casino tourism-related crime, in particular. Furthermore, their variable accounting for casinos is a simple dummy variable; they do not distinguish among different sizes of casino industries or casino volume. So in their model Clark County, Nevada (Las Vegas), would be treated the same as a county in Colorado with a single small casino. Taken together, these issues raise serious questions as to whether the crime effect Grinols and Mustard attribute to casinos can be believed.

A recent study by William Reece (2010) helps to alleviate some of the concerns with the Grinols and Mustard study. Although Reece studied only one state (Indiana), he posited a more careful model. First, he attempted to control for casino volume by including casino turnstile count as an explanatory variable. Reece also included the number of hotel rooms as a variable; this helped to control for tourism in general. Reece's results indicate that increased casino activity reduces crime rates, except for burglary. He specifically indicates that leaving out a measure of casino activity creates a "serious specification error." Finally, the results indicate that the building of hotel rooms subsequent to casinos opening tends to reduce crime. Overall, Reece's study provides important evidence that counters the claims by Grinols and Mustard (2006) that casinos attract crime.

There is clearly good reason to expect some relationship between casinos and criminal activity. However, based on an overview of the literature, there is no conclusive evidence on the relationship between casinos and crime. More careful econometric analyses are needed.

5 Measurement Problems in Cost-Benefit Analysis

It is doubtful that researchers will adopt a pure economic conception of social cost such as that described by Walker and Barnett (1999).[13] Even if the proper definition of social cost were clear, there are other obstacles to actually measuring the social costs of gambling. Of course, the inability to measure costs does not mean that the costs do not exist. Rather, it simply means that researchers and policy makers must be careful in interpreting social cost studies.

Perhaps the most serious obstacle in performing valid social cost estimates is the issue of comorbidity. That is, pathological gamblers may have other problems that contribute to their socially costly behavior, meaning that the costly behavior results from multiple disorders rather than one. One study found that almost 75 percent of pathological gamblers also have alcohol use disorders, and almost 40 percent have drug problems (Petry, Stinson, and Grant 2005). Consider a problem gambler who is also an alcoholic. Suppose his behavior results in social costs of $1,000. Most gambling researchers, including Thompson, Gazel, and Rickman (1997), Thompson and Schwer (2005), and Grinols (2004) would simply attribute the entire $1,000 cost to gambling, even though the drinking may be responsible for some (or even most) of the total cost. How should comorbidity be handled, in terms of estimating the costs due to a particular affliction? This question has not been addressed by researchers, but it is critical to creating valid social cost estimates.

A second issue is the *counterfactual scenario*. What if casinos were not legal? Would pathological gambling and the associated social costs disappear? Probably not. A valid estimate of the costs of pathological gambling, as it relates to government policy, is not the total cost of pathological gambling behaviors. Rather, the relevant cost is the difference between the costs when casinos are legal and when they are not. Unfortunately, it is very difficult to know with accuracy the counterfactual scenario. Since most social cost estimates do not consider this, they must be viewed with skepticism.

A third problem with estimating the social costs of gambling is that many of the published estimates have been based on unreliable survey data. In some studies authors have based their cost estimates on diagnostic tools like the DSM-IV or SOGS.[14] Some papers use original surveys in which problem gamblers are asked about the extent of their gambling losses or the sources of their money used for gambling. The study by Alex Blaszczynski et al. (2006) found that many survey respondents are unable to estimate their gambling losses, even if given instructions on how to do so. This evidence suggests that it would be difficult for the same individuals to reliably report the source of their gambling losses. This is because budgets are fungible (Walker 2007a, 121). Yes, a person may gamble too much. But he or she also may have a very high car payment. How confident are we that this person (or the researcher for that matter) could accurately identify what source of income—paycheck, bank loan, cash gifts, theft, and so on—was used to finance a specific expenditure? Those taking a survey on problem gambling may be predisposed to blame all of their problems on gambling even though they have other problems as well. This suggests that estimated costs for bad debts, theft, and the like are likely invalid.

The fourth and final problem discussed here relates to how government expenditures are handled. A large portion of the social costs of gambling may be related to government expenditures. For example, suppose government-provided treatment is available and many pathological gamblers commit crimes that create legal costs. Most social cost estimates simply take the value of these government expenditures and call them social costs. It seems obvious to some analysts that, since government spending requires taxes, these expenditures should be considered social costs. Indeed, most people would agree that lower spending on these sorts of things would be preferred to higher spending. But the same is not necessarily true of, say, education. People often vote for more public education spending. The point is that government expenditures are *not* equivalent to social costs. If they were, then we could reduce the social costs of gambling by simply reducing spending on gambling-related problems! Unfortunately, this does not leave us with a clear and appropriate way to classify gambling-related government expenditures.

Yes, such expenditures may be a reflection of social costs, but they may also represent the social costs of *policy decisions*. This issue has been addressed by Edgar Browning (1999). As with the previous issues discussed here, there is no ideal way to deal with this one.

Finally, since the above problems (among others) make it very difficult to obtain credible data on the social costs of gambling, many researchers rely on a variety of wildly arbitrary assumptions in performing their analyses. The result is sometimes nothing but meaningless cost estimates.

There are other problems in the gambling literature that make cost-benefit analyses unreliable. This section has focused only on cost-side issues, but there are benefit-side problems as well. Until researchers can adequately deal with some of these problems, policy makers and voters must be cautious in how they interpret and use the cost-benefit analyses of gambling studies. In many ways the problem gambling literature parallels the substance abuse literature, both reflecting a "cost-of-illness" approach. That work provides a possible path for gambling researchers to follow. But even the better established substance abuse literature has its critics (e.g., Reuter 1999 and Kleiman 1999).

6 Conclusion

This chapter summarizes my previous research on the economic and social impacts of casino gambling in the United States. Empirical studies have demonstrated a short-term positive economic growth effect from casinos, at least those using state-level data. Subsequent studies have confirmed that, in the case of recovery from Hurricane Katrina, casinos seem to have made significant positive contributions to state-level personal income. This is likely due to straightforward economic development effects resulting from capital investment and the employment provided by casinos.

What is less clear is whether casinos make a net positive contribution to state-level revenues. Our empirical analysis indicates that casinos actually detract from state government revenues, perhaps due to a large substitution away from other types of spending. Evidence on lotteries appears to be consistent with the common sense notion that lotteries enhance state revenue. These contradictory results suggest that legalized gambling is not always a positive contributor to state government coffers. It likely depends on what types of gambling the jurisdiction has as well as the specifics of the tax policies.

One of the primary arguments used to oppose the introduction of casinos is that they may be the catalyst for significant social costs. Most of these costs are attributable to such pathological gambling behaviors as criminal activities and to requiring treatment. There has been much debate over the magnitude of such costs, and I have argued that most of the empirical estimates of the social costs are largely arbitrary. When we focus on the specific issue of crime, we find that many of the studies from the literature may have overstated the crime effect from casinos. This is because studies often use a flawed measure of the crime rate in their analyses. The best evidence suggests an unclear crime effect from casinos. Although there certainly are social costs attributable to pathological gambling, most published estimates are wildly inaccurate. Finally, I have identified some fundamental problems with the nature of social costs and data availability that make it unlikely that significant improvement in such estimates will happen anytime soon.

The economics of gambling is a fascinating area of research. It is a young field with countless potential research topics. More developed casino markets, such as Australia, Canada, Macau (China), the United Kingdom and the United States, have been the recipients of most of the research attention. But all casino jurisdictions deserve more attention from economists. It is my hope that economic research on casino gambling will continue to expand as data availability improves, covering more markets around the world. Only with more attention from researchers can we develop a better understanding of the actual economic and social effects of casino gambling.

Notes

This chapter is based on a paper presented at the Macao Polytechnic Institute Global Gaming Management Seminar Series, Oct. 23, 2009. The author gratefully acknowledges financial support from the Macao Polytechnic Institute; the U.S. Department of Education, Title VI-A, International Studies and Foreign Language Program; and the Global Trade Resource Center at the College of Charleston.

1. These studies include Anderson (2005), Anders, Siegel, and Yacoub (1998), Elliott and Navin (2002), Fink, Marco, and Rork (2004), Kearney (2005), Mobilia (1992), Popp and Stehwien (2002), Ray (2001), Siegel and Anders (1999, 2001), and Thalheimer and Ali (1995).
2. Unfortunately, volume is not measured or reported the same in each industry. Volume for lotteries is total ticket sales; for commercial casinos it is net revenue; for horse and greyhound racing it is handle (the total dollar amount of bets placed). Because Indian casinos are not required to publicly disclose casino data, there is no obvious or easy way to measure their volume. Instead, we collected data on the total square footage of Indian casino floor space for each state. The casino industry uses a basic formula for casino layout, and revenues would be expected to vary directly with square footage. Obviously this is not a perfect measure, but it perhaps the best proxy for Indian casino volume that has been developed in the literature.
3. State sales taxes are obviously more complicated than this statement implies. The complexities of state sales taxes are not important for the argument being made here, so they are ignored.
4. Casino taxes are complex in some states, with graduated marginal tax rates, various fees, and so on.
5. See (Walker and Jackson 2011) for an explanation of how we estimated handle from the revenue data, in light of the problem discussed above in note 2.

6. It should be noted that we are not confident in the validity of the large negative marginal impact from greyhound racing. Nor can we explain it. Yet we get similarly large and negative results whatever model specification is used.
7. For examples of studies that examine particular states' or localities' tax impacts from legalized gambling, see Anders, Siegel, and Yacoub (1998), Borg, Mason, and Shapiro (1993), Fink, Marco, and Rork (2004), Siegel and Anders (1999), and Thalheimer and Ali (1995).
8. There are different degrees of pathological gambling and corresponding different terms. We ignore these details here. It is not clear the extent to which legalizing gambling causes an increase in the prevalence of pathological gambling. However, it seems reasonable that we would see an increased prevalence closer to casinos, for example. This does not necessarily imply, however, that eliminating casinos or other forms of gambling would decrease the prevalence rate markedly, since individuals could still gamble illegally or travel to nearby legal establishments.
9. For this discussion we will ignore the potentially serious problems of using survey data from Gamblers Anonymous members to generate an estimate of social costs for the representative pathological gambler.
10. This is not a generally a controversial statement for economists. For example, the Federal Reserve Bank of Minneapolis's gambling issue of *Fedgazette* (2003) cited the Walker and Barnett paper in discussing transfers.
11. This estimate excludes other legitimate social costs that have been ignored in the literature. For example, lobbying on the part of the casino industry and casino opponents could be classified as social costs related to gambling.
12. Relatively good crime-related studies include Albanese (1985), Curran and Scarpitti (1991), and Stitt, Nichols, and Giacopassi (2003). For a review of the casinos and crime literature, see Walker (2010).
13. This section is adapted from Walker (2007b, 2007c).
14. The DSM-IV is American Psychiatric Association's *Diagnostic and Statistical Manual of Mental Disorders* 4e. The SOGS is the South Oaks Gambling Screen developed by H. R. Lesieur and S. B. Blume (1987). Both instruments indicate likely pathological gamblers based on their responses to a variety of questions about their gambling-related behavior.

References

Albanese, Jay S. 1985. The effect of casino gambling on crime. *Federal Probation* 48:39–44.

Anders, Gary C., Donald Siegel, and Munther Yacoub. 1998. Does Indian casino gambling reduce state revenues? Evidence from Arizona. *Contemporary Economic Policy* 16:347–355.

Anderson, John E. 2005. Casino taxation in the United States. *National Tax Journal* 58: 303–324.

Arthur Andersen. 1996. *Economic impacts of casino gambling in the United States. Vol. 1: Macro study*. Study conducted for the American Gaming Association. Las Vegas, Nev.: Arthur Andersen.

Australian Government Productivity Commission. 1999. *Australia's gambling industries*. 3 vols. Canberra: AusInfo; http://www.pc.gov.au/projects/inquiry/gambling/docs/report.

Blaszczynski, Alex, Robert Ladouceur, Annie Goulet, and Cathy Savard. 2006. "How much do you spend gambling?" Ambiguities in questionnaire items assessing expenditure. *International Gambling Studies* 6(2):123–128.

Borg, Mary O., Paul M. Mason, and Stephen L. Shapiro. 1993. The cross effect of lottery taxes on alternative tax revenue. *Public Finance Review* 21(2):123–140.

Browning, Edgar K. 1999. The myth of fiscal externalities. *Public Finance Review* 27(1):3–18.

Calcagno, Peter T., Douglas M. Walker, D., and John D. Jackson. 2010. Determinants of the probability and timing of commercial casino legalization in the United States. *Public Choice* 142(1):69–90.

Curran, Daniel, and Frank Scarpitti. 1991. Crime in Atlantic City: Do casinos make a difference? *Deviant Behavior* 12(4):431–449.

Elliott, Donald S., and John C. Navin. 2002. Has riverboat gambling reduced state lottery revenue? *Public Finance Review* 30(3):235–247.

Federal Reserve Bank of Minneapolis. 2003. Gambling: A sure thing? *Fedgazette*, March; http://www.minneapolisfed.org/publications_papers/issue.cfm?id= 45.

Fink, Stephen, Alan C. Marco, and Jonathan Rork. 2004. Lotto nothing? The budgetary impact of state lotteries. *Applied Economics* 36(1):2357–2367.

Grinols, Earl L. 2004. *Gambling in America: Costs and benefits.* New York: Cambridge University Press.

Grinols, Earl L., and David B. Mustard. 2006. Casinos, crime, and community costs. *Review of Economics and Statistics* 88(1):28–45.

Heckman, James J. 1979. Sample selection bias as a specification error. *Econometrica* 47(1):153–161.

Jackson, John D, David S. Saurman, and William F. Shughart II. 1994. Instant winners: Legal change in transition and the diffusion of state lotteries. *Public Choice* 80:245–263.

Kearney, Melissa Schettini. 2005. State lotteries and consumer behavior. *Journal of Public Economics* 89(11–12):2269–2299.

Kleiman, M. A., 1999. "Economic cost" measurements, damage minimization and drug abuse control policy. *Addiction* 94(5):638–641.

Lesieur, H. R., and S. B. Blume 1987. The South Oaks Gambling Screen (SOGS): A new instrument for the identification of pathological gamblers. *American Journal of Psychiatry* 144(9):1184–1188.

Mobilia, Pamela. 1992. Trends in gambling: The pari-mutuel racing industry and effect of state lotteries, a new market definition. *Journal of Cultural Economics* 16(2):51–62.

National Gambling Impact Study Commission (NGISC). 1999. *National Gambling Impact Study Commission final report.* Washington, D.C. NGISC; http://govinfo.library.unt.edu/ngisc.

National Research Council. 1999. *Pathological gambling: A critical review*. Washington, D.C.: National Academies Press.

Petry, N. M., F. S. Stinson, and B. F. Grant. 2005. Comorbidity of DSM-IV pathological gambling and other psychiatric disorders: Results from the National Epidemiologic Survey on Alcohol and Related Conditions. *Journal of Clinical Psychiatry* 66(5):564–574.

Popp, Anthony V., and Charles Stehwien 2002. Indian casino gambling and state revenue: Some further evidence. *Public Finance Review* 30(4):320–330.

Pumper, Dror. 2011. A complete ban of all casino opportunities in Kyrgyzstan. Casino.com, Nov. 14; http://www.casino.com/news/online-casino-news/a-complete-ban-of-all-casino-opportunities-in-kyrgyzstan-00356720.html.

Ray, Margaret A. 2001. How much on that doggie at the window? An analysis of the decline in greyhound racing handle. *Review of Regional Studies* 31(2):165–176.

Reece, William. 2010. Casinos, hotels, and crime. *Contemporary Economic Policy* 28(2): 145–161.

Reuter, P. (1999). Are calculations of the economic costs of drug abuse either possible or useful? *Addiction* 94(5):635–638.

Schumpeter, Joseph A. [1934] 1993. *The theory of economic development.* Reprint, New Brunswick, N.J.: Transaction.

Siegel, Donald, and Gary Anders. 1999. Public policy and the displacement effects of casinos: A case study of riverboat gambling in Missouri. *Journal of Gambling Studies* 15(2):105–121.

———. 2001. The impact of Indian casinos on state lotteries: A case study of Arizona. *Public Finance Review* 29(2):139–147.

Stitt, B. Grant, Mark Nichols, and David Giacopassi. 2003. Does the presence of casinos increase crime? An examination of casino and control communities. *Crime & Delinquency* 49(2):253–284.

Thalheimer, Richard, and Mukhtar M. Ali. 1995. The demand for parimutuel horse race wagering and attendance. *Management Science* 41(1):129–143.

Thompson, William N., and R. Keith Schwer. 2005. Beyond the limits of recreation: Social costs of gambling in southern Nevada. *Journal of Public Budgeting, Accounting & Financial Management* 17(1):62–93.

Thompson, William N., Ricardo Gazel, and Dan Rickman. 1997. Social and legal costs of compulsive gambling. *Gaming Law Review* 1(1):81–89.

Tullock, Gordon. 1967. The welfare costs of tariffs, monopolies, and theft. *Western Economic Journal* 5(3):224–232.

Walker, Douglas M. 2003. Methodological issues in the social cost of gambling studies. *Journal of Gambling Studies* 19(2):149–184.

———. 2007a. *The economics of casino gambling.* New York: Springer.

———. 2007b. Problems in quantifying the social costs and benefits of gambling. *American Journal of Economics and Sociology*, *66*, 609–645.

———. 2007c. Hurdles in gambling research. *The WAGER*, Feb. 7; http://www.basisonline.org/2007/02/op-ededitorials.html.

———. 2008a. Clarification of the social costs of gambling. *Journal of Public Budgeting, Accounting, & Financial Management* 20(2):141–152.

———. 2008b. Evaluating crime attributable to casinos in the U.S.: A closer look at Grinols and Mustard's "Casinos, crime, and community costs." *Journal of Gambling Business and Economics* 2(3):23–52.

———. 2010. Casinos and crime in the USA. In *Handbook on the economics of crime*, edited by Bruce L. Benson and Paul R. Zimmerman. Northampton, Mass.: Edward Elgar, 488–517.

Walker, Douglas M., and A. H. Barnett. 1999. The social costs of gambling: An economic perspective. *Journal of Gambling Studies* 15(3):181–212.

Walker, Douglas M., and John D. Jackson. 1998. New goods and economic growth: Evidence from legalized gambling. *Review of Regional Studies* 28:47–69.

———. 2007. Do casinos cause economic growth? *American Journal of Economics and Sociology* 66(3):593–607.

———. 2008a. Market-based "disaster relief": Katrina and the casino industry. *International Journal of Social Economics* 35(7):521–530.

——. 2008b. Do U.S. gambling industries cannibalize each other? *Public Finance Review* 36(3):308–333.

——. 2009. Katrina and the Gulf states casino industry. *Journal of Business Valuation and Economic Loss Analysis* 4(2): article 9.

——. 2011. The effect of legalized gambling on state government revenue. *Contemporary Economic Policy* 29(1):101–114.

SECTION II

SPORTS BETTING

This page intentionally left blank

CHAPTER 7

THE ECONOMICS OF ONLINE SPORTS BETTING

GEORGE DIEMER AND RYAN M. RODENBERG

1 Introduction

The technological shock ushered in by the Internet has brought about significant structural changes in many industries. The sportsbook industry is one such example. Michael Smith and Leighton Vaughan Williams (2008, 404) have concluded that it is "now widely accepted that the technologies associated with the internet have spawned radical new business models and formats [and a]spects of the gambling industry are being transformed by these pervasive technologies." More pointedly, Mark Davies et al. (2005) have explained how five distinct technological forces have revolutionized sports wagering—(1) Moore's Law pertaining to computing power, (2) Metcalfe's Law regarding networks, (3) decreased transaction costs as explained by the Coase Theorem, (4 increased potential for customer clustering (also known as the "flock of birds phenomenon"), and (5) diminished barriers to entry via the so-called "fish tank phenomenon." The technology-induced boom in Internet sports gambling has, in turn, rendered the somewhat antiquated laws "virtually unenforceable" (Cabot and Faiss 2002, 2).

The pre-Internet economic and legal environment of sports wagering was relatively clear, as the industry was heavily concentrated. In this previous era, with full-scale wagering on sports illegal in 49 out of 50 states in the United States,[1] American gamblers had only two options: (1) they could travel to Nevada and gamble on sports legally (at what are known as "terrestrial" gambling sites) or (2) they could place illegal wagers with their local bookie. In the post-Internet age, however, the legal and economic issues involved in sports gambling are much more complex and uncertain. The sports gambling sector has changed dramatically from an oligopolistic to a monopolistically competitive industry unhampered by geography. The geographical sportsbook market was local in nature. The Internet changed that market to a global one. Although there have been sporadic legislative attempts to regulate and/or prosecute the industry in the United States, the legality of the sports gambling industry is currently in a state of flux.

The purpose of this chapter is to provide an overview of the online sports betting industry from an economic and a legal perspective. Our dual focus is one of necessity, as the overlapping, and sometimes conflicting, regulatory environment lends itself to a decidedly "law and economics" inquiry. After an overview of the pre-Internet environment, we proceed to a multifaceted treatment of the online sportsbook, explaining the decided shift in its economic characteristics and the ambiguous legal status in which the industry resides. For the avoidance of doubt, our economic-based coverage is applicable globally. In contrast, the accompanying legal discussion will focus primarily on jurisprudence in the United States, a nation that holds the somewhat ironic distinction of being the world's largest single-nation sports gambling market and the only major industrial jurisdiction whose laws leave bettors and gambling-related businesses in a legal grey area regarding the permissibility of sports wagering activities.

II Pre-Internet Legal and Economic Environment

The pre-Internet legal and economic environment for sports gambling was highly regulated and heavily concentrated among a few key businesses and locations. It was dominated by terrestrial gambling sites and illegal sportsbooks, both of which operated within relatively stable geographical and legal constraints. While the operations of the illegal sportsbooks are largely clandestine with a few exceptions (Strumpf 2003), the licensed terrestrial gambling sites "present significant opportunities for economic analysis" (Sauer 1998, 2021).

A Terrestrial Sports Gambling Sites

Terrestrial gambling sites, such as the legal sportsbook casinos in Las Vegas and other locations in Nevada, operate in an oligopolistic industry shaped by large barriers to entry. Significant start-up costs, regulatory licenses, and a heavy tax rate serve as barriers to entry into the industry and result in a limited number of competitors in the highly regulated jurisdiction. This may be changing, however. First, Cantor Fitzgerald, a well-known Wall Street firm, has recently made inroads into the Vegas-based sportsbook industry through its subsidiary affiliate Cantor Gaming, operating the upscale M Resort and licensing its real-time "Inside Wagers" technology to a handful of other sportsbooks around Las Vegas. Second, William Hill, a prominent British bookmaking firm, announced the 2011 acquisitions of Nevada-based sportsbooks American Wagering, Club Cal Neva Satellite Race and Sportsbook Division, and Brandywine Bookmaking LLC.

As is the case with most oligopolistic industries, these firms can earn an excessive economic profit. Terrestrial sports gambling sites use a traditional fee, known as vigorish, of 10 percent on all losses. This fee is often thought of as a commission. Traditional literature assumes that the objective of a sportsbook is to balance the action or amount wagered on each side of the game or event. If this objective is attained, the sportsbook takes no risk on the game, as sportsbooks are neutral regarding the actual outcome of the game or event, collect all losing bets (plus the 10 percent vigorish), and pay out the winning bets. This leaves the sportsbook with commission revenue of 5 percent of all action.[2] Terrestrial sportsbooks are often content with this commission, especially when it brings in additional action to the table games on-site. In addition to traditional wagers, most sportsbooks offer a limited number of so-called exotic wagers similar in form to the pari-mutuel system used by a racetrack to handle its gambling. These systems take in 14 to 20 percent before distributing the winnings (Davies et al. 2005).

B Local Bookies

Prior to the 1840s, most gambling in the United States was heavily rigged and consisted of poker games that took place on steamboats. Gambling was disorganized and decentralized, but this all changed with the construction of major horseracing tracks, such as Saratoga in 1864 and Churchill Downs in 1875. In the 1870s racetracks began introducing Paris Mutual (or pari-mutuel) betting. Pari-mutuel wagering is a type of bet in which the payout odds are tabulated after the race has completed. These odds are a function of where the money was wagered and in what amounts. Updated odds give the gamblers an idea of the expected payout. This type of wager took time to gain popularity but remains the dominant type of racetrack wager today.

Peter Reuter (1983) estimated that the illegal sportsbook industry in New York City's four-firm concentration ratio was no more than 35 percent. He concluded that bookmaker operations were relatively small, with frequent entry barriers but almost no exit barriers. Reuter also found that efforts toward collusion often failed and that the market seemed to operate competitively. However, one signal of monopoly power in the industry still existed—third-degree price discrimination. Koleman Strumpf (2003) found evidence of such third-degree price discrimination within the illegal sportsbook industry. Using a unique data set provided by a county district attorney following several high-profile arrests, Strumpf reached this conclusion by examining six bookmakers who operated in and around the New York City area in the 1990s. Strumpf described the client differentiation that is required in such activity as being relatively easy. Simply by moving the point spread of the local team to make it more expensive to bet on them, the local sportsbooks discriminate against the local geographical bias of the fan base in that area.

C Sportsbook Corruption Pre-Internet

The industry of the racetrack bookmaker relied on an information network just at a time when communication technology was evolving. Bettors with results of prior races that were allowed to wager on off-track locations opened the door for "past-posts." Exploiting information gaps became a lucrative business for professional gamblers. Filling any existing information gaps became a great concern for bookmakers as they used the services of Western Union for up-to-date information. Federal authorities opted against cracking down on legal or illegal gamblers at these early racetracks, largely so that they could go after affiliated services provided by legally operating businesses like Western Union. By pressuring companies that contributed to illegal gambling, the government sought to indirectly curb the bookmaking business, a trend that continues today.[3] Finally, in 1904 Western Union announced it would end the collection and dissemination of race results. This move resulted in large profits for illegal wire transfers companies that practiced price discrimination in connection with the information.[4]

From 1950 to 1951, the U.S. Senate, the nation's most powerful legislative body, conducted what was officially called the Senate Special Committee to Investigate Crime in Interstate Commerce and commonly referred to as the Kefauver Committee (U.S. Congress 1951). The primary objective of this committee was to investigate organized crime. The investigation led to the conclusion that illicit bookmaking resulted in illegal monopolization of racing information and, in turn, police corruption. This finding put a spotlight on the link between organized crime and bookmaking and served as the impetus for the Wire Act of 1961 (Wire Act), one of the most prominent federal anti-gambling statutes.

The National Gambling Impact Study Commission (NGISC) was authorized by Congress in 1996 and completed its work in 1999 (NGISC 1999). Except for tribal and Internet gambling, the NGISC recommended that gambling be given statewide autonomy. Furthermore, the NGISC advised curtailing statewide lotteries, especially to the extent that they target lower income neighborhoods. The NGISC recommended that collegiate and other amateur sports gambling be reduced or eliminated, a recommendation that was not enacted as legislation at the federal level.[5]

The illegal bookmaking industry changed in many ways between the Kefauver Committee and the time the NGISC was commissioned. What was once dominated by horse races slowly began moving toward what is now known as traditional sports betting. Information gaps were no longer problematic for bookmakers, as transmitting information about the games by broadcasters was acceptable. In addition, with the popularity of the point spread system, taking bets became simpler with fewer events and only two possible outcomes to wager on. Finally, terminology began to change with legitimate bookmaking rings being known as sports books.

III Post-Internet Legal and Economic Environment

Defining the geographical and legal markets for Internet sportsbooks is a complex task. Since its enactment, the Wire Act has prohibited gamblers and bookies from making transactions across state lines over telephone wires.[6] As such, to steer clear of the statute, bettors and bookies had to be in the same place at the same time. Today, however, there are up to three locations for gambling activity: (1) a domain name registered location, (2) a place of business operation, and (3) a target market where individual gamblers reside. One marketplace often covers multiple geographical locations. As Mark Wilson (2003, 1246) explained: "An example of online gambling is Casino Australia, which has a name suggesting Australia, a domain name registered in New Hampshire in the United States, and a website based in the Netherlands Antilles."

The example of Casino Australia is illustrative of how uncertain the legal environment has become, notwithstanding the fact that a number of Internet sports books are publicly traded in well-respected stock exchanges, such as the London Stock Exchange. In these cases, it is hard to initially identify which nation is responsible for enforcing which nation's gambling laws. Given the reactive nature of lawmaking, statutes and code are just now being codified with Internet-based transactions in mind. For example, on August 7, 2007, after several court rulings had chipped away at the law, the U.S. Congress amended the Foreign Intelligence Surveillance Act of 1978 (FISA) to better fight terrorism. This update allows American investigative bodies to use modern technology in a more expansive way.[7] It is conceivable that a similar update to the Wire Act could be forthcoming in the not too distant future.[8] Subsequent interpretations of FISA, the Wire Act, and similarly situated laws enacted before the emergence of the Internet will shape the industry moving forward.

In their examination of international legal regulations in the Internet age, Stephan Wilske and Teresa Schiller (1997) set forth several legal principles for regulating Internet activity within and across national borders. The first and most relevant to Internet gambling is the territoriality principle, which states that a nation has authority over locally produced Web content. This effectively grants a nation the power to have laws restricting Internet businesses from offering services to its citizens, even if the business operates within the nation's borders and provides the services of foreigners legally. A business can comply with such laws by identifying the geographic location of potential customers through their Internet addresses. This practice is common in the Internet sports industry. For example, Australia allows local online gambling sites to operate provided no Australian citizens are allowed to gamble on their sites. In the United States, the territoriality principle is unevenly applied. As a result, many Internet sportsbooks have decided to restrict access to American customers.[9]

A considerable portion of the aforementioned NGISC centered on the then-novel type of gambling that appeared on the landscape—Internet gambling. In its recommendations, the NGISC report concluded:

> The Commission recommends to the President, Congress, and the Department of Justice (DOJ) that the federal government should prohibit, without allowing new exemptions or the expansion of existing federal exemptions to other jurisdictions, internet gambling not already authorized within the United States or among parties in the United States and any foreign jurisdiction. Further, the Commission recommends that the President and Congress direct the DOJ to develop enforcement strategies that include, but are not limited to, internet service providers, credit card providers,[10] money transfer agencies, makers of wireless communications systems, and others who intentionally or unintentionally facilitate internet gambling transactions. Because it crosses state lines, it is difficult for states to adequately monitor and regulate such gambling.[11]

Related NGISC findings included a call for passage of legislation that would curb or prohibit wire transfers involving gambling websites and a recommendation to the president and Congress to facilitate cooperation with foreign governments.

As for the economic environment, the industry appears to be in transition. The post-Internet era of sportsbooks is one that is not oligopolistic or localized. Instead, it is highly monopolistically competitive with a global marketplace. Complicating matters are two non-American judicial proceedings with far-reaching free trade and consumer protection implications—the European Court of Justice's *Liga Portuguesa* decision in 2009 and the 2003 World Trade Organization's *Antigua and Barbuda v. United States* case.[12] The multi-jurisdictional nature of these disputes, coupled with the now monopolistically competitive nature of the industry, suggest an uncertain and likely litigious future.

A Entry Barriers

In the post-Internet era, there are few significant entry barriers in this industry. Starting up a no-frills Internet sportsbook costs approximately ten thousand U.S. dollars, which is used to handle such tasks as acquiring domain names, obtaining the use of servers and phone lines, purchasing or licensing gambling software, and advertising. By operating in Internet-friendly nations (such as Antigua or Costa Rica), these new Internet sportsbooks conduct business with less regulation, lower tax rates, fewer licensing requirements, and little to no threat of legal prosecution vis-à-vis online sportsbooks in jurisdictions with stricter laws and policies.

B Market Shares

The Internet sportsbook industry is currently in a state of flux for a number of reasons. As the market for online sportsbooks matured, the number of virtual sports gambling portals increased rapidly and subsequently stabilized. Whether this trend will continue is unclear because of the increasing threat of prosecution and the uncertainty with which it is associated. For example, on May 23, 2011, an American federal grand jury returned indictments charging various online sports betting businesses and individuals with illegal gambling and money laundering (Millman 2011). The domain names for 10 online sportsbooks were seized, including the popular website Bookmaker.com. Several prominent online sports betting portals responded to the crackdown by rerouting traffic to domain names with different website suffixes, such as .ag or .eu instead of the ubiquitous .com.

Wilson (2003) cites 1,122 gambling websites in June 2001, though his number likely includes non-sports-related gaming sites as well. In the summer of 2007, SportsbookReview.com (SBR) graded 755 sportsbooks as acceptable and 886 sportsbooks as blacklisted, which in this context refers to a sportsbook that has liquidity problems and/or may be fraudulent. In June 2011, according to SBR, the number of blacklisted sites increased to 1,052 (many presumably have since closed down) and 579 other sites were currently in operation. Accurate revenue figures are impossible to establish in this industry due to existing regulations (or lack thereof) and an absence of incentive to disclose financial figures. Estimates by Christiansen Capital Advisors (CCA) indicate enormous growth in the industry.[13] Revenues increased from USD3 billion in 2001 to an estimated USD18 billion in 2007. These figures include non-sports-related gambling but nonetheless provide some insight into the growth of the industry. Time will tell how new legislation affects the industry's expansion. In the post-Internet era, sportsbooks operate in complete autonomy compared to U.S. corporations. If nothing else, local land-based bookmakers now face increased competition, though there is some evidence that terrestrial gambling companies support online betting (Richtel 2001).

C Product Differentiation

In the pre-Internet sportsbook market, a bookie would only collect money after the sporting event was over. Thus the burden of establishing credibility lay solely on the bookie, who had to decide if a gambler was worthy of credit. Information gathered by the gamblers about the bookies was obtained by word of mouth and only really concerned the bookies' reputations of paying out any winning wagers. In contrast, Internet sportsbooks require fully funded accounts to be set up by gamblers before they place wagers. The burden of establishing credibility has shifted to the gambler, who must establish that the sportsbook site will pay out on wins. In addition to financial credibility, gamblers must be concerned also with the stability of the

Internet sportsbook as legislative bodies attempt to regulate the industry. Information gathered by the gamblers can take many forms, including rating websites, such as SBR, that the credit worthiness and customer service track record of each Internet sportsbook.

Classic signs of competition in the marketplace have appeared online, for example, the introduction of new forums for gambling. The traditional system of a bookmaker setting odds that a gambler wagers on has been challenged by online exchanges that act more like stock exchanges than traditional sportsbooks. Instead of the traditional gambling methods, online gambling can now take the form of playing the bookie or investing in a predictive market. Also, one would expect the product's price to decrease as competition is introduced. This reduction in the price of gambling has taken place as online sportsbooks have begun to decrease the vigorish they take in. Aside from the new predictive market forums, traditional sportsbooks also offer more types of wagers. In-game style sports betting (wagering on a game while the event is taking place) has increased in popularity. Research has found that such real-time markets operate just as efficiently as traditional wagers (Debnath et al. 2003).

D Demand Elasticity

Due to the elimination of geographical restrictions, demand is more elastic to price in the post-Internet sports gambling era. Currently, small point-spread fluctuations change bettor tendencies, as bettors can have multiple accounts and move quickly from sportsbook to sportsbook. In addition, if the point spread differential is too great, there exists the opportunity for what is known as arbitrage, in which a bettor wagers on both sides of a game with different point spreads (Griffin 2011). If the game in question results in a score between the two spreads, the bettor wins both wagers. This kind of activity creates an incentive for sportsbooks' point spreads to converge. The end result is multiple firms with prices (or point spreads) very close to each other and the end of third-degree price discrimination, as we would expect in a more competitive economic environment.

E Economies of Scope

Some economies of scope do exist between the Internet sportsbook industry and other Internet gaming industries. An Internet sportsbook will often offer more traditional Internet games, such as blackjack, slots, or poker. Calls to regulate the industry could not be more understandable in this case, as the recreational gambler likely has no idea of the true expected payouts, which in regulated terrestrial gambling casinos are closely monitored (and transparently displayed). Overall the impact of the economies of scope compared to the terrestrial games has greatly decreased.

F Sportsbook Efficiency

In the classic framework of the sportsbook, the objective was simple. By balancing the action on each side of the game the sportsbook is guaranteed no risk and a safe (5% on all action) return. Recent literature has challenged this assumption, a change likely attributable to greater competition in the online sportsbook industry. Steve Levitt (2004) has suggested that oddsmakers adjust the lines to take advantage of their superior forecasting ability, a suggestion supported by Wayne Winston (2009). George Diemer (2009) followed up this research and identified two types of sportsbooks, a profit-maximizing type and an exposure-minimizing type. Rodney Paul and Andrew Weinbach (2011) also found imbalance but attributed it to bettor's bias toward the favorite teams. Deniz Igan, Marcelo Pinheiro, and John Smith (2011) highlighted bookies' recognition of bettor racial biases as influencing unbalanced books and the initial setting of point spreads. No matter the cause, the assumption of the traditional terrestrial sportsbook's balanced action is being reevaluated.

G Online Sports Betting Exchanges as an Anti-Corruption Tool

Stephen Easton and Katherine Uylangco (2010) found a high level of efficiency in the online tennis betting market, making it a useful predictor of match outcomes. Given the betting market's forecasting ability, it could be used as a statistical screening device for nefarious conduct in the market. Similarly, Karen Croxson and J. James Reade (2011) pinpointed prompt and accurate price updates in the online soccer wagering market following any change in scoring, indicating considerable market efficiency. Unexpected deviations from the documented efficiency could trigger a follow-up investigation regarding whether such deviation was benign or indicative of chicanery. Using Betfair-derived data, Alasdair Brown (2012) found trace amounts of inside information in trading during the epic 2008 Wimbledon men's final between Roger Federer and Rafael Nadal. Betting exchanges such as Betfair have also contributed to the detection of corruption in sport (Drape 2008). As quoted by Davies et al. (2005, 539), a Betfair spokesperson said, "We are putting a searchlight on the sport and helping it clean up its act. There is a clear paper trail on our site that doesn't exist in high street [betting] shops. We are entirely transparent. We have no vested interest in the outcome of a horse race" or other sporting event.

In 2007, Betfair detected suspicious betting patterns involving a tennis match between Russian Nikolay Davydenko and Italian Martin Vassallo Argüello and took the then unprecedented step of halting further trading on the match and turning over transaction records to the appropriate tennis authorities. The subsequent investigation cleared both players of wrongdoing but resulted in Ben Gunn and Jeff Rees's (2008) "Environmental Review of Integrity in Professional Tennis," which (1) identified 45 matches with prima facie evidence of fraud that merited further inquiry and (2) led to the establishment in 2009 of the Tennis Integrity Unit (TIU), a joint venture of the International Tennis Federation, the four Grand Slam tournaments, the ATP World Tour, and the Women's Tennis Association, each of the governing bodies in global elite-level tennis.[14] The TIU rendered its first significant enforcement action on May 31, 2011, when the organization placed a lifetime ban on Austrian player Daniel Kolleher for gambling-related match-fixing.

IV The Future of the Internet Sportsbook Industry

Sports gambling is now in its third incarnation. The first generation of sports gambling has existed for centuries. Whether the wagers take place directly between bettors without an intermediary, at a government-regulated brick-and-mortar establishment, or through a neighborhood bookie operating illegally, this first-generation sports betting continues to be popular. The second generation of sports-based gambling was described by Steven Crist (1998) in a *Sports Illustrated* cover story that reported on bettors placing wagers based on fixed odds remotely through the Internet. By offering odds, point spreads, totals, and other wagering opportunities, online (and often offshore) sportsbooks mimicked terrestrial sites in Las Vegas and were more convenient because they do not require the bettor to be on-site to place a wager.[15] During this time, only one American sportsbook operator conducting business offshore was convicted by a jury of violating the Wire Act,[16] and only one individual bettor was charged with placing a sports wager over the Internet.[17] The third generation, sports gambling 3.0, characterized by real-time online betting exchanges with attributes analogous to any modern stock exchange, has been described by Smith and Vaughan Williams (2008, 404–405) as "an interactive web-based platform for placing and laying bets on sporting events" that offers in-running markets and lower commission rates. Betfair in London is the most recognizable (and popular) example of a sports betting exchange (Davies et al. 2005). Recognizing this as a potential growth area, at least one land-based, in-person sportsbook, Cantor Gaming's upscale M Resort in Las Vegas, has begun offering similar options for patrons (Benston 2010). Given this backdrop, a number of noteworthy developments over the course of the past decade have illustrated the uncertainty currently associated with Internet gambling and the online sportsbook industry.

A BetOnSports.com

No case study explains the uncertainly in the industry as well as what transpired in connection with BetOnSports.com. Between 2002 and 2004, the Costa Rica–based company listed on the London Stock Exchange took in almost USD4 billion in wagers, 98 percent of which came from Americans. On July 17, 2006, company CEO David Carruthers was arrested at the airport in Dallas, Texas, en route to Costa Rica. He was charged with racketeering, conspiracy, and fraud connected to an illegal gambling enterprise. In addition, the government charged BetOnSports.com founder Stephen Kaplan with multiple counts of promoting illegal gambling. These charges illustrate how current laws are being enforced in connection with Internet sports gambling businesses and are suggestive of how future cases may be prosecuted.

B Unlawful Internet Gambling Enforcement Act of 2006

The latest attempt by Congress to curtail sports gambling was made in October 2006, when it passed the Unlawful Internet Gambling Enforcement Act (UIGEA) as part of an otherwise unrelated anti-terrorism bill.[18] UIGEA was signed into law in 2006 but did not go into effect until June 1, 2010. The statute's preamble acknowledges: "New mechanisms for enforcing gambling laws on the Internet are necessary because traditional law enforcement mechanisms are often inadequate for enforcing gambling prohibitions or regulations on the Internet, especially where such gambling crosses State or national borders."

UIGEA prohibits payment processing entities, such as banks and credit card companies, from accepting or distributing monies derived from gambling (Ramasastry 2006). Prior to UIGEA, a gambler could fund his or her account by using a credit card. The intent of Congress was to eliminate the ease of financing overseas sports gambling and to cut sportsbooks' liquidity. A number of Internet gambling sites responded by withdrawing from U.S. markets. Certain private websites continue to access the American market, but they face the threat of prosecution. Similarly, bettors with money on deposit risk being unable to redeem their winnings and account funds if the website is shut down and assets are seized.

UIGEA contains a number of important exemptions. Such carve-outs include wildly popular online fantasy sports, government-run lotteries (some of which have intellectual property licensing deals with professional sports teams and leagues), entirely intrastate gambling, certain interstate horse racing, and gambling taking place on Indian reservations. The indirect anti-gambling enforcement mechanism inherent in UIGEA and the numerous exceptions to what is defined as "gambling" therein have resulted in a situation described as "the absence of a comprehensive federal policy toward internet and sports wagering" (Cabot 2010, 271), "the troubled legal status of internet gambling casinos in the United States in the wake of [UIGEA]" (Nelson 2007, 39), and "an exercise in futility" (Kelly 2011, 257). In addition, Jennifer Chiang (2007) highlighted how compliance with federal laws such as UIGEA may still result in violations of state law. Finally, the passage of UIGEA has re-ignited calls for Internet gambling to be regulated at the state, not federal, level (Wajda 2007).

C Galileo Fund

The proliferation of online sportsbooks and the accompanying decreased concentration of market power have resulted in a considerable increase in the number of "outs" available to both casual bettors and professionals who make a living wagering on sporting events. Such professional bettors gamble systematically in a way resembling stock market investors and include individuals with large bankrolls as well as syndicates using sophisticated computer-driven algorithms. Accordingly, it was only mildly surprising when Gibraltar-based Quay Financials and London-headquartered Centaur introduced the Galileo Managed Sports Fund (Galileo), a hedge fund intending to focus on the sports exchange market (Millman 2010; Popper 2010; Wachter 2010). Darren Rovell (2010) reported that Galileo would bet on outcomes in "soccer, tennis, cricket, horse racing and golf, with plans to expand to [American football] and baseball over the next year." The minimum initial investment in Galileo is €100,000. As of June 1, 2011, Galileo did not accept investors from the United States, although fund managers were reportedly seeking the government's stamp of approval.[19] If the Galileo fund is successful (in terms of investment volume and returns), online sportsbooks and exchanges almost certainly will respond.

V Conclusion

The Internet sportsbook industry is in transition. As some American-based sports leagues shift to tepidly supporting legalized gambling (McKelvey 2004a, 2004b; Thomsen 2009), the sometimes polarizing discussion continues regarding the (beneficial and detrimental) impact gambling has on economies and financial systems (Kindt 2010). Future legal developments, in both the United States and elsewhere, will continue to shape the industry (Cabot 2010; Chan 2010). Absent American intervention in the form of legislation and prosecution, the Internet sportsbook industry will probably converge toward a competitive industry with normal economic profits and wagering options available to the consumer. That likelihood notwithstanding, government regulation for U.S.-based bettors and sportsbooks would almost certainly be welcomed by those in the Web-based sports gambling industry, as such oversight would likely increase the legitimacy, transparency, and long-run certainty for the industry globally.

Notes

1. Three other U.S. states (Delaware, Oregon, and Montana) are permitted to offer limited-scope sports wagering under the Professional and Amateur Sports Protection Act of 1992 (PASPA) (Galasso 2010). Although largely outside the scope of this book chapter, it is noteworthy that a lawsuit brought by a New Jersey senator and an Internet gaming trade

association challenging the legality of PASPA was dismissed in March 2011 (Rodenberg and Young 2011).
2. If the sportsbook is successful in balancing the action, 50 percent of the money will be winners and 50 percent will be losers. The 10 percent vigorish on the losses comes to 5 percent of aggregate money wagered on each event.
3. As discussed in detail below, government pressure on legal businesses indirectly affiliated with sports gambling continues today as evidenced by the Unlawful Internet Gambling Enforcement Act of 2006 (UIGEA), which makes it illegal for companies to process payments for gambling companies.
4. For more information, see Peter Reuter's (1983) discussion of John Payne and Mont Tennes.
5. The scope of the NGISC's coverage regarding gambling was extensive, and Internet gambling was but one section.
6. In relevant part, Section 1084(a) of the Wire Act states: "Whoever being engaged in the business of betting or wagering knowingly uses a wire communication facility for the transmission in interstate or foreign commerce of bets or wagers on any sporting event or contest, or for the transmission of a wire communication which entitles the recipient to receive money or credit as a result of bets or wagers, or for information assisting in the placing of bets or wagers, shall be fined under this title or imprisoned not more than two years, or both." David Schwartz (2005) has provided a comprehensive analysis of the Wire Act.
7. For example, prior to being amended, FISA was interpreted to prohibit examination of an e-mail exchange between two foreign individuals even if both were outside American territory. The amendments closed such apparent loopholes.
8. Although none has passed into law, several bills to amend the Wire Act have been considered by Congress over the past 15 years.
9. As Chad Millman (2011) summarizes: "The decision of whether to take a bet from the United States seems to be based as much on a founder's tolerance for risk as it is on existing law. If the boss is willing to taunt the authorities, the billion-dollar American betting market is something they'll openly tap."
10. This recommendation may have led, at least partially, to passage of UIGEA in 2006.
11. The complete NGISC report can be found online at http://govinfo.library.unt.edu/ngisc/reports/ngisc-frr.pdf.
12. Edward Morse (2010) and Luca Rebeggaini (2010) discuss both cases in detail.
13. Full details of CCA's estimates can be found online at http://www.cca-i.com.
14. Shortly after the report by Gunn and Rees (2008), gambling-related suspensions were levied against several professional tennis players from Italy. The Italian players subsequently sued the ATP World Tour challenging their suspensions (Bass and Rodenberg 2011).
15. Steve Budin and Bob Schaller (2007), Sean Patrick Griffin (2011), and Michael Konik (2006) have provided intricate accounts of how online sportsbooks operate.
16. Antigua-based World Sports Exchange founder and president Jay Cohen was convicted by a trial court jury of violating the Wire Act when his company took bets from Americans in New York and elsewhere (United States v. Cohen, 260 F.3d 68 (2nd Cir. 2001); Schwartz 2010). An in-depth discussion of the case is presented by David McGinty (2003).
17. Dave Forster (2003) has explained how Jeffrey Trauman is believed to be the first American individual bettor to be charged with a crime for placing a sports wager over the Internet.

18. I. Nelson Rose and Martin D. Owens (2009) provide comprehensive coverage of UIGEA and related federal anti-gambling statutes.
19. On a related note, Michael Macchiarola (2010) has suggested that federal securities laws could be used to regulate sports betting.

References

Bass, Jordan R., and Ryan M. Rodenberg. 2011. Tennis + gambling = A looming double fault? *Sports Litigation Alert* 8(1).

Benston, Liz. 2010. Technology is king at the M Resort sports book. *Las Vegas Sun*, March 20.

Brown, Alasdair. 2012. Evidence of in-play insider trading on a UK betting exchange. *Applied Economics* 44(9):1169–1175.

Budin, Steve, and Bob Schaller. 2007. *Bets, drugs, and rock & roll: The rise and fall of the world's first offshore sports gambling empire.* New York: Skyhorse.

Cabot, Anthony. 2010. The absence of a comprehensive federal policy toward internet and sports wagering and a proposal for change. *Villanova Sports & Entertainment Law Journal* 17:271–308.

Cabot, Anthony N., and Robert D. Faiss 2002. Sports gambling in the cyberspace era. *Chapman Law Review* 5(1):1–45.

Chan, Sewell. 2010. Congress rethinks its ban on Internet gambling. *New York Times*, July 28.

Chiang, Jennifer W. 2007. Don't bet on it: How complying with federal internet gambling law is not enough. *Shidler Journal of Law, Commerce & Technology* 4(2).

Crist, Steven. 1998. All bets are off. *Sports Illustrated*, Jan. 26.

Croxson, Karen, and J. James Reade. 2011. Information and efficiency: Goal arrival in soccer betting. Discussion Paper 11-01. Birmingham, U.K.: Department of Economics, University of Birmingham.

Davies, Mark, Leyland Pitt, Daniel Shapiro, and Richard Watson. 2005. Betfair.com: Five techonology forces revolutionize worldwide wagering. *European Management Journal* 23(5):533–541.

Debnath, Sandip, David M. Pennock, C. Lee Giles, and Steve Lawrence. 2003. Information incorporation in online in-game sports betting markets. In *Proceedings of the 4th ACM conference on electronic commerce.* New York: ACM, 258–259.

Diemer, George. 2009. Internet sports books operating in an inefficient marketplace. *Journal of Gambling Business and Economics* 3(1):1–22.

Drape, Joe. 2008. Web site puts focus on the fix in sports bets. *New York Times*, May 25.

Easton, Stephen, and Katherine Uylangco. 2010. Forecasting outcomes in tennis matches using within-match betting markets. *International Journal of Forecasting* 26(3):564–575.

Foster, Dave. 2003. Online bettor pleads guilty. Major Wager Forum, Aug. 12.

Galasso, Anthony G., Jr. 2010. Betting against the House (and Senate): The case for legal, state-sponsored sports wagering in a post-PASPA world. *Kentucky Law Journal* 99, 163–183.

Griffin, Sean Patrick. 2011. *Gaming the game: The story behind the NBA betting scandal and the gambler who made it happen.* Fort Lee, N.J.: Barricade Books.

Gunn, Ben, and Jeff Rees. 2008. *Environmental review of integrity in professional tennis.* London: Tennis Integrity Unit; http://www.howtofixasoccergame.com/Report%20on%20possible%20fixing%20in%20Tennis%20for%20the%20ATP.pdf.

Igan, Deniz, Marcelo Pinheiro, and John Smith. 2011. *"White men can't jump," but would you bet on it?* MPRA Paper No. 31469; http://mpra.ub.uni-muenchen.de/31469/1/MPRA_paper_31469.pdf.

Kelly, Joseph M. 2011. Sports wagering prohibition in the United States: An exercise in futility and the best development for organized crime. *Gaming Law Review and Economics* 15(5):257–259.

Kindt, John Warren, ed. 2010. *The gambling threat to economies and financial systems: Internet gambling*. Buffalo, N.Y.: William S. Hein & Co.

Konik, Michael. 2006. *The smart money: How the world's best sports bettors beat the bookies out of millions*. New York: Simon & Schuster.

Levitt, Steve D. 2004. Why are gambling markets organised so differently from financial markets? *Economic Journal* 114(April):223–246.

Macchiarola, Michael C. 2010. Rethinking sports wagering. *Indiana Law Journal Supplement* 85(1):1–14.

McGinty, David B. (2003). The near-regulation of online sports wagering by United States v. Cohen. *Gaming Law Review* 7(3):205–214.

McKelvey, Stephen M. 2004a. U.S. professional sport organization policies shift to embrace legalized gambling entities: A roll of the dice? *Journal of Legal Aspects of Sport* 14(1):23–45.

———. 2004b. The growth in marketing alliances between US professional sport and legalized gambling entities: Are we putting sport consumers at risk? *Sport Management Review*, 7(2):193–210.

Millman, Chad. 2010. Don't trust Wall Street? Try Galileo fund. Chad Millman blog. ESPN.com, April 14.

———. 2011. Despite blow, online betting still strong. Chad Millman blog. ESPN.com, May 25.

Morse, Edward A. 2010. Free trade and consumer protection: Competing interests in the regulation of Internet gambling. *European Journal of Consumer Law* 2010(2):289–304.

National Gambling Impact Study Commission (NGISC). 1999. *National Gambling Impact Study Commission*. 1999. *National Gambling Impact Study Commission final report*. Washington, D.C.: NGISC; http://govinfo.library.unt.edu/ngisc.

Nelson, Benjamin B. 2007. Regulation or prohibition? The troubled legal status of Internet gambling casinos in the United States in the wake of the Unlawful Internet Gambling Enforcement Act of 2006. *Texas Review of Entertainment & Sports Law* 9:39–65.

Paul. Rodney J., and Andrew P. Weinbach. 2011. Investigating allegations of pointshaving in NCAA basketball using actual sportsbook betting percentages. *Journal of Sports Economics* 12(4):432–447.

Popper, Nathaniel. 2010. New hedge funds bets on sports, literally. Los Angeles Times, April 17.

Ramasastry, Anita. 2006. The new federal law banning payments for online gambling: Why it's the wrong choice. FindLaw.com, Oct. 10.

Rebeggiani, Luca. 2009. The Liga Portuguesa decision of the European Court of Justice—An economist's view. *Revista di Diritto ed Economia dello Sport* 5(3):111–122.

Reuter, Peter. 1983. *Disorganized crime: The economics of the visible hand*. Cambridge, Mass.: MIT Press.

Richtel, Matt. 2001. Las Vegas casinos, in shift of position, back online betting. *New York Times*, May 17.

Rodenberg, Ryan M., and Young, S. J. (2011). Ruling impacts New Jersey sports gambling litigation. *Sports Litigation Alert* 8(6).

Rose, I. Nelson, and Martin D. Owens, Jr. 2009. *Internet gaming law*. 2nd ed. New Rochelle, N.Y.: Mary Ann Liebert, Inc.

Rovell, Darren. 2010. Sports betting hedge fund becomes reality. CNBC.com, April 7.

Sauer, Raymond D. 1998. The economics of wagering markets. *Journal of Economic Literature* 36(4):2021–2064.

Schwartz, David G. 2005. *Cutting the wire: Gambling prohibition and the Internet*. Reno: University of Nevada Press.

——. 2010. Not undertaking the almost-impossible task: The 1961 Wire Act's development, initial applications, and ultimate purpose. *Gaming Law Review and Economics*, 14(7): 533–540.

Smith, Michael A., and Leighton Vaughan Williams. 2008. "Betting exchanges: A technological revolution in sports betting." In *Handbook of sports and lottery markets*, edited by Donald B. Hausch and William T. Ziemba. Burlington, Mass.: Elsevier, 403–418.

Strumpf, Koleman S. 2003. *Illegal sports bookmakers*. Working paper; http://www.unc.edu/~cigar/papers/Bookie4b.pdf.

Thomsen, Ian. 2009. Stern open to legalized betting, rule changes. *Sports Illustrated*, Dec. 11.

United States Congress. 1951. *Senate Special Committee to Investigate Organized Crime in Interstate Commerce*. Washington, D.C.: U.S. Government Printing Office.

Wachter, Paul. 2010. A league of their own. *Bloomberg BusinessWeek*, July 8.

Wajda, Nicholas M. 2007. Over-playing a weak hand: Why giving individual states a choice is a better bet for Internet gambling in the United States. *Thomas Jefferson Law Review* 29:313–340.

Wilske, Stephan, and Teresa Schiller. 1997. International jurisdiction in cyberspace: Which states may regulate the Internet? *Federal Communications Law Journal* 50(1):117–178.

Wilson, Mark. 2003. Chips, bits, and the law: An economic geography of Internet gambling. *Environment and Planning A* 35(7):1245–1260.

Winston, Wayne L. 2009. *Mathletics: How gamblers, managers, and sports enthusiasts use mathematics in baseball, basketball, and football*. Princeton, N.J.: Princeton University Press.

CHAPTER 8

THE FOOTBALL POOLS

DAVID FORREST AND LEVI PÉREZ

Introduction

In principle, the term *football pools* could be applied to any pari-mutuel wagering concerning the outcomes, or any other aspects, of football (soccer) matches. However, it has long been used more specifically to refer to long-odds, high-prize gambling games where entitlement to a share of the jackpot is linked to football results. Long odds are inherent in the product and are achieved by requiring players to match their guesses or forecasts with the results of a long list of fixtures. This form of gambling therefore closely resembles lotto, the principal difference being dependence of winning on football results rather than the drawing of numbers by random process.

The similarity of football pools to lotto games has shaped both the history of the pools and academic research on this form of gambling. The history is one of rise and fall, with the decline of the pools from its status as a mass participation activity evidently triggered by the introduction of lotto to the gambling market. Academic research on football pools has been influenced by the literature on lottery demand. e.g., the way in the price variable has been specified, and calculated, to facilitate estimation of own- and cross-price elasticities of demand. In turn, academic research on the pools has turned up findings potentially relevant to modeling lotto sales and to public policy issues relating to lotto games.

Academic work on football pools has focused on two markets, the United Kingdom and Spain. So too does this chapter, reflecting not only that these are the countries where research has taken place but also that they present two particularly interesting cases. The British pools industry is interesting for its history. As recently as the early 1990s it was by far the dominant mode of gambling in the country, attracting a clear majority of the male population and a large proportion of the female as well. By 2010 it had all but faded away, with barely 3 percent of adults participating and turnover likewise reduced (Gambling Commission 2010). The Spanish case is interesting because it is the largest pools market in Europe, accounting for about one-half of sales on the continent as a whole. A point for investigation is why Britain's industry almost disappeared in the age of lotto games whereas that in Spain was able to establish a significant niche in the gambling market despite substantial initial cannibalization when lotto was first introduced.

The Games Explained

Football pools are offered in a variety of jurisdictions outside Europe, for example, in Argentina, China, Singapore, and (with other sports substituted) in parts of the United States and Canada. However, it is a form of gambling most closely associated with Western Europe, where pools usually are marketed by the operator of the state lottery. In Britain, however, it is run by a private sector enterprise (which is also a monopoly but, in this case, a non-statutory one).

The football pools emerged as a popular gambling medium in the years immediately after World War II, when the British firm Littlewoods invented the game known as the treble chance. This was the first (very) long-odds gambling game on football and remains the core product in Britain today. Rules have varied only in detail over the years. From the mid-1990s the number of football matches included on the weekly coupon was 49. A single entry consists of eight selections, though customers are expected to make multiple entries by choosing 10 or 11 matches (with every combination of 8 matches within the set then comprising one play). If a player enters a combination of eight matches where all eight subsequently result in score-draws (any tie except for 0–0, for example, 1–1 or 2–2), he wins a share of the jackpot prize. Those close to having picked eight score-draws are rewarded with a share of a lower tier prize. In the event that there are no winners of the jackpot, there is a rollover, with the prize available on the jackpot added to that for the next edition of the pools. In the event that any of the football matches included in the game do not in fact take place, due to adverse weather for example, the competition still goes ahead with the outcome then related to football results invented by a panel of experts appointed by the operator (Forrest and Simmons 2000, modeled how the panel reaches its forecasts).

The football pools also arrived in Continental Europe in the 1940s, but the core game offered then and now differs from that in Britain. Here, the core toto game has a 1X2 formula whereby a player is required to guess or forecast whether each match in a named set of 12–15 (the number differs by country and has also varied over time) will end in a home win (1), a draw (X), or an away win (2). In countries with relatively low status football leagues, the set of matches may include fixtures from higher profile foreign competitions, such as the premier leagues of England or

Spain, or even from alternative, locally popular sports. The number of matches to be guessed tends to be lower in small market countries, such as Austria and Finland, where otherwise the jackpot would be won too infrequently because the game would be too hard.

Everywhere, a share of the jackpot is claimed on the basis of the player matching all of his or her forecasts with the football results, and there are lower tier prizes for those coming close. Commonly, if a jackpot is not won, it rolls over to the next time. As in Britain, this is an adaptation of the original rules, which had provided for funds from the jackpot pool to cascade down to lower tier prizes when there was no jackpot winner. The change to rollovers postdates the introduction of lotto and made game structure resemble lotto even more closely.

Football Pools in Britain

Pools were offered from the 1920s but in a different form from the current version of this gambling product. During that period, players were asked to forecast results in small groups of matches for small stakes and small prizes. The industry structure was competitive, with the number of firms in the market in the hundreds. In 1946 one of those firms introduced the long-odds treble chance game, and it quickly became popular. The shift of demand to the new long-odds game then both grew the total market and led to substantial market concentration. Roger Munting (1996) noted that there were still 231 companies in 1948 but that the number fell to 42 by 1950 and rapidly thereafter, so that two companies, Littlewoods and Vernons, soon claimed virtually the entire market (Competition Commission of Great Britain 2007). The shift away from competition may be understood as a consequence of the product transforming from essentially a betting product to one akin to a lottery, where the appeal of the size of the prize offset an extremely low probability of winning. Philip Cook and Charles Clotfelter (1993) noted the importance of economies of scale in consumption in lotto markets. Their research showed that the utility derived by a purchaser depends on the number of other entries because it is that which determines the size of the jackpot pool. Consequently, in long-odds competitions, like lotto but also like the pools, there is a tendency to natural monopoly, since whichever firm has the largest, and therefore most exciting, prize will attract more customers from smaller companies, making the games of the latter even less appealing by the next period, and so on, until all competitors collapse. This is consistent with what happened in the pools sector. It is somewhat surprising that only two firms, rather than one, eventually dominated the market. They finally merged in 2007, under the ownership of Sportech plc, their coexistence until then explained by one firm compensating for lower jackpots with a much lower entry fee (Competition Commission of Great Britain 2007).

The Heyday of the Pools in Britain

Such was the rapidity of development of the pools market once it was based on the new long-odds game that it soon became (at least for men) a majority pastime, "more so than elsewhere deeply rooted in the nation's recreational and social fabric" (Miers 1996, 360). Even by the late 1980s, more than one-third of adults were still playing (Sharpe 1997, 360). They were playing for payouts exceeding £2m when there was an outright winner, comparable in real terms to U.K. lotto payouts today. Such life-changing wins attracted considerable popular interest, and a celebrated stage show and subsequent film tracked the "spend, spend, spend" lifestyle of one (female) winner.

David Forrest (1999) has noted that, in this period, the pools served the role of a National Lottery from the perspective of both the government and the public. The government exploited the willingness of players to tolerate high takeout rates in gambling games where a large top prize is available by extracting significant tax revenue. Within two years of the emergence of the treble chance game in the 1940s, it had introduced a new pools tax, set at 10 percent of stakes. This was increased successively to 20, 25, 33, 40 and (in 1982) 42.5 percent. It was still 42.5 percent at the time of the launch of the National Lottery, though with the variation that 5 percent was by then hypothecated to (mainly sports-related) good causes nominated by the pools industry foundation. The proportion of the stake accruing to tax was therefore very close to that applied subsequently to the National Lottery (40–41%). The relative lateness of the United Kingdom in introducing a state lottery was no doubt related to the fact that the country already effectively had a proxy lottery with similar fiscal function.

Certainly the public tended to treat the pools as an alternative form of lottery product rather than as sports betting. Nearly half of all players purchased in advance, for a block of weeks, with the same numbers entered for each of those weeks, such that their entries were made without even knowledge of which football games would correspond to which numbered lines on the entry form. Even those who purchased their entries week by week appear to have chosen numbers rather than matches: the operators themselves suggested that 80–95 percent of all customers chose numbers on the coupon without regard to which teams were covered (Forrest 1999). Thus there may have been conscious selection rather than random play, but it was largely based on selecting numbers of personal significance rather than by assessing teams' prospects as a sports bettor would.

It was by no means unreasonable for pools customers to treat the pools product as if it were a numbers game. The rules require players to guess which matches will be drawn. Perhaps the most detailed statistical forecasting model in all of sport, the football forecasting model outlined in its first version by Stephen Dobson and John Goddard (2001), proves to be effective in picking which team in a match is more likely to win but very ineffective in calling draws, which the authors regarded as near random events. Similarly, fixed-odds bookmakers offer odds for draws that display very low variance across matches. This implies that there is almost no scope for the application of skill in selecting *eight* matches as draws, let alone score-draws. That there was no advantage to being knowledgeable about football will have been helpful in developing the pools product as a mass market activity, but this feature may have been a hindrance later when the pools needed to find a niche in the face of migration of players to lotto.

The Impact of the Lottery on the Football Pools in Britain

The Royal Commission on Gambling (1978) had foreseen such migration and had forecast that any future introduction of a national lottery would damage the pools. But the pace with which they declined was nevertheless striking. The 6/49 lotto game was launched in the course of the 1994–1995 football season, and pools turnover for the whole season was 12 percent lower than in the season before. In 1995–1996 there was a further decline of 28 percent. By 1996–1997, turnover was about £400m, compared with well over £900m in the last full season without lotto, for a cumulative decline of close to 60 percent.

Analysis below will show an immediate effect of similar magnitude in the case of the Spanish pools when that country launched its first lotto game. But, whereas in Spain the pools market then stabilized, in Britain relentless decline continued. By the time of the Competition Commission (1997) inquiry into the pools industry, annual sales were down to £77m. Statistics from regulatory returns (Gambling Commission 2010) indicated turnover of only £52m in 2009–2010, two years after the merger of Littlewoods and Vernons. Thus, even without adjustment for inflation, 95 percent of the market had been lost over the period since the introduction of a national lottery. The pools had become marginal to the gambling industry. For example, the regulatory statistics indicated annual turnover in football bookmaker betting of over £1b and in total non-remote betting of more than £9b. The adult four-week participation rate for the pools had fallen to 3.1 percent (Gambling Commission 2010).

Substantial cannibalization of the pools by the lottery was, with hindsight, inevitable since demand for the pools product likely largely represented suppressed demand for a lottery. Promoters of an actual lottery have advantages over the pools. First, their product is more consistent in that it can be offered year-round and more than once a week if they deem that likely to be profitable. Second, both lotto and pools games can produce large variations in the number of winners from week to week, with disappointing individual payouts if popular numbers are drawn in the case of lotto or if draws occur in a large number of matches in the case of the pools; but the problem is easier to control in the case of a pure numbers game, and the risk of multiple winners each receiving very little may be reduced by choice of an appropriate game matrix. Third, the lottery can attract additional players, particularly among women, whose aversion to football may deter them from playing the pools. These inherent advantages made it likely that the new lottery would gain a larger market than the pools, threatening the latter's survival if initially loyal customers subsequently switched their expenditures because the lottery had emerged, as it did immediately and by a substantial margin, as the medium with the larger jackpot. However, the pools industry also displayed weaknesses of its own making that made it still more vulnerable to this possibility of continuing decline. A fundamental problem to retaining customers was that it offered poor value for money, with a takeout rate of 70 percent compared with 50 percent for the lottery. This gap was so wide, despite similarity in tax rates, because operating costs in the pools accounted for 30 percent (rather than 10%, as in the lottery) of sales revenue, reflecting a failure to modernize distribution and processing systems. Most pools customers still purchased from an army of 96,000 door-to-door collectors (Munting 1996), which was not only costly for the companies but also inconvenient for customers relative to the possibility offered by lotto to buy tickets at computer terminals operated in supermarkets, convenience stores and petrol stations.

It is therefore completely understandable why the introduction of lotto initiated a rapid decline in the pools sector. Of more interest is why demand continued to decline precipitously in a continuous fashion. It might have been thought that the pools sector would respond to the new environment by improving its product and that market size would eventually stabilize when it found a core of customers for whom a link with football was an appealing feature of a long-odds product.

The pools companies did respond over time to the lottery. From government, they won such concessions as a lower tax rate, the freedom to advertise on television, and the lowering of the minimum age for playing, from 18 to 16. To boost headline prizes, they offered players a "free" chance to win £2m prizes (unlikely to be paid out, as they were related to freakish sets of football results) and initiated rollovers of the jackpot. They modernized their distribution systems to embrace, for example, entering through the Internet or at a bookmaker's shop. All these measures were no doubt positive. However, the companies also raised the takeout rate, despite a fall in the rate of tax charged to pools. The Competition Commission of Great Britain (2007) inquiry noted that value for money had worsened in the period since the introduction of lotto, with a takeout of 75–80 percent in the immediately preceding years to its Report. Post-merger, takeout remained highly uncompetitive with any other gaming product. In 2009–10, Gross Gaming Yield was £41.0m on stakes of £52.4m (Gambling Commission 2010). This represents a takeout rate of 78.2 percent.

It is instructive to ask who plays the pools given the poor value it offers compared with other modes of gambling. The Competition Commission of Great Britain (2007) commissioned research that provides a snapshot profile of customers at that time, based on a sample of 1,100 participants (nearly all of whom proved to play weekly, consistent with the commission's observation that weekly sales appeared to be invariant to the size of lotto jackpots). The customer base was strongly skewed toward older males (in fact, only 6% of players were below age 45). Most customers had played for a very long time, 70 percent for over 20 years and more than half for over 30 years. This appears to be illustrative of a striking failure to refresh the customer base to replace those lost through mortality or defection.

The remaining players of the pools seemed largely ignorant of the effective price of entering the game. Effective price, rather than entry fee, is commonly used by economists to model demand for lottery-style products and is the expected player loss per unit stake. But only 1 percent of this sample was able correctly to identify the proportion of the stake paid out in prizes as in the band 21–30 percent. Those who have stayed with the pools seem to have done so out of both inertia and a lack of awareness of the high effective price.

The failure of the industry to recruit new customers is not surprising, given that for those seeking gambling opportunities with high potential winnings, the portfolio of games available from the National Lottery offers far superior value. For those with a taste for football, the pools offer a main game that does not challenge them to employ their football knowledge because it cannot realistically be claimed to be based on skill. Moreover, long-odds wagers can be constructed by wagering on football with fixed odds bookmakers who, in a vigorously competitive sector, have been offering much fairer bets than in the past (Forrest 2012). It is hard not to speculate that the pools industry strategy was not to pursue the market captured by football bookmakers but to be content to extract maximum revenue from existing players attached to the pools by habit and poor awareness of price.

The Football Pools in Spain

Based on information for 2007, provided by the World Lottery Association, Spain is clearly the leading football pools market in Europe, with annual sales of €547m, which we calculate to represent €12.24 of stakes per inhabitant. France, the second-ranked country by absolute turnover, reported a market only about one quarter as large. Some smaller countries, particularly the Nordic group, recorded high expenditures in per capita terms: Norway (€18.87), Sweden (€9.27), Finland (€6.73), Denmark (€4.45), Greece (€4.11), and Morocco (€2.30) (included in the data as a member of European Lotteries) exhibited greatest enthusiasm. For comparison, the sales of the shrunken British industry were equivalent to only about one euro per person per year.

Our focus is on Spain, where the relative importance of the pools has stimulated the only formal econometric modeling of pools games. The core product, known as La Quiniela, was launched in the 1946–1947 season into a market that already had passive national lottery games (but no legal bookmaking). The first coupon invited players to forecast results (1, X, or 2) for just 7 matches, but the list was soon increased to 14 (Pujol 2009), possibly in light of the success of the new long-odds treble chance game in Britain. The format remained at 14 matches until season 1987–1988, but a 15th match was added in 1988–1989 (it reverted to 14 in 2003–2004 but only for two seasons). Rule changes have included provision for the top prize to roll over to the next game (from 1988–1989) and adding a new lower-prize tier for those with 11 (1991–1992) or 10 (2003–2004) correct selections. The entry fee was increased regularly, from 2 pesetas at the beginning to €0.50 from 2003–2004. The takeout rate is 45 percent, of which 11 percent accrues to provincial governments to provide sports facilities, 1 percent to the National Council of Sports, and 10 percent to the Football League. The pools are operated by the same state-sanctioned agency that runs national lotto games. Lotto was first introduced in October 1985 in the form of the weekly La Primitiva. Subsequently, similar to the evolution of lotto in other countries, a second weekly draw was added, and over time a variety of alternative lotto games augmented the portfolio. Legal sports betting appeared in some regions from 2008.

With this background in mind, figure 8.1 shows real sales revenue (at 2007 prices), on a game-by-game basis, from 1970. Econometric analysis reviewed below attempts to account for the considerable variation from coupon to coupon. But the most striking features are the sharp fall in sales following the introduction of lotto and the subsequent recovery and stabilization. Stabilization was achieved at a level at which the pools remained a significant part of the operator's portfolio, in-season attracting similar weekly sales as one of the two weekly drawings of the core lotto product, La Primitiva (Forrest and Pérez 2011).

Who is entering the game? Brad Humphreys and Levi Pérez (2010) analyzed the findings from two random sample telephone surveys conducted in Spain in 2005 and 2006, with over 2,600 observations. Of the respondents, 49.7 percent reported having played La Quiniela at some time in their lives, with one-fifth of these playing weekly. A probit model showed that the likelihood of participation was positively related to income and negatively related to age. Compared with the British pools, the Spanish game attracted a much higher participation rate and disproportionately appealed to the

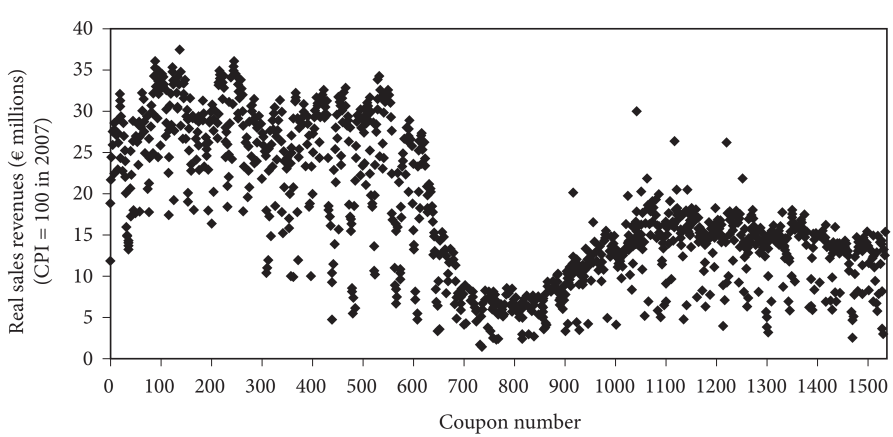


FIGURE 8.1 Real sales revenues of Spanish football pools

young and affluent. This makes the customer profile more like that for British sports betting. The Spanish pools no doubt benefited from the absence of a legal bookmaking sector but also had a product that was more skill-based than the British treble chance and one with a takeout rate approximately 30 percentage points lower. These seem likely reasons for the Spanish operator's success in maintaining a niche for the pools in a lotto world. Its principal challenge for the future may be to maintain this position as the emerging legal bookmaker betting market in Spain matures. Legalization of sports betting in some provinces appears to have coincided already with some fall off in demand.

Modeling Sales of La Quiniela

A number of papers (García and Rodríguez 2007; García, Pérez, and Rodríguez 2008 and 2011; Forrest and Pérez 2011) have analyzed pools sales using the same framework as followed in the lotto demand literature since the model developed by David Gulley and Frank Scott (1993). The Gulley-Scott model regresses the number of entries in a draw on the effective price of entries to that edition of the game (and controls). Effective price is the expected loss per unit stake from a single entry, assuming that all players select their entries randomly, and shows considerable variation according to the pattern of rollovers. The problem of endogeneity (sales themselves influence expected loss) is addressed by instrumenting effective price on the size of rollover. Variations include substituting the size of jackpot for effective price (Forrest, Simmons, and Chesters 2002) and adding variance and skewness of returns to the specification (Walker and Young 2001).

Applying the basic model for lotto to the pools game is problematic in terms of the calculation of effective price. It is straightforward to calculate the expected value of entering a randomly selected combination of numbers into a lotto game but not self-evident what would be meant by random choice of a set of football results. The studies reviewed here calculate expected loss on the basis that any entry will have the same probability of success of winning at each prize tier as is observed across entries in the long run (for example, the whole data period). It therefore treats players as if each perceives his or her chance of success as equal to that of anyone else. It could be argued that this is less plausible in a game which purports to require skill than in a pure numbers game, but it is a necessary simplification and one that, on the whole, appears not to prevent the models from achieving satisfactory fit.

On the other hand, modeling pools sales in the context of Spain offers some advantages over much of the literature on lotto, which may allow the exercise to yield stronger evidence than hitherto concerning the behavior of players of lotto-style games. First, most modeling of lotto sales has been conducted in jurisdictions where the entry fee has never varied from one unit of local currency, preventing results from informing policy debate on whether there should now be an updating of the price of a ticket, given high cumulative inflation since the inception of lotto. Second, the Gulley-Scott model abstracts from prize structure, and though this varies from draw to draw everywhere, effective price, size of jackpot, and variance and skewness of returns always move in sympathy with each other as size of rollover varies, given constancy in game rules. This makes it hard to distinguish the effects of each of these variables and therefore frustrates attempts to evaluate the merits of alternative game designs. But the history of La Quiniela incorporates multiple changes of entry fee, game design, and prize structure, which induce additional variation in the data and make it potentially easier to test hypotheses compared with lotto markets in America and Britain. Econometric modeling of pools demand in Spain may therefore serve more than a local purpose.

The Relationship between the Pools and Lotto Markets

Controlling for features of the coupon (which teams are represented), within-season trend, real entry fee and effective price, Forrest and Pérez (2011) focused on how pools sales have been and are influenced by the parallel existence of the core lotto product in Spain, La Primitiva. They investigated two issues. First, to what extent did the introduction of the lotto game displace pools demand? Second, do consumers switch between pools and lotto on a week-by-week basis, depending on the incidence of rollovers, making them substitutes? The two issues are frequently conflated in the literature on gambling but are in fact quite distinct. For example, a new product may cause structural change in the market for an existing product because a group of consumers changes permanently to buying the new product. But that does not imply that the two products will be substitutes for each other in the economic sense of cross-price elasticity henceforth being positive.

The analysis was based on observations from 1,514 editions of the pools game between April 1970 and December 2007. The lotto game was present in the market from observation 618, and the introduction of a second weekly draw coincided with observation 857. It should be noted that the new competing game was sold at exactly the same outlets as the pools and that consumers could therefore readily choose for which product to purchase tickets. It should further be noted that the second weekly lotto draw was timed for the weekend, when, in season, there is always a pools coupon.

Cannibalization was measured employing two dummy variables, respectively indicating whether it was a period with a Thursday-only lotto game or a period with Thursday and Saturday lotto. The results indicate a 52 percent fall in the number of pools entries as a result of the introduction of the lotto game, broadly similar to what was observed in Britain in the aftermath of the launch of its national lottery. With two lotto draws per week available, pools entries were estimated to be 87 percent lower than in the no-lotto case. The consistency of the order of magnitude of impacts in Britain and Spain suggests that, pre-lotto, more than half of the pools market was expressing purely a demand for long-odds gambling rather than for a specifically football-related game. Perhaps indeed this applied more strongly in Spain to the extent that, in contrast to Britain, there was no inducement to switch because lotto had a lower takeout rate (in Spain, both the pools and lotto offered a 45% takeout).

After cannibalization, whether consumers subsequently switch expenditure between pools and lotto on a week-to-week basis, as rollovers modify the relative value of the two products, is of practical importance to operators in Europe because, typically, the two types of game are provided by the same agency. Game formats are designed to produce periodic rollovers to boost sales above the normal level, but if extra sales were achieved just because players switched from an alternative product for that week, the rollover money would be wasted in terms of boosting the agency's global revenue. Forrest and Pérez (2011) estimated the elasticity of pools sales with respect to the effective price of Thursday lotto and the effective price of Saturday lotto. It was possible of course that the pools and lotto could be either substitutes or complements (the latter could apply, for example, if extra customers drawn to the lottery shop by a lotto rollover also bought a pools entry given that there would be no additional transaction costs). In the event, cross-price elasticity was zero for Thursday lotto. For Saturday lotto, it was positive and statistically ($p = .02$), but not economically, significant (i.e., numerically the estimate was close to zero). For practical purposes, it appears that the two gaming opportunities are not regarded as substitutes by consumers, consistent with the observation by the Competition Commission of Great Britain (2007) that in the United Kingdom national lottery rollovers do not deflate pools sales.

Sensitivity of Demand to Own-Price and to Game Design

The structure of lotto-style games lends itself to modeling demand, given that the exogenous pattern of rollovers induces strong variation in the value for money offered from draw to draw. Typically, value for money is captured by effective price and estimates of the elasticity of sales with respect to effective price are used to evaluate whether the takeout has been selected at a level where net revenue for the state is maximized (this is the approach, for example, in Farrell, Morgenroth, and Walker 1999 and Forrest, Gulley, and Simmons 2000). But the approach has problems to the extent that it is implausible that consumption decisions are driven only by effective price. It abstracts altogether from entry fee, which in most jurisdictions has been constant in nominal terms and declining in real terms for the whole period of any data set. It also ignores the potential importance of prize structure. Whenever there is a large rollover, lotto tickets become better value for money, but this happens only because of augmentation of the top tier prize; what is perceived by the model as a response to value for money may in fact just be a response to the size of a jackpot.

The papers here have the opportunity to address these two problems directly given the variability in the data on the pools and the length of the data set (records of sales are available from 1970). Regarding entry fees, multiple changes have been made over time, the overall trend being upward (García and Rodríguez 2007) such that periodic adjustments have more than compensated for inflation. Between adjustments of course real entry fee depreciates given that inflation in Spain has been consistently positive. With effective price included as one of their controls, Forrest and Pérez (2011) estimated the elasticity of the number of entries with respect to the real entry fee as −0.66 over the entire data period. This implies that, over the period, pushing up real entry fees lowered sales less than proportionately, allowing real revenue to rise. Successive, sometimes substantial, entry fee increases after lotto was introduced (with takeout held) therefore compensated the pools to some extent for the loss from cannibalization. The converse of the result is that allowing inflation to erode real entry fees indefinitely would have lowered real revenue (and the erosion of the real jackpot would then play a role in deterring consumers). This finding may be relevant to jurisdictions where the real price of a lotto ticket has fallen substantially since the game first appeared.

Forrest and Pérez (2011) also evaluated elasticity with respect to own-effective price as inelastic (-0.74), indicating that the pools game may, over the period, have been too generous from the perspective of raising money for the state and other claimants on the profits. But, in their case, there were no proxies in the specification for features of the game, such as prize structure or how hard it is to win. Jauma García and Plácido Rodríguez (2007) included in their specification proxies for the prize structure (real jackpot) and the probability of winning (a dummy representing when the game was based on 15 rather than 14 matches). Sales were shown to respond positively to the size of the jackpot and positively to the hardness of the game, but they found little role for effective price. This illustrates the danger of modelers misinterpreting the significance of effective price in a basic model. If effective price merely proxies other variables which influence sales, coefficient estimates cannot reliably be employed to predict the effect of any variations in takeout that would alter the size of all the prize pools and not just the jackpot. Success in game design requires careful attention to prize structure and game matrix, and simulations based on a model with only effective price will be inadequate for this purpose. In terms of pools policy, the importance of game design was underlined when sales were observed to fall after the pools reverted temporarily to 14 rather than fifteen 15 in 2003–2004 (the season after the end of the period employed in the econometric analysis).

The findings have implications here for lotto analysts outside Spain. Forrest, Robert Simmons and Neil Chesters (2002) also drew attention to the inadequacy of models based only on effective price by showing, tentatively, that U.K. Lotto sales were tracked more closely when the size of the jackpot pool replaced effective price in the specification. However, they were unable to include both in the same model because of collinearity. In the Spanish pools case, adequate variation in the data permitted García and Rodríguez (2007) to distinguish between the effects of the two variables, and they found that the size of the jackpot pool mattered to consumers. Omission of this variable or alternative variables related to prize structure is therefore likely to bias results in modeling lotto demand, and it is important to attempt to capture prize structure as well as effective price if findings are to offer useful guidance to operators and governments. In this spirit, Ian Walker and Juliet Young (2001) attempted to represent the various dimensions of a lottery ticket by including the mean, variance, and skewness of returns but were hampered by the problem of insufficient instruments for the number of endogenous variables.

Does the Football Matter?

The modeling reported so far treats the football pools as if they were a lottery game, and indeed it appears that players respond similarly as lotto players to stimuli linked to the costs and returns of playing. However, the ability of the pools in Spain to sustain a market, even with lotto readily available, suggests that basing a game on football results is an attraction to a segment of the market. Therefore, how football features in the game is liable to influence demand.

In their models, García and Rodríguez (2007) and Forrest and Pérez (2011) included as controls variables related to which matches feature on the coupon. In a full program of fixtures, the 15 matches chosen for La Quiniela constitute all Spanish First Division matches and some from the Second Division. However, a full program of matches is not always scheduled, for example, because the national team is playing. In this case there might be no First Division games and even no First or Second Division games on the coupon. Were the game regarded universally as just another lottery, this would not impact sales, but, in fact, faced with less familiar clubs making up the coupon, both papers find that players collectively purchased far fewer entries. For example, Forrest and Pérez (2011) predicted a 38 percent decrease in sales if there were no First Division matches, with the impact rising to 60 percent if the Second Division also was not featured. These estimates are for the entire period from 1970. García and Rodríguez (2007) reported slightly lower and slightly higher estimates, respectively, for subperiods when lotto was not and was available, consistent with a higher proportion of players in the later period treating the pools as a football-related product rather than a surrogate lotto game. In any case, it is clear that the link of the pools to football is important to those who buy. It is less clear why fewer purchase when the grade of football covered by the coupon is lower. One explanation is that a significant proportion of entries is made by customers who actively attempt to apply skill to the process of selecting their sets of results. They may feel less confident about having a chance of winning a prize when the teams featured are unfamiliar to them, or they may find it less fun if they cannot draw on their knowledge of the top teams in the top competitions. Alternatively the reason may be less tangible, as no doubt there are also many (non-gambling) products where brand association with the glamorous end of football increases sales.

A different but related issue is investigated in García, Pérez, and Rodríguez (2008). They employed panel data to model annual per capita sales of La Quiniela by province over the period 1985–2005. They found considerable variation from province to province, which is explained in their model by variables including per capita income (which has a positive effect) but also the number of teams in the province playing in the First Division and the number playing in the Second Division. The system of promotion and relegation provides that these variable show variation across time as well as across space. Both variables are signed positive, that for the First Division proving strongly significant ($p < .01$) but that for the Second Division only marginally so ($p = .06$). The long-run impact from a province having a team playing in the First Division was estimated as raising per capita sales by 11 percent. Again the reason for the finding is unclear. Players may simply be attracted more to products with local relevance, or they may think they can win with higher probability when they have been following the league closely, acquiring relevant knowledge on the way. Alternatively the probability of a First Division club being present in a province may be correlated with other characteristics of the province which, independently, predispose it to high pools sales. Despite uncertainty over causation, the evidence points to a complementarity between the sport and gambling on the sport. This may justify the payment of a significant part of pools gross revenue to the football league.

If association with football is part of the appeal of the pools game, which does indeed appear now not to be just another lottery, does this imply that significant numbers of players in Spain actively seek to forecast rather than merely guess the results? And, if they do, are they any good at it? These are less researched questions. García and Rodríguez (2007, 336) claimed that "the number of players getting all forecasts right is usually much higher than the number we would expect if the final results were completely random." This remark implies that the game is skills-based and that many players possess and apply the requisite skills. However, no detail is provided to justify that claim. It is not obvious what the benchmark proportion of winners is or how this would be calculated. How would a random set of results or forecasts be defined? Would it be with equal probabilities of the three results (1, X, 2), or would it be with matches allocated results randomly within the constraint that specified proportions would be allocated as home win/ draw/ away win?

There could be many routes to players scoring a higher success rate than would be expected from some definition of random play. For example, Barcelona and Real Madrid attract nationwide support, and their large numbers of fans may always select their team to win out of loyalty rather than because they are actively forecasting the result. But most of the time, these teams do win, therefore raising the aggregate success rate of pools winners. Consequently it is very hard to test whether skill is being successfully employed in the game. A further complication is that some players may seek out unlikely combinations of results because they take into account that they may win more money if no one else has matched all the results. Such behavior, though based on well-informed play, would tend to lower overall measured success rates in the game.

Francesc Pujol (2009) modeled the proportion of bettors who win over the course of more than 1,000 editions of the game. Controlling for the number of home wins on the coupon and the number of entries in the game, he identified a weak negative long-run trend, which he attributes to the decline in home advantage making matches harder to forecast and a strong positive within-season trend. The latter he interprets as reflecting pools players learning as the season goes on. At the beginning of a season, many clubs will essentially have a new squad, and some will have a new coach or be playing in a new division. Results will then be relatively hard to forecast, but as the season's events unfold, information is revealed about team strengths, and Pujol argues that this information is absorbed and processed efficiently by enough pools players that their collective win rate increases with time. The findings are certainly consistent with this explanation, though others might be investigated (for example, there may be less randomness in results as the season progresses if players on the top teams become more incentivized as the shape of the league table hardens or if players in new squads gradually gel to produce results commensurate with their total valuation on the transfer market).

How pools players make their selections is an area that requires additional research. As noted earlier, there have been few empirical studies of pools. Any future research will continue to focus on Spain because it is perhaps the one pools market where the size of the sector is large enough for investment of academic resources to be worthwhile.

References

Competition Commission of Great Britain. 2007. *Anticipated acquisition by Sportech plc of the Vernons Football pools business from Ladbrokes plc: Final report.* London: Competition Commission.

Cook, Philip J., and Charles T. Clotfelter. 1993. The peculiar scale economies of lotto. *American Economic Review* 83:634–643.

Dobson, Stephen, and John Goddard. 2001. *The economics of football.* Cambridge: Cambridge University Press.

Farrell, Lisa, Edgar L. W. Morgenroth, and Ian Walker. 1999. A time-series analysis of U.K. lottery sales: Long run and short run elasticities. *Oxford Bulletin of Economics and Statistics* 61(4):513–526.

Forrest, David. 1999. The past and future of the British football pools. *Journal of Gambling Studies* 15(2):161–176.

——. 2012. Online gambling: An economics perspective. In *Routledge International Handbook of Internet Gambling*, edited by Robert Williams, Robert Wood, and Jonathan Parke. London: Routledge 29–45.

Forrest, David, and Levi Pérez. 2011. Football pools and lotteries: Substitute roads to riches? *Applied Economics Letters* 18(3):1253–1257.

Forrest, David, and Robert Simmons. 2000. Making up the results: The work of the Football Pools Panel, 1963–1997. *The Statistician* 49(2):253–260.

Forrest, David, O. David Gulley, and Robert Simmons. 2000. Elasticity of demand for UK National Lottery tickets. *National Tax Journal* 53(4, pt. 1):853–863.

Forrest, David, Robert Simmons, and Neil Chesters. 2002. Buying a dream: Alternative models of demand for lotto. *Economic Inquiry* 40(3):485–496.

Gambling Commission (2010). *Industry Statistics 2009/10.* Birmingham, U.K.: Gambling Commission.

García, Jaume, and Plácido Rodríguez. 2007. The demand for football pools in Spain: The role of prices, prizes, and the composition of the coupon. *Journal of Sports Economics* 10(8):335–354

García, Jaume, Levi Pérez, and Plácido Rodríguez. 2008. Football pool sales: How important is a football club in the top divisions? *International Journal of Sport Finance* 3(3):119–126.

———. 2011. Guessing who wins or predicting the exact score: Does it make any difference in terms of the demand for football pools? In *Contemporary issues in sports economics: Participation and professional team sports*, edited by Wladimir Andreff. Cheltham, U.K.: Edward Elgar, 114–130.

Gulley, O. David, and Frank A. Scott Jr. 1993. The demand for wagering on state-operated lotto games. *National Tax Journal* 46(1):13–22

Humphreys, Brad R., and Levi Pérez. 2010. *A microeconometric analysis of participation in sports betting markets.* Economic Discussion Paper 02/2010. Oviedo, Spain: Department of Economics, University of Oviedo.

Miers, David. 1996. The implementation and effects of Great Britain's National Lottery. *Journal of Gambling Studies* 12(4):343–373.

Munting, Roger. 1996. *An economic and social history of gambling in Britain and the USA.* Manchester, U.K.: Manchester University Press.

Pujol, Francesc. 2009. *Football betting as a cyclical learning process* Faculty Working Paper 05/09. Pamplona, Spain: Department of Economics, University of Navarra.

Royal Commission on Gambling. 1978. *Report of the Royal Commission on Gambling.* London: H.M. Stationery Office.

Sharpe, Graham. 1997. *Gambling on Goals: A Century of Football Betting.* Edinburgh: Mainstream Publishing.

Walker, Ian, and Juliet Young. 2001. An economist's guide to lottery design. *Economic Journal* 111(475):F701–F722.

CHAPTER 9

THE EFFICIENCY OF SOCCER BETTING MARKETS

JOHN GODDARD

1 Introduction

Academic interest in the informational efficiency of fixed-odds betting markets for the results of professional soccer (football) matches dates back to the 1980s. In the first study of this topic, Peter Pope and David Peel (1989) analyzed the prices set by four national High Street bookmakers for English Football League matches played during the 1981/1982 season. As a basis for tests of the weak-form efficiency hypothesis, regressions of match outcomes against implicit bookmaker's probabilities are reported. Let $H_{i,j} = 1$ if the match between team i and team j results in a home win and 0 otherwise, and let $\varphi_{i,j}^{H}$ denote the bookmaker's implied probability for a home win. In the linear probability model $H_{i,j} = \rho_1 + \rho_2\varphi_{i,j}^{H} + v_{i,j}$, a necessary condition for weak-form efficiency is $\{\rho_1 = 0,\ \rho_2 = 1\}$. Equivalent conditions apply in the corresponding regressions for the draw and away-win dummy variables. In general, Pope and Peel failed to reject the condition $\{\rho_1 = 0,\ \rho_2 = 1\}$ for the odds on home-win and away-win match outcomes. The odds for draws have no significant predictive content for draw outcomes, however, suggesting a departure from weak-form informational efficiency conditions.

In a later study using similar methods, Michael Cain, David Law, and Peel (2000) reported evidence of favorite-longshot bias in the fixed-odds betting market for match results and scores in English soccer. The odds available for bets on specific scores are dependent only on the odds posted for the three possible match results (home win/draw/away win). Estimates of the "fair" odds for specific scores are computed, conditional on the bookmaker's posted home-win odds. In comparisons of the estimated "fair" odds with the bookmaker's actual odds for specific scores, the former are generally found to be significantly longer than the latter for long-shot bets. The "fair" odds are sometimes shorter than the bookmaker's odds for bets on strong favorites. Bets on strong favorites may offer limited profitable betting opportunities.

A second generation of soccer betting market efficiency studies, published in the late 1990s and early 2000s, investigated similar questions using forecasting models estimated from past match-results data to test the profitability of betting strategies informed by a comparison of bookmaker odds with the expected returns implied by probabilities for match results obtained from the estimated forecasting model. For example, Mark Dixon and Stuart Coles (1997) and Dixon and Peter Pope (2004) reported that bets for which the ratio of their forecasting model's probability to the bookmaker's implicit probability exceeded a threshold of 20 percent would have produced a positive return to the bettor. John Goddard and Ioannis Asimakopoulos (2004) compared probabilistic forecasts obtained from their model with a High Street bookmaker's fixed odds. A strategy of selecting bets ranked in the top 15 percent by expected return according to the model's probabilities would have generated a positive return. (For further discussion of soccer match–results forecasting models and betting market applications see Crowder et al. 2002 and Graham and Stott 2008.)

During the 2000s, the growth of the Internet and the migration of a significant share of sports betting business from High Street bookmakers' shops to websites operated by established bookmakers and new online betting firms likely have had profound implications for the informational efficiency of all sports betting markets, including those for soccer. The odds offered by competing bookmakers are easily compared, by visiting either the bookmakers' own websites or price comparison websites, such as oddschecker.com. The spectacular growth of the online betting exchange Betfair, which charges a commission of between 2 and 5 percent on the winnings of its successful clients, has placed downward pressure on the much larger margins that, traditionally, were built into the betting industry's fixed-odds betting pricing structure. Rapid enhancements in computing power have greatly facilitated the capability of industry members and sophisticated bettors to process large volumes of sports data in an effort to identify both inefficiencies in the market and opportunities for profitable trading that will, in time, tend to eliminate the inefficiencies. More generally, the Internet has increased the flow and quality of information on all aspects of sporting competition that is available to industry members and to sophisticated and leisure bettors alike.

In the light of these developments, this chapter evaluates the performance of forecasting models of the kind used in several of the academic studies reviewed briefly above when confronted with recent fixed-odds betting market prices data. Section 2 provides a brief, nontechnical description of a forecasting model for soccer match results that is based on a large-scale number-crunching exercise using historical match results and other relevant data. Section 3 examines whether the claims made in several earlier academic studies, that forecasting models of this kind can provide the basis for the development of profitable fixed-odds betting strategies, are sustainable when the model is confronted with recent fixed-odds betting prices. A prior hypothesis is that opportunities for profitable betting that were available during earlier periods might have been largely eliminated by the increased sophistication of contemporary sports betting markets, which have been greatly enhanced by advances in computing technology. Finally, Section 4 provides a brief summary and conclusion.

2 Statistical Analysis of Patterns in English Soccer Match Results

This section provides a brief, nontechnical description of the match-results forecasting model whose capability to inform the development of a profitable fixed-odds betting strategy is examined later in the chapter. A full technical description of the specification and estimation of this model is presented by Stephen Dobson and Goddard (2011). Early prototypes of the model, constructed along similar lines but with some variations in the choice of covariates, are described by Goddard and Asimakopoulos (2004) and Goddard (2005).

The variant of the forecasting model evaluated in this chapter is derived from an ordered probit regression in which match results in the home win/draw/away win format are the dependent variable. It is assumed that the result of the match between home team i and away team j, denoted here as $R_{i,j} = 1$ for a home team win, 0.5 for a draw, and 0 for an away team win, depends on an unobserved or latent variable denoted $y^*_{i,j}$ and an independent and identically distributed disturbance term, $\varepsilon_{i,j}$, which follows the standard normal distribution:

$$\text{Homewin}: R_{i,j} = 1 \text{ if } \mu_2 < y^*_{i,j} + \varepsilon_{i,j}$$

$$\text{Draw}: R_{i,j} = 0.5 \text{ if } \mu_1 < y^*_{i,j} + \varepsilon_{i,j} < \mu_2$$

$$\text{Awaywin}: R_{i,j} = 0 \text{ if } y^*_{i,j} + \varepsilon_{i,j} < \mu_1$$

The two cut-off parameters μ_1 and μ_2 are to be estimated. The latent variable $y^*_{i,j}$ is a linear function of a set of covariates, which are defined using match results and other data that are easily observable before the match between teams i and j is played. The coefficients of the linear equation determining $y^*_{i,j}$ are to be estimated. The covariates of this equation include the following:

- All previous league match results of teams i and j over the 24 months prior to the current match, which are used to calculate win ratios (the ratio of wins to matches played, with draws treated as half wins) indexed according to their timing within the 24-month period. More recent win ratio data have stronger predictive content for the current match than do earlier data. The indexing is also by division in which the team was playing at the time each win ratio was recorded. A discount is applied to win ratio data from a lower division, and a premium to data from a higher division, in evaluating the contribution of past performance to the probability of winning the current match.

- The few most recent league match results of teams i and j appear separately and individually among the covariates to allow for any patterns in sequences of results deriving from winning or losing streaks. Dobson and Goddard (2003, 2011) found that the durations of both good and poor sequences tended to be shorter than would be expected if streaks were irrelevant. In other words, a team currently enjoying a run of good results has a higher probability (relative to the probability conditioned only on that team's underlying quality) of a poor result in its next match and vice versa.
- The geographical distance between the stadiums of home team i and away team j. The phenomenon of home field advantage is pervasive throughout many professional team sports. Despite some erosion of the importance of home field advantage for English soccer match results over recent decades, a large gap still remains between average performance at home and away from home. The magnitude of home field advantage is, to a modest extent, dependent on the geographical distance the away team has to travel in order to complete the fixture: the greater the geographical distance, the stronger the home field advantage effect.
- Dummy variables are used to identify matches that are relevant or irrelevant for end-of-season championship, promotion, or relegation issues for either team. Teams that have end-of-season issues at stake are more likely to win matches played late in the season against teams with no issues at stake.
- Dummy variables are used to distinguish between teams that are currently involved in the FA Cup and European club competitions, and teams whose involvement has ended at the time the current league match is played.
- Recent average home league match attendance data for teams i and j, which (after adjustment for league position) are used as a proxy for any big club/small club effect on match results.

3 Evaluation of the Forecasting Model's Information Content Relative to Bookmakers' Fixed-Odds Betting Prices

This section examines the extent to which the match-results forecasting model described in the previous section could have informed the development of a profitable fixed-odds betting strategy had it been employed for this purpose in real time over the course of the three English soccer seasons 2008/2009 to 2010/2011 (inclusive).

After the estimation of the forecasting model, the fitted home win/draw/away win probabilities for any out-of-sample match for which a probabilistic forecast is required are obtained by substituting the evaluated $\hat{y}^*_{i,j}$ for the match in question, and the estimated cut-off parameters $\hat{\mu}_1$ and $\hat{\mu}_2$, into the following expressions:

$$\begin{aligned}
\text{home win probability} = p^H_{i,j} &= P(\varepsilon_{i,j} > \hat{\mu}_2 - \hat{y}^*_{i,j}) = 1 - \Phi(\hat{\mu}_2 - \hat{y}^*_{i,j}) \\
\text{draw probability} = p^D_{i,j} &= P(\hat{\mu}_1 - \hat{y}^*_{i,j} < \varepsilon_{i,j} < \hat{\mu}_2 - \hat{y}^*_{i,j}) \\
&= \Phi(\hat{\mu}_2 - \hat{y}^*_{i,j}) - \Phi(\hat{\mu}_1 - \hat{y}^*_{i,j}) \\
\text{away win probability} = p^A_{i,j} &= P(\varepsilon_{i,j} < \hat{\mu}_1 - \hat{y}^*_{i,j}) = \Phi(\hat{\mu}_1 - \hat{y}^*_{i,j})
\end{aligned}$$

$\Phi(\ )$ denotes the standard normal distribution function. Let $O^H_{i,j}$ denote the bookmaker's decimal odds for a bet on a home team win in the match between home team i and away team j. If the home team wins, the bettor's net profit on a unit bet is $O^H_{i,j} - 1$; if the home team fails to win, the bettor's net profit is -1. The bettor's expected profit from a bet on a home team win, according to the probabilities obtained from the model, is

$$E^H_{i,j} = p^H_{i,j}(O^H_{i,j} - 1) - (1 - p^H_{i,j}) = p^H_{i,j}O^H_{i,j} - 1$$

$O^D_{i,j}$, $E^D_{i,j}$, $O^A_{i,j}$ and $E^A_{i,j}$ are defined similarly for the bets on a draw and on an away win.

The empirical investigation reported in this section examines fixed-odds betting on match results in all four divisions of the English Premier League and Football League, expressed in home win/draw/away win format, during the 2008/2009, 2009/2010 and 2010/2011 soccer seasons. The fixed odds are those published by 10 High Street and online bookmakers: Bet365, Bet&Win, Gamebookers, Interwetten, Ladbrokes, Sportingbet, William Hill, Stan James, VC bet, and Blue Square. The data source is the soccer website www.soccer-data.co.uk.

Table 9.1 reports the mean overround (bookmaker's margin) of each bookmaker per season, where the overround $1/O^H_{i,j} + 1/O^D_{i,j} + 1/O^A_{i,j} - 1$ is the difference between the sum of the bookmaker's implied "probabilities" (defined as the reciprocals of the decimal odds) and the sum of "fair" probabilities (which must add up to one). The final row of table 9.1 reports the mean overround based on a set of "best odds," constructed by taking the longest of the decimal odds for each outcome quoted by any of the 10 bookmakers. Naturally the overround calculated from "best odds" is smaller than the overrounds calculated from the odds of each of bookmaker individually. It is interesting to note, however, that the "best odds" overround was positive for every match and was never zero or negative. This means that there were no "pure" arbitrage opportunities such that a bettor could have guaranteed a positive return by placing suitably calibrated bets on all three possible outcomes at "best odds." Table 9.1 also illustrates the general trend, identified previously, toward the erosion of the overround or the bookmakers' margins, with some (but not all) of the 10 bookmakers recording significant reductions in their mean overrounds during the three-season observation period.

Table 9.1 Mean Overrounds, 10 High Street and Online Bookmakers, 2008/2009 to 2010/2011 Seasons (Inclusive)

Bookmakers	2008/2009	2009/2010	2010/2011	All
Bet365	0.063	0.063	0.063	0.063
Bet & Win	0.109	0.107	0.106	0.107
Gamebookers	0.091	0.093	0.093	0.092
Interwetten	0.130	0.127	0.130	0.129
Ladbrokes	0.095	0.082	0.065	0.081
Sportingbet	0.101	0.101	0.101	0.101
William Hill	0.103	0.107	0.067	0.092
Stan James	0.091	0.089	0.078	0.086
VC bet	0.096	0.083	0.041	0.073
Blue Square	0.078	0.075	0.074	0.076
"best odds"	0.040	0.039	0.025	0.035

For the development of a profitable betting strategy, the information content of $\{p_{i,j}^k\}$ for k = H,D,A must be sufficiently strong for the bettor to generate a positive realized return by selecting those bets for which $E_{i,j}^k$ exceeds a certain threshold value, overcoming the in-built bias against the bettor caused by the bookmaker's overround. Tables 9.2 and 9.3 report the outcomes that would have been achieved through the implementation of such a strategy using the "best odds" for various threshold values for $E_{i,j}^k$. The threshold values are selected by ranking all available bets at "best odds" (three for each match) in descending order of $E_{i,j}^k$ and setting the threshold at each of the following quantiles of the ranked distribution of $E_{i,j}^k$: 0.5, 1, 1.5, 2, 3, 4, 5, 10 and 20 percent. Table 9.2 reports the total realized returns, numbers of bets, and average returns per bet for bets that fall within each quantile band. Table 9.3 reports the corresponding cumulative realized returns, the cumulative number of bets, and the cumulative average returns per bet, assuming the bettor had selected all bets with expected returns larger than the threshold value. For completeness, the final two rows of tables 9.2 and 9.3 report the corresponding figures at the 50 and 100 percent quantiles, even though bets falling into these bands normally would not be selected in view of their large, negative expected returns.

The results reported in tables 9.2 and 9.3 suggest that the forecasting model does contain relevant information not captured by the bookmakers' odds but that the model does not always produce forecasts that can be relied upon to deliver positive returns, especially over relatively small numbers of bets. According to table 9.3, setting the threshold parameter at the 2 percent quantile of the ranked distribution of $E_{i,j}^k$ over all three soccer seasons would have delivered a positive return of just over 1 percent on 366 bets. The same strategy applied season by season would have delivered positive returns

Table 9.2 Total Return and Average Return for Bets in Various Quantiles of the Distribution of Bets Ranked by Expected Returns Obtained from the Forecasting Model

Bets Ranked by Expected Return	Total Return and Number of Bets				Average Return per Bet			
	2008/2009	2009/2010	2010/2011	All	2008/2009	2009/2010	2010/2011	All
top 0.5%	16.2	−3.4	−11.5	1.3	0.675	−0.071	−0.575	0.014
	24	*48*	*20*	*92*				
0.5%–1%	−2.5	−12.8	12.0	−3.3	−0.105	−0.283	0.545	−0.036
	24	*45*	*22*	*91*				
1%–1.5%	8.7	−19.3	7.5	−3.1	0.433	−0.410	0.301	−0.033
	20	*47*	*25*	*92*				
1.5%–2%	−6.8	16.4	−0.8	8.8	−0.261	0.443	−0.027	0.097
	26	*37*	*28*	*91*				
2%–3%	12.5	−46.4	−18.9	−52.7	0.228	−0.760	−0.278	−0.287
	55	*61*	*68*	*184*				
3%–4%	−3.2	19.1	−10.4	5.5	−0.067	0.258	−0.170	0.030
	48	*74*	*61*	*183*				
4%–5%	3.0	−22.7	2.8	−16.8	0.055	−0.343	0.046	−0.092
	55	*66*	*62*	*183*				
5%–10%	11.6	−29.0	40.3	22.9	0.046	−0.093	0.114	0.025
	251	*313*	*352*	*916*				
10%–20%	−43.7	−33.8	−58.5	−136.0	−0.077	−0.060	−0.083	−0.074
	564	*560*	*709*	*1833*				
20%–50%	−12.2	−176.9	−22.7	−211.8	−0.008	−0.099	−0.011	−0.039
	1628	*1792*	*2077*	*5497*				
50%–100%	−237.2	−28.8	−30.6	−296.6	−0.069	−0.009	−0.011	−0.032
	3413	*3065*	*2684*	*9162*				

of 16.5 percent over 94 bets in 2008/2009 and 7.7 percent over 95 bets in 2010/2011, partially offset by a negative return of −10.8 percent over 177 bets in 2009/2010. Clearly the returns realized in each year are highly variable, and the bettor would need to adhere to the prescribed betting strategy for a period of several years' duration, in order to see the emergence of a more stable long-run average return.

These results suggest that using a forecasting model based on extrapolation from large volumes of historical match results and other data to identify fixed-odds bets offering relatively good value for money, which will either reduce or eliminate the losses that would otherwise be expected owing to the bookmaker's overround, is a feasible proposition. Developing a betting strategy that will reliably deliver a positive return in the long run is a more difficult task, however, though perhaps not an impossible one. In pursuit of a positive return, the statistician-bettor must be willing to adhere patiently to his or her chosen strategy for a long period and must be prepared to tolerate the possibility of realizing significant short-run losses from time to time.

Table 9.3 Total Cumulative Return and Average Cumulative Return for Bets in Various Quantiles of the Distribution of Bets Ranked by Expected Returns Obtained from the Forecasting Model

Bets Ranked by Expected Return	Total Cumulative Return and Cumulative Number of Bets				Average Cumulative Return per Bet			
	2008/2009	2009/2010	2010/2011	All	2008/2009	2009/2010	2010/2011	All
top 0.5%	16.2	−3.4	−11.5	1.3	0.675	−0.071	−0.575	0.014
	24	*48*	*20*	*92*				
top 1%	13.7	−16.2	0.5	−2.0	0.285	−0.174	0.012	−0.011
	48	*93*	*42*	*183*				
top 1.5%	22.3	−35.4	8.0	−5.1	0.328	−0.253	0.120	−0.018
	68	*140*	*67*	*275*				
top 2%	15.5	−19.0	7.3	3.8	0.165	−0.108	0.077	0.010
	94	*177*	*95*	*366*				
top 3%	28.1	−65.4	−11.6	−49.0	0.188	−0.275	−0.071	−0.089
	149	*238*	*163*	*550*				
top 4%	24.9	−46.3	−22.0	−43.5	0.126	−0.148	−0.098	−0.059
	197	*312*	*224*	*733*				
top 5%	27.9	−69.0	−19.2	−60.3	0.111	−0.182	−0.067	−0.066
	252	*378*	*286*	*916*				
top 10%	39.4	−97.9	21.1	−37.4	0.078	−0.142	0.033	−0.020
	503	*691*	*638*	*1832*				
top 20%	−4.2	−131.8	−37.4	−173.4	−0.004	−0.105	−0.028	−0.047
	1067	*1251*	*1347*	*3665*				
top 50%	−16.5	−308.7	−60.1	−385.2	−0.006	−0.101	−0.018	−0.042
	2695	*3043*	*3424*	*9162*				
all bets	−253.7	−337.5	−90.7	−681.8	−0.042	−0.055	−0.015	−0.037
	6108	*6108*	*6108*	*18324*				

Finally, to illustrate the degree of patience that might be required, consider the relationship between the number of bets placed (denoted N), the expected return from N bets, and the 95 percent confidence interval for the actual proceeds from N bets, assuming that the bettor employs a betting strategy that offers a small expected return of μ on each bet. For simplicity, assume all bets are placed at the same "fair" odds. The variance of the proceeds from any bet is $\sigma^2 = (O^k_{i,j} - 1 - \mu)^2 p^k_{i,j} + (-1-\mu)^2(1-p^k_{i,j})$, where $p^k_{i,j}$ is the probability that the bet wins. A 95 percent confidence interval for the actual proceeds from N bets is $\mathrm{N}\mu \pm 1.96\sigma\sqrt{N}$. Suppose $\mu = 0.05$ and bets are placed at "fair" decimal odds of $O^k_{i,j} = 3$ so that $p^k_{i,j} = 0.35$. With $\mathrm{N} = 300$ bets (for example), the expected proceeds are +15, but the 95 percent confidence interval extends from −33.6 to +63.6. In this case, no fewer than $N = 3,146$ bets would be required in order to make the lower limit of this confidence interval zero, implying a 0.95 probability of achieving a positive return.

4 Conclusion

This chapter evaluates the performance of a statistical match-results forecasting model that relies on a large-scale number-crunching exercise using historical match results and other relevant data when confronted with recent fixed-odds betting prices. Our prior hypothesis, that opportunities for profitable betting that were available during earlier periods might have been reduced or eliminated by the increased sophistication of contemporary sports betting markets, appears to be substantiated to some extent. While it appears that the forecasting model contains some relevant information that is not captured by the bookmakers' published fixed odds, the model does not always produce forecasts that will reliably deliver a positive return, especially if the number of bets is relatively small. And while there is some evidence for the existence of profitable betting opportunities, the statistician-bettor would need to adhere patiently to the chosen strategy for a long period in order to be confident of achieving a positive average or total return. The evidence for the existence of profitable betting opportunities appears somewhat weaker than claims that were made concerning informational inefficiency in fixed-odds betting markets by the authors of several papers of late-1990s and early-2000s vintage that adopted a similar modeling approach. This observation is consistent with the hypothesis that there has been an improvement over time in the informational efficiency of fixed-odds soccer betting markets.

References

Cain, Michael, David Law, and David Peel. 2000. The favourite-longshot bias and market efficiency in UK football betting. *Scottish Journal of Political Economy* 47(1):25–36.

Crowder, Martin, Mark Dixon, Anthony Ledford, and Mike Robinson. 2002. Dynamic modelling and prediction of English Football League matches for betting. *The Statistician* 51(2):157–168.

Dixon, Mark J., and Stuart G. Coles. 1997. Modelling association football scores and inefficiencies in the football betting market. *Applied Statistics* 46(2):265–280.

Dixon, Mark J., and Peter F. Pope. 2004. The value of statistical forecasts in the UK association football betting market. *International Journal of Forecasting* 20(4):697–711.

Dobson, Stephen, and John Goddard. 2003. Persistence in sequences of football match results: A Monte Carlo analysis. *European Journal of Operational Research* 148(2):247–256.

——. 2011. *The economics of football.* 2nd ed. Cambridge: Cambridge University Press.

Goddard, John. 2005. Regression models for forecasting goals and match results in association football. *International Journal of Forecasting* 21(2):331–340.

Goddard, John, and Ioannis Asimakopoulos. 2004. Forecasting football results and the efficiency of fixed-odds betting. *Journal of Forecasting* 23(1):51–66.

Graham, I., and H. Stott. 2008. Predicting bookmaker odds and efficiency for UK football. *Applied Economics* 40(1):99–109.

Pope, Peter F., and David A. Peel. 1989. Information, prices and efficiency in a fixed-odds betting market. *Economica* 56(223):323–341.

CHAPTER 10

THE EFFICIENCY OF PELOTA BETTING MARKETS

LORETO LLORENTE, JOSEMARI AIZPURUA, AND JAVIER PUÉRTOLAS

Introduction

We study the very attractive betting system used in pelota games in the Navarra, Basque Country, and La Rioja regions of Spain. In what follows we call this the pelota betting system. This betting system may also be of interest due to its similarities to person-to-person betting, that is, betting on the exchanges. Before presenting the main work, we provide a brief introduction regarding pelota games and the betting system.

Most authors who have studied the origins of the Basque sport of pelota believe it is linked to the medieval hand-ball game of *jeu de paume* or "real tennis." The history of this game was passed down orally, so the first written references to it do not appear until 1331. In the eighteenth century the game began to die out in France, Italy, and England, but it grew more popular in the Basque Country, where increased economic prosperity led to the building of more courts.[1]

Although there is evidence of betting on ball games in ancient times, there is no evidence of what type of wagers were made. We do not know when the betting system currently used on the pelota courts of the Basque Country first came into being, but we can say that no evidence has been found of this system operating outside this region. Traditionally the rules of the system were not written down but were passed on orally, which makes it harder to study. We have found a similar betting system that has been implemented on the Internet. The betting system followed on betting exchanges, for example Betfair, is similar to the pelota betting system in that bettors can make as many bets as they want provided there is another bettor on the other side and that the market maker takes a percentage of the money as a commission. As shown by Edelman and Llorente (2010), the primary difference between the two systems is the pelota betting system's odds scale. Michael Smith, David Paton, and Leighton Vaughan

Williams (2006) have studied market efficiency in person-to-person betting, and David Marginson (2010) has studied efficiency in the exchanges, pointing out the impact of betting on a "known loser." The latter takes on importance when a sport can have more than two mutually exclusive and exhaustive outcomes, as in horse races.

In America a different kind of game known as jai alai or cesta punta also is played. Jai alai originated from the version of pelota called cesta punta, but the sport is now different. Bets are also made in jai alai, but the betting system in America is totally different from the system described here. A more related topic, betting on team jai alai, the game played in Mexico City, Connecticut, Florida, Nevada, Rhode Island, and other locales, has been studied by Daniel Lane and William Ziemba (2008). But the betting system they analyzed also differs from the pelota betting system in terms of the study of efficiency.

In this chapter, we examine efficiency in the pelota betting system under different scenarios. We make use of the three concepts of efficiency widely utilized in various empirical settings; Constant Returns, Absence of a Profit Opportunity, Equilibrium Pricing Functions (Sauer 1998, 2024), and give some suggestions for further research.

In the first section, we describe the game and the betting system (a complete description can be found in Llorente and Aizpurua 2008). In section 2, we position the pelota betting system within the framework of the economic literature. In section 3, following Llorente 2007, we exploit peculiarities of the betting system to make some assumptions that allow us to analyze what we will call the *general odds rule*. Analyzing the way odds are determined in this market, we find inefficiencies both assuming equal return of bets and when no profitable betting strategies are allowed. A third notion of efficiency, reporting findings from an analysis of field data in these markets, is presented in section 4, and section 5 considers hedging strategies. The final section contains a summary discussion and conclusions.

1 A Description of the Pelota Betting System

All pelota (pelota vasca) matches are played by two teams: reds (R) and blues (B) play against each other by hitting a ball in turn against a wall on a court called a *frontón*. The team that serves first is chosen by tossing a coin. When a team makes an error the opponent scores one point and serves to start the next point. The team that reaches a pre-set number of points wins the match.

A bet on these matches is made based on two quantities that inform the odds: the amount of money the bettor loses for failing to predict the winning team and the amount the bettor wins for guessing right. Bets can be made during the entire match, so as points are scored, odds change. A bettor can place as many bets as he or she wants, provided someone can be found to accept those bets.

This betting system is popular in the Basque Country, Navarra, and La Rioja, where several types of pelota games are played. There are slight differences between the different types. In what follows we will describe the betting system in particular to the type called remonte. The rules of the betting system are not written, thus all our explanations are based on information obtained at the *frontón*. More specifically, our study is based on information collected at the Euskal Jai Berri, a *frontón* where the pelota game played is remonte. This *frontón* was chosen because it has screens that display the odds as they change during the game, a peculiarity that is very helpful for obtaining field data.

1.1 A Brief Description of Remonte

Remonte is a type of pelota game that looks like a jai alai game. The *frontón* on which it is played consists of a playing court limited by three walls at the front, the left, and the back. It measures about 54 meters long, 12 meters wide, and 11 meters high.

In this game, reds and blues hit a ball with a wicker scoop called the *cesta* that is attached to the players' hands. The teams usually have two players each but may occasionally have one or three. Each team has to hit the ball in turn, starting with the team chosen by the field judge by tossing a coin. When one player fails to hit correctly the opponents score one point and serve the ball in order to start the next point. The first team to reach 40 points wins the match.

To play a point, each team has two possibilities. The first and most common one is to hit the ball against the front wall so that the ball bounces on the floor inside the limits marked. The second one is to hit the ball against the left wall so that it rebounds against the front wall and then bounces on the floor inside the court. Once the ball hits the front wall the other team is allowed to hit the ball either before or after it touches the floor (only once) or even after it bounces off the back wall. Each game usually lasts around one hour.

1.2 A Brief Description of the Betting Market in Remonte

Throughout the game spectators are allowed to place as many bets as they want. On *each bet* the bettor chooses either the red team or the blue team and wagers an amount of money against another spectator who chooses the other team and wagers a different amount of money (these amounts become the odds). The bettor who guesses correctly wins the money that the other bettor loses. For example, if you bet 100 euros on the reds against an opponent who bets 100 on the blues and the blues win the match, you pay your opponent 100 euros. But all bets are placed through a middleman who works for the organizers and takes 16 percent of the winnings of the successful bettor (in what follows this is called the commission); so in the previous example, you would pay

Table 10.1 The Screen at Euskal Jai Berri *frontón*

	Odds	Score	O_R is the amount a bettor risks by betting on reds.
(reds)	O_R	s_r	O_B is the amount a bettor risks by betting on blues.
(blues)	O_B	s_b	s_r (s_b) is the red (blue) team's score.

Table 10.2 An Example of the Presentation of Odds and Score

	Odds	Score
(reds)	100	5
(blues)	80	2

your opponent 100 euros, but he or she would only receive 84 because the middleman takes 16.

Throughout the remonte match a screen (see table 10.1) shows the effective odds in the market and the current score.

The odds consist of two numbers, with the bigger one always being 100 euros and the smaller one varying between 2 and 100 as points are played. Generally the smaller odd is one of the set {2, 3, 4, 5, 6, 8, 10, 12, 15, 20, 25, 30, 40, 50, 60, 70, 80, 90, 100}.

In the example in table 10.2 the red team has scored 5 and the blue team 2. The odds are 100 to 80, denoted by (100, 80). The bettor who bets on reds risks 100 euro to win 80, and the one who bets on blues risks 80 euro to win 100. From here on the chapter will follow this convention in describing the various bets, and we will denote these odds (OR, OB). Bets are always between spectators, so if one spectator places a bet on red there must be another who bets on blue. What happens when the game is over?

Bet on reds: A bettor on reds will lose 100 euros if blues win the match. Otherwise, if reds win the match the bettor will win 80 euros minus the 16 percent commission, that is, 67.2 euros.

Bet on blues: A bettor on blues will lose 80 euros if reds win the match. Otherwise, if blue wins the match the bettor will win 100 euros minus 16 percent, that is, 84 euros.

This describes a single bet. A bettor can make as many bets as he or she wants to, provided there is someone on the other side who will take the bets. For example, with the same screen as represented in table 10.2, if a bettor places 10 bets on the reds he or she will lose $100 \times 10 = 1{,}000$ euros if blues win the match. Otherwise, if reds win the match the bettor will win $67.2 \times 10 = 672$ euros.

Moreover, a bettor can bet at different times during the game, choosing one team in one period and the other team in another. Therefore the result above concerns only the particular bet analyzed here.

1.3 The Way the Odds Are Fixed in the Market

In remonte there is an auctioneer, who posts the odds that appear on the screen. We call him the coordinator. He is usually someone who has been a player and a middleman for many years. This coordinator is an expert on *remonte*, usually the person at the *frontón* who is most knowledgeable about the game. The coordinator chooses the handmade balls used by players to play the match, and he posts the odds at which people bet. The coordinator sits in front of a computer, in a privileged place behind the spectators, where he can follow the match and see the spectators and all the middlemen. He posts the odds that appear on the screens. After a team scores a point it is closer to winning and thus more likely to win the match than before the score was made. Therefore the money a bettor risks betting on that team should be higher in order to maintain the expected value of the bet.

There are some general rules that the coordinator follows to set and change the odds:

If there is no reason to think of either team as the favorite just before the match starts, the score is zero-zero and the odds (O_R, O_B) are (100, 100). If there is a favorite team, the odds may be different. For example, if red is favorite, with the same score as above, we would have odds (O_R, O_B) of (100, 80).

Once it is clear what the initial odds are, the match goes on and points are scored. The general rule is that the difference between the amounts of the odds increases by 10 euros on the team that has just scored a point, keeping in mind that the larger amount in the odds is always 100.

If the odds differ by more than 70 (100, 30), they change by only 5 euros for each point. If they differ by more than 90, the odds change by only 2 euros for each point. When one team has accumulated approximately 30 points, the change in the odds doubles for each point then scored and trebles or quadruples when the end of the match is very near. For example, if the score is (38, 39) the odds would be (40, 100).

Of course, sometimes these rules are modified because of changes in supply and demand among spectators. When a middleman finds two people who want to bet at odds different from those on the screen, the middleman has to ask the coordinator to change the odds on the screen so that he can print the receipts for the bettors. There are no bets on the *frontón* at odds different from those on the screens. In general the odds vary mainly as the above rules indicate.

It is important to realize that the coordinator works for the firm that organizes pelota matches, so his goal could be described as making people bet as much as possible. We can confidently assume therefore that the odds he posts are those at which people are most willing to bet, that is, equilibrium odds where there is no excess of demand of bets.

2 The Pelota Betting System: Similarities and Differences with Other Studies

Studies of different sports and differing betting systems can require different analyses and/or hypotheses, and this affects how efficiency in the market is analyzed. Raymond Sauer (1998) mentions different definitions of market efficiency that are widely utilized in empirical settings.

Before presenting different ways of analyzing efficiency in the pelota betting market, this section presents an overview of the betting systems that were prominent in the economic literature before betting exchanges were established on the Internet. It also will be useful to show the framework of our study within the economic literature, pointing out, according to our point of view, which are the relevant characteristics of other studies related to the pelota betting system.

2.1 Summary of Betting Systems Prominent in the Economic Literature

According to Sauer (1998) and Richard E. Quandt (1986), three forms of wagering are prominent in the relevant economic literature: pari-mutuel odds, odds offered by bookmakers, and point spread, also offered by bookmakers. The pari-mutuel system is used exclusively by racetracks in North America, France, Hong Kong, and Japan and coexists with the bookmaking market in Australia and Great Britain. Nevada bookmakers take bets on races at major tracks and offer odds on point spreads on team sports such as baseball, basketball, and football. The legal bookmaking market is less restricted, more extensive, and more liquid in Great Britain and Australia.

In pari-mutuel betting individuals invest in shares of the various horses. The prices of the shares are standardized, but the payoffs depend on the amount bet on a particular horse relative to the amount bet on all horses. Particularly in a pari-mutuel market people place wagers on which of two or more mutually exclusive and exhaustive outcomes will occur at some time in the future. A predetermined percentage is taken out of the betting pool to cover the market maker's costs, and after the true outcome becomes known, the remainder is returned to winning bettors in proportion to their individual stakes. Therefore the possible payoffs are not known until the last bet is placed. Pari-mutuel markets are common at horse races, dog races, and jai alai games.

Bookmakers offer their customers a set of payoffs conditioned on the outcome of a given event. The payoffs offered may change during the betting period—in general the more likely an outcome, the lower the payoff—but the payoff for each bet is determined at the time the bet is placed. The return conditioned on winning is thus known at the time of the wager; this is in contrast to the pari-mutuel system, where heavy betting late in the period can reduce returns below acceptable levels. Individuals who make bets large enough to affect the odds may prefer to bet with a bookmaker.

Point spread betting on football games is the staple of the Las Vegas sports betting market. In a point spread wager the payoff depends on the difference in points scored by the two opposing teams. Point spreads (PS) are typically reported as the number of points by which one team is favored to beat another. The actual difference in points (DP) is defined as the points scored by the favored team less those scored by the underdog. Bets on the favorite pay off when DP−PS >0, bets on the underdog pay off when DP−PS <0, and all bets are refunded when DP = PS. The eleven for ten rule characterizes standard terms in the Las Vegas market. Where commission $ts = 0.1$ and successful bets return net winnings of \$1 to every $\$(1 + ts)$ wagered. The point spread represents a price in this market. Let p represent the probability that wagers on the favorite will pay off, that is, that $p =$ prob (DP−PS) >0. The expected cost of an attempt to gain \$1 by betting on the favorite is the amount wagered times the probability of losing or $\$(1 + ts)(1 - p)$. Since p falls as PS increases, the expected cost of a wager on the favored team, that is, its price, increases as PS increases.

The pelota betting system fills an unrealized gap in the relevant literature before the advent of online betting exchanges. Although its territorial scope is limited, its cultural importance is high. Many fans follow pelota matches not only at the *frontón* but also on television. We emphasize that although the Pelota betting system has operated for centuries in the regions of Navarra, the Basque Country, and La Rioja, its rules are not written down but passed on verbally. This has hindered its spread to other regions, and thus it has not been studied, even though its importance in the area is remarkable. Both its peculiarities and its theoretical simplicity make analyzing the pelota betting market an interesting exercise. Nowadays the relevance of this system is more obvious due to its similarity with the betting exchanges (see Edelman and Llorente 2010).

2.2 The Pelota Betting System in the Economic Literature

Stephen Skiena (1988) studied a jai alai game where both the game and the betting system are completely different from those described here. Conversely, the team jai alai game studied by Lane and Ziemba (2008) is similar to the sport of pelota, but the betting system they studied is bookmaking.

The pelota betting system is similar to that on betting exchanges. Edelman and Llorente (2010) compared both and showed that normalizing the odds in the pelota betting system (dividing them by O_R) makes *a bet on the reds* similar to a *Back on the reds at odds x* and betting on the blues similar to *Laying a bet on the reds at odds x* whenever $\frac{O_B}{O_R} = x - 1$. Smith, Paton, and Vaughan Williams (2006) and Marginson (2010) studied efficiency in person-to-person betting or betting exchanges. The latter study focused on the case of the "known loser" in sports, where there are more than two possible results, as in horse races. In pelota there are only two states of nature, either the reds win or the blues win. Another point of view is to model pelota bets as a game between two bettors, as done by Werner Güth, Anthony Ziegelmeyer, and Llorente (2009), who used pelota bets to study experimentally inconsistent beliefs in a betting game.

3 The General Odds Rule

In this section, following Llorente (2007), we take advantage of the peculiarities of the betting system to make some assumptions that allow us to analyze what we call the *general odds rule*.

We were told by the coordinator in the market that games are arranged in such a way that the chances of winning for each team are as similar as possible. From this peculiarity we assume that, in general, the probability of the following point being scored by each team is 50:50. As we will see, this assumption allows us to obtain each team's theoretical probability of winning (see subsection 3.1).

The coordinator also told us that he follows a general rule to set odds in the market so long as nothing atypical happens (where atypical means that there is an excess of people willing to bet on one color). Odds differ from those of the general rule if, for example, it can be seen that a player has lost his ability to play or something happens that causes more bettors to bet on one team more than the other. From that we can infer that odds in the general rule are equilibrium odds (i.e., there is no excess of willingness to bet on one of the two teams), when players have the same probability of scoring the following point. Therefore we posit the general odds rule as the equilibrium price in the market when the probability of scoring the following point by each team is 50:50. From these market odds, and under some assumptions that will be explained in subsection 3.2, we can obtain the probabilities inferred from market odds.

We will compare the theoretical probability with the probability inferred from market odds to check whether the market is efficient or not.

3.1 Teams' Theoretical Probabilities of Winning the Match

Assuming that the likelihood of a team scoring the following point is the same throughout the match, the theoretical probability of the team's winning the match can be derived at any time. With no loss of generality we perform such a calculation for the reds. Given the reds' score, s_r, the blues' score, s_b, and the reds' probability of scoring a point at any moment during the match, p, the reds' probability of winning the match, p_r, is given by

$$p_r(s_r, s_b, p) = \sum_{i=0}^{40-s_b-1} \binom{i+40-s_r-1}{i} p^{40-s_r}(1-p)^i. \tag{10.1}$$

This equation is obtained by the addition of as many amounts as there are different possible final scores where the reds are the winners (i is the number of points the blues could score, from 0 to $39-s_b$). For each of these possible final scores, the amount added is obtained by multiplying the probability of this final score by the number of different ways in which it can be reached.

In pelota games the probability of each team scoring the following point and the probability of winning the match are unknown, but we know that matches are arranged so that the chances of each team winning are as similar as possible according to the subjective perception of the organizer of the match: when one team is superior the odds are shortened by using match-balls that favor the worse team and so on. When one team is superior the other may even be allowed to use one more player. Thus for the sake of simplicity we set the reds' probability of scoring the following point at $p = 0.5$. This in conjunction with equation 10.1, above, allows us to calculate, for a given score, the reds' probability of wining the match, p_r, as

$$p_r(s_r, s_b) = \sum_{i=0}^{39-s_b} \binom{i+39-s_r}{i} 0.5^{40-s_r} 0.5^i.$$

We call this the reds' *theoretical probability* of winning the match.

These theoretical probabilities for some possible scores are shown in table 10.3.

Table 10.3 Reds' Theoretical Probabilities of Winning Given the Current Score

(*sr*,*sb*)	Theoretical p_r
(1,0)	0, 545
(2,0)	0, 5901
(3,0)	0, 6345
(4,0)	0, 6778
(5,0)	0, 7193
(6,0)	0, 7586
(7,0)	0, 7952
(8,0)	0, 8288
(9,0)	0, 859
(10,0)	0, 8858
(11,0)	0, 9091
(12,0)	0, 929
(13,0)	0, 9456
(14,0)	0, 9592
(15,0)	0, 97
(16,0)	0, 9785
(17,0)	0, 985

Actually, the particular problem of betting analyzed here is closely connected with the so-called problem of points that initiated the study of probability. That problem, also known as the division problem, was proposed by Antoine Gombaud, chevalier de Méré, to Blaise Pascal and led to correspondence between the latter and Pierre de Fermat in the summer of 1654 (see Edwards 1982). Based on this correspondence Pascal and Fermat are said to be among the joint discoverers of probability calculus (see David [1962] 1998, 75). According to their discussion it can be concluded that the probability of winning for a player who needs j points to win (each obtained with probability p) against another who needs k points to win (each point obtained with probability $q = 1 - p$) is given formally[2] by $P_{j,k} = \sum_{l=j}^{j+k-1} \binom{l-1}{j-1} p^j q^{l-j}$. As can be easily checked, the equation that we propose, (10.1), is very similar. Indeed, it is the same formula after a change of variables. We use equation (10.1) for fluency to obtain the theoretical probability given the score in our particular betting market.

3.2 Teams' Probabilities of Winning the Match Inferred from Market Odds

In this subsection we try to analyze the market with orthodox methods; particularly we analyze the general odds rule followed in the pelota betting system under certain assumptions that may be somewhat strong but are also usually made in studies of other wagering markets. Thus we start by assuming that bettors are expected value maximizers. A condition for equilibrium is that the expected value of a bet on the reds should be equal to the expected value of a bet on the blues; if not, all bettors prefer to bet on the color with the higher return. In subsection 3.2.2 we discuss the probability inferred from the market assuming equal returns on each bet and determine that there is a difference between the probability inferred from market odds and the theoretical probability of a team winning the match. Low probabilities are overestimated while high probabilities are underestimated.

In these markets there are commissions, so the equilibrium condition of equal return of bets implies that each bet has a negative expected return. Therefore it seems more convenient to introduce the less restrictive restriction of not allowing profitable bets in the market. This is done in subsection 3.2.3, where we calculate what the probabilities inferred from the market odds must be to satisfy this less restrictive restriction of no profitable bets, that is, to satisfy the condition of the expected value of each bet being lower than or at most equal to zero. Comparing this probability inferred from market odds with the actual probabilities we find that in these markets there are profitable betting strategies.

Table 10.4 General Odds Rule

Score (s_r, s_b)	Odds (O_R, O_B)
(1,0)	(100,90)
(2,0)	(100,80)
(3,0)	(100,70)
(4,0)	(100,60)
(5,0)	(100,50)
(6,0)	(100,40)
(7,0)	(100,30)
(8,0)	(100,25)
(9,0)	(100,20)
(10,0)	(100,15)
(11,0)	(100,10)
(12,0)	(100,8)
(13,0)	(100,6)
(14,0)	(100,4)
(15,0)	(100,2)
(16,0)	(100,2)
(17,0)	(100,2)

3.2.1 *The General Odds Rule*

As already mentioned, for each score we can derive the odds in the market by applying the general odds rule. Table 10.4 shows some scores and their corresponding market odds.

3.2.2 *Efficiency of the General Odds Rule Assuming Equal Return on Bets*

Now that we know the odds in the market for each score, we can derive the probabilities inferred from these odds as follows. If a pelota betting market is efficient, where "efficiency" means that the expected returns are equal on the various bets (see the different meanings of "efficiency" in Sauer 1998, 2024), then equation (10.2) should apply.

$$\frac{p_r}{1-p_r} = \frac{O_R}{O_B}, \tag{10.2}$$

where p_r is the likelihood of the reds winning the match, O_R is the money risked in a bet on the reds and O_B is the money risked in a bet on the blues.

Proof. Denote by EVR the expected value of a bet on the reds and EVB the expected value of a bet on the blues. We know that $EVR = p_r\, O_B(1-t) - (1-p_r)\, O_R$ and that

$EVB = (1 - p_r)\, O_R(1 - t) - p_r O_B$, where t is the middleman's commission. If the market is efficient the two expected values should be equal, thus $p_r O_B(1 - t) - (1 - p_r)O_R = (1 - p_r)O_R(1 - t) - p_r O_B$, and by operating we obtain $\frac{p_r}{1-p_r} = \frac{(2-t)O_R}{(2-t)O_B}$, which proves that equation (10.2) is true no matter what the commission is.

Rearranging equation (10.2), $p_r = \frac{O_R}{O_B}(1 - p_r)$; $p_r\left(1 + \frac{O_R}{O_B}\right) = \frac{O_R}{O_B}$; $p_r = \frac{O_R/O_B}{\left(\frac{O_B+O_R}{O_B}\right)}$; therefore

$$p_r = \frac{O_R}{O_B + O_R}. \tag{10.3}$$

The probability inferred from the odds is obtained by equation (10.3). These probabilities are shown in table 10.5.

In table 10.6 we put together four columns: first, of all the scores; second, the market odds for each score applying the general odds rule; third, the reds' theoretical probability of winning the match obtained for each score; and fourth, the reds' probability of winning the match obtained from the market odds by applying equation (10.3).

In figure 10.1 a scatterplot is shown of the probabilities inferred from markets odds, π_r, against the corresponding *theoretical* probability, p_r.

Table 10.5 Probability of Reds Winning Inferred from Market odds Assuming Equal Return of Bets
***Note*: Obtained by equation (10.3).**

Odds (O_R,O_B)	p_r Derived from Market Odds $= \pi_r$
(100,90)	0,5263
(100,80)	0,5556
(100,70)	0,5882
(100,60)	0,625
(100,50)	0,6667
(100,40)	0,7143
(100,30)	0,7692
(100,25)	0,8
(100,20)	0,8333
(100,15)	0,8696
(100,10)	0,9091
(100,8)	0,9259
(100,6)	0,9434
(100,4)	0,9615
(100,2)	0,9804
(100,2)	0,9804
(100,2)	0,9804

Table 10.6 Probability of Reds Winning; Theoretical and Derived from Market Odds

Score (sr,sb)	Odds (O_R,O_B)	Theoretical p_r	p_r Derived from Market Odds
(1,0)	(100,90)	0,545	0,5263
(2,0)	(100,80)	0,5901	0,5556
(3,0)	(100,70)	0,6345	0,5882
(4,0)	(100,60)	0,6778	0,625
(5,0)	(100,50)	0,7193	0,6667
(6,0)	(100,40)	0,7586	0,7143
(7,0)	(100,30)	0,7952	0,7692
(8,0)	(100,25)	0,8288	0,8
(9,0)	(100,20)	0,859	0,8333
(10,0)	(100,15)	0,8858	0,8696
(11,0)	(100,10)	0,9091	0,9091
(12,0)	(100,8)	0,929	0,9259
(13,0)	(100,6)	0,9456	0,9434
(14,0)	(100,4)	0,9592	0,9615
(15,0)	(100,2)	0,97	0,9804
(16,0)	(100,2)	0,9785	0,9804
(17,0)	(100,2)	0,985	0,9804
(0,1)	(90,100)	0,455	0,4737
(0,2)	(80,100)	0,4099	0,4444
(0,3)	(70,100)	0,3655	0,4118
(0,4)	(60,100)	0,3222	0,375
(0,5)	(50,100)	0,2807	0,3333
(0,6)	(40,100)	0,2414	0,2857
(0,7)	(30,100)	0,2048	0,2308
(0,8)	(25,100)	0,1712	0,2
(0,9)	(20,100)	0,141	0,1667
(0,10)	(15,100)	0,1142	0,1304
(0,11)	(10,100)	0,0909	0,0909
(0,12)	(8,100)	0,071	0,0741
(0,13)	(6,100)	0,0544	0,0566
(0,14)	(4,100)	0,0408	0,0385
(0,15)	(2,100)	0,03	0,0196
(0,16)	(2,100)	0,0215	0,0196
(0,17)	(2,100)	0,015	0,0196

The graph shows that for low theoretical probabilities the probability inferred from market odds is higher, and for high probabilities the probability inferred from market odds is lower than is actually the case. Thus this analysis, as well as that in other studies, supports the long-shot bias. Empirical evidence on horseracing is found in Dowie (1976), Henery (1985), Thaler and Ziemba (1988), and Vaughan Williams and Paton (1997). And empirical evidence of the long-shot bias in greyhound racing is found in Cain, Law, and Peel (1992).

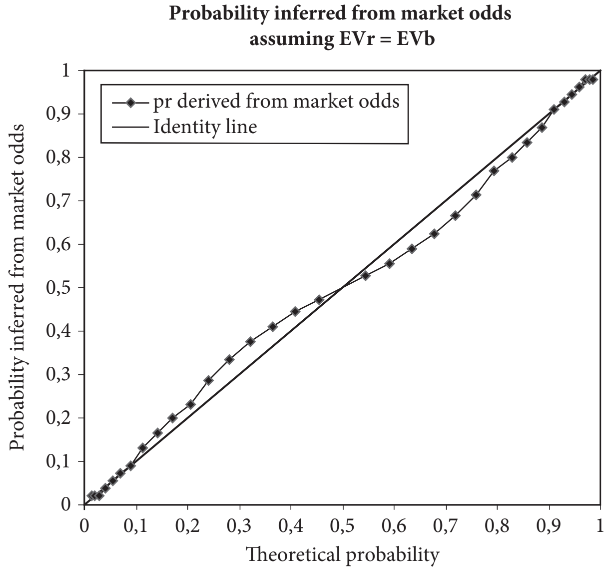


FIGURE 10.1 A scatterplot of the *theoretical probability against probability inferred from market odds* (table 10.6) together with the identity line.

3.2.3 *Efficiency of the General Odds Rule Assuming No Profitable Bets*

Nevertheless, as there are commissions, $t = 0.16$, when a bet takes place both bettors' expected values add up to a negative amount. Thus when analysing efficiency it is more convenient to ask for no possible profitable bets in the market. This implies that the expected value of a bet, both on the reds and on the blues, has to be lower than or equal to zero, and as shown in the following proposition, this implies that both equations (10.4) and (10.5) have to be fulfilled.

In a pelota betting market there are no profitable bets if and only if

$$p_r \leq \frac{O_R}{O_B(1-t)+O_R} \tag{10.4}$$

and

$$\frac{O_R(1-t)}{O_B+O_R(1-t)} \leq p_r. \tag{10.5}$$

Proof. The expected value of a bet on the reds lower than zero and the expected value of a bet on the blues lower than zero implies

$$p_r O_B(1-t) - (1-p)O_R \leq 0. \tag{10.6}$$

and

$$(1-p_r)O_R(1-t) - p_r O_B \leq 0, \tag{10.7}$$

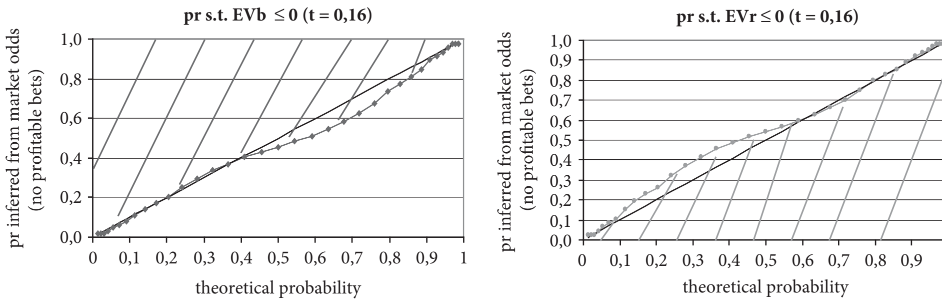


FIGURE 10.2 On the left, p_r assuming that the expected value of a bet on the reds is at most zero: equation (10.4). On the right, p_r assuming expected value of a bet on the blues is at most zero: equation (10.5).

respectively. Rearranging equations (10.6) and (10.7) we obtain equations (10.4) and (10.5).

We represent equations (10.4) and (10.5) first in two separate graphs, Figure 10.2, and then together, Figure 10.3.

Given a score, we know both the theoretical probability of winning for the reds (applying equation (10.1)) and the market odds (applying the general odds rule). Therefore in figure 10.2 (left), the horizontal axis shows the theoretical probability and the vertical axis shows the upper bound probability inferred from market odds (applying equation (10.4)). The marked area corresponds to the probabilities inferred from markets odds assuming that the expected value of a bet on the reds is lower than zero.

Given a score, we know both the theoretical probability of winning for the reds (applying equation (10.1)) and the market odds (applying the general odds rule). Therefore in figure 10.2 (right) the horizontal axis shows the theoretical probability of winning for the reds. The vertical axis shows the lower bound probability inferred from market odds (applying equation (10.5)). The marked area corresponds to the probabilities inferred from market odds assuming that the expected value of a bet on the blues is lower than zero. Because the conditions in both equations (10.4) and (10.5) are necessary for the market to have no possible profitable bets, the probability inferred from market odds should be at the intersection of both areas.

From figure 10.3 it can be seen that there are odds in the market at which profitable bets could be made. It can be seen that when the theoretical probability of the reds winning is {.68, .72, .75} it is profitable to bet on the reds: the odds are respectively {(100, 60), (100, 50), (100, 40)}. Symmetrically, when the theoretical probability of the reds winning is {.24, .28, .32}the odds are respectively {(40, 100), (50, 100), (60, 100)}, and it is profitable to bet on the blues, whose probabilities of winning are complementary. Overall, when the odds differ by 40, 50 or 60 euros, it is profitable to bet on the favorite.

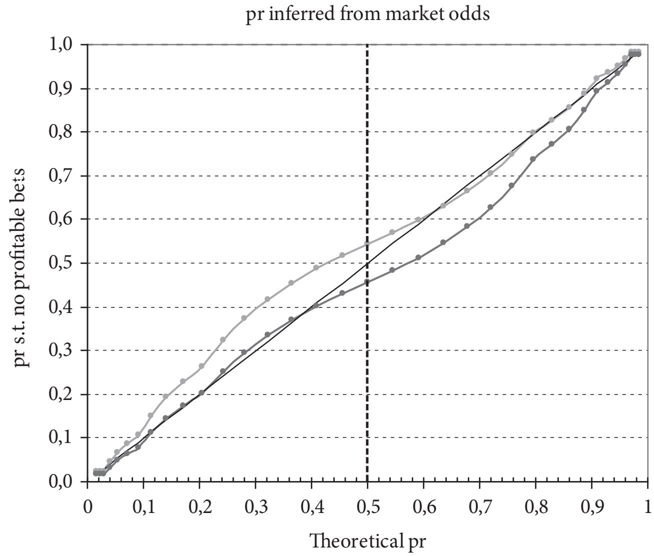


FIGURE 10.3 The probability inferred from market odds, such that no profitable bets are possible, has to be in the area between both curves. In this area, equations (10.4) and (10.5) are fulfilled.

Linda Woodland and Bill Woodland (1994) found deviation from efficiency in the baseball betting market, but their analysis failed to allow for profitable betting strategies when commissions are considered. Here we find profitable betting strategies taking commissions into account.

If the expected value of a bet on the reds must be equal to the expected value of a bet on the blues, equation (10.2), the probability inferred from market odds is just in the middle of the area between the two lines, a probability for which the expected value of a bet is negative. This is shown in figure 10.4.

4 Field Data Analysis

Llorente and Aizpurua (2008) found a theoretical explanation for the existence of this market in a world where both sides of the market are not different in beliefs and preferences. They found that for a bet to exist when bettors are rank-dependent expected utility maximizers, they have to be optimistic; that is to say, they underestimate the probability of the worst result occurring when they bet. Based on this theoretical explanation, the authors conducted a preliminary analysis of field data. Although the sample was small, their analysis was based on actual, observed bets. They found evidence of underestimation of the favorite's chances of winning, supporting the well-known long-shot bias.

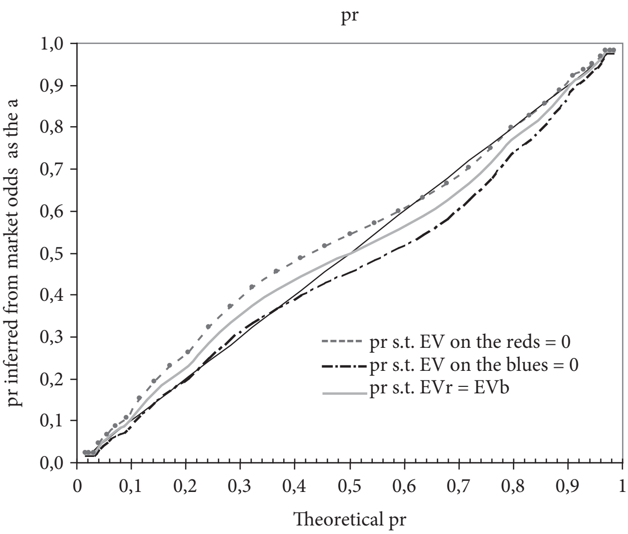


FIGURE 10.4 The solid line represents the probability inferred from market odds when bets have the same return, i.e. equation (10.2) is fulfilled. In this case the expected value of each bet is negative.

Here we explain the methodology they used. They collected data at the *frontón*, for each score, the odds and the approximate volume of bets. The odds are presented as pairs (O_f, O_l) where f denotes the favorite team and to which the higher number of the odds is associated, and l corresponds to the non-favorite or long-shot team.

For each set of odds, the total number of bets made at those odds was tabulated. Then it was counted for each given odds, the number of times the favorite team turned out to be the winner, and the number of times the favorite was defeated. Then frequencies were obtained (favorite wins and favorite loses).

For each given odds betting on l means accepting that O_l will be lost with some probability denoted by π_f^l and that $O_f(1-t)$ will be gained with probability $1-\pi_l^f$. A bettor who bets on l indicates that his subjective probability of the worse event (the favorite team wins) is $\pi_f^l \leq \frac{O_f(1-t)}{O_f(1-t)+O_l}$. This is the upper bound on the worse event for a bettor on l, with commission $t = 0.16$.

Similarly, for a given odds betting on f means accepting that O_f will be lost with some probability denoted by π_l^f and that $O_l(1-t)$ will be gained with probability $1-\pi_l^f$. A bettor who bets on f indicates that his subjective probability of the worse event (the non-favorite team wins) is $\pi_l^f \leq \frac{O_l(1-t)}{O_l(1-t)+O_f}$. This is the upper bound on the worse event for a bettor on f, with commission $t = 0.16$.

For any given odds the worse event for a bet on l is the favorite wins. And π_l^f is interpreted as the weight a bettor on l attaches to this frequency.

Similarly, for any given odds the worse event for a bettor on f is the favorite loses. And π_l^f is interpreted as the weight a bettor on f attaches to this frequency.

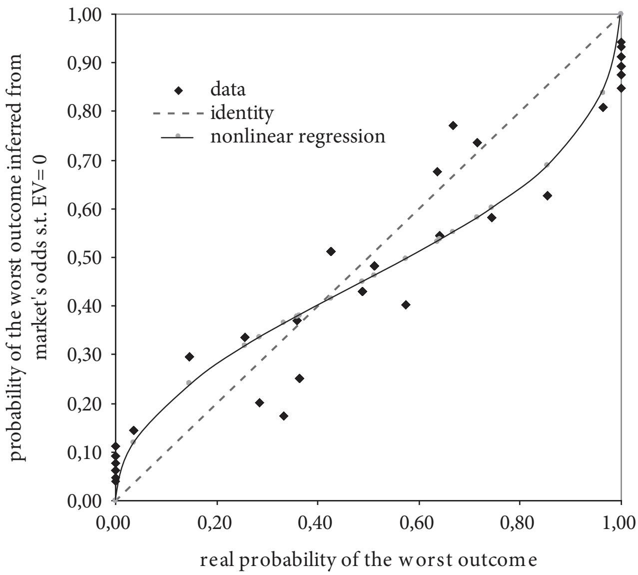


FIGURE 10.5

Source: Llorente and Aizpurua (2008)

The frequency with which an event occurs and the weights that different bettors place on that probability are represented in figure 10.5. High probabilities that the favorite will win are associated with weights that are lower than those probabilities; that is, the market odds underestimate the real chances that the favorite will win the match, and low probabilities that the favorite will lose are associated with weights that are higher than those probabilities, evidence of the well-known long-shot bias studied in Sauer (1998) and Woodland and Woodland (1994).

The following function is estimated

$$\pi = \frac{\delta(\text{“real } p\text{”})^{\gamma}}{\delta(\text{“real } p\text{”})^{\gamma} + (1 - \text{“real } p\text{”})^{\gamma}} \qquad R^2 = .839,$$

with $\delta = .84$ and $\gamma = .55$, both significantly lower than 1 at the 95 percent confidence level (figure 10.5). For this parameter value $\pi(\text{“real } p\text{”}) + \pi(1 - \text{“real } p\text{”}) < 1$ for all “real p.”

The result is a weighting function of optimistic bettors who overestimate low probabilities and underestimate high ones. In addition p^* (cross point with the identity line) is 0.38, very close to that obtained by Richard Gonzalez and George Wu (1999). These parameter values are similar to those obtained in the parameter-free elicitation of the probability weighting function studied by Mohammed Abdellaoui (2000). Their findings then are similar to those obtained in studies of weighting functions.

5 Hedging

Many bettors follow a strategy of hedging their bets. In this context, hedging refers to betting at two different moments during the game in such a way that the second bet intends to offset potential losses that may be incurred by the first. The bettor could then receive a payout independently of which team wins the match.

In subsection 2.2 we mentioned the Lane and Ziemba (2008) study on team jai alai, where the game is similar to pelota but the betting system is bookmaking. Differences between the betting system they analyzed and ours affect the meaning of what a consistent bet is. Therefore the arbitrage strategy they employ when the bookies' odds make a positive expected value possible due to the minimum payout on a bet makes no sense in pelota betting systems. On the other hand, when they use what they call risk arbitrage strategies, even though the betting game is different, the sport is similar and their idea of studying all feasible paths could help in the study of efficiency in the pelota betting system. Calling *G* the set of scenario game paths from (0,0) to the final outcome, they show the cardinality of G (262), which can be of interest in our study. Lane and Ziemba also discuss the applicability of jai alai study in traded options and warrants markets (270). Although we do not have details to present here, we actually are researching hedging strategies.

6 Summary and Concluding Remarks

There are two peculiarities that differentiate the pelota betting system from other well-known betting systems. Unlike pari-mutuel betting systems, the odds in a pelota market are established when the wager is made, and bets are arranged by middlemen. However, unlike conventional forms of bookmaking, for a bet to be placed one bettor bets on one team and another bettor bets on the other team; the middleman does not bet at all.

There is a similar betting system that has been implemented on the Internet, online betting exchanges. The betting system followed by betting exchanges is similar to the pelota betting system in that bettors can make as many bets as they want provided there is another bettor on the other side, and the market maker takes a percentage of the money as a commission. As we have noted, the main difference with the pelota betting system is the odds scale.

In this chapter, we have positioned the pelota betting system within the framework of the economic literature and we have presented the results obtained applying three different concepts of efficiency (defined in Sauer 1998) in this betting market: *constant return of bets, absence of profit opportunity, and equilibrium pricing function.* First, we conducted an examination of the market using orthodox methods, in particular, an analysis of the general odds rule followed in the pelota betting system under certain assumptions. Although these may be somewhat heroic assumptions, they have been invoked in studies of other wagering markets. Thus we start by assuming that bettors are expected value maximizers. A condition for equilibrium is that the expected value of a bet on the reds must be equal to the expected value of a bet on the blues because, if not, all bettors prefer to bet on the color with the higher return. We also discussed the probability inferred from the market odds assuming equal returns on each bet and found that there is a difference between the probability inferred from market odds and the theoretical probability of a team winning the match. Low probabilities are overestimated while high probabilities are underestimated. Therefore, in the pelota betting market we find evidence of the long-shot bias.

Second, in these markets there are commissions, so the equilibrium condition of equal returns on bets implies that each bet has a negative expected return. Thus it seems more convenient to introduce the less restrictive restriction of not allowing profitable bets in the market. We checked what the probability inferred from market odds is to satisfy this less restrictive restriction of no profitable bets, that is, to satisfy the condition of expected value of a bet lower than or at most equal to zero. We found that the probability inferred from market odds must be in an area satisfying equations (10.4) and (10.5). Under the assumption of no profitable bets there are odds at which subjective probabilities differ from real ones. Thus there are odds in the market at which profitable bets could be made; when the odds differ by 40, 50, or 60 euros, it is profitable to bet on the favorite. We would like to point out that the theoretically positive expected value bets that we found are relatively small. It is nonetheless an achievement to find any positive expectation bets in a game with a 16 percent commission.

Thirdly, the chapter also contains an empirical analysis of the efficiency of this market based on a standard assumption of equilibrium in an explicit model assuming identical bettors. Llorente and Aizpurua (2008), assuming all bettors are identical rank-dependent expected utility maximizers, found evidence in the market of underestimation of the favorite's chances of winning, supporting the well-known long-shot bias. More accurately, the effect found is a kind of long-shot bias. However, it is a bit different from the long-shot bias in horse racing. In horse racing, the long-shot bias gets more extreme as p gets smaller. The pelota bias found reaches a maximum in the medium long-shot range and then gets smaller again for extreme long shots.

Finally, we also provide some insights on hedging strategies.

Notes

1. Data obtained on-line from the paper "Origen y desarrollo de las distintas modalidades del juego de la pelota vasca" on the website of the Confederación Internacional del Juego de la Pelota; http://www.cijb.info.
2. I am grateful to a referee for showing me this connection and for his detailed explanation of it.

References

Abdellaoui, Mohammed. 2000. "Parameter-free elicitation of utility and probability weighting functions," *Management Science* 46(11):1497–1512.

Cain, M., D. Law, and D. A. Peel. 1992. *Greyhound racing: Further empirical evidence on market efficiency in a wagering market.* Aberystwyth Economic Research Papers No. 92-04, March. Aberystwyth: Department of Economics and Agricultural Economics, University College of Wales.

David, F.N. [1962] 1998. *Games, gods and gambling: A history of probability and statistical ideas.* Reprint, Mineola, N.Y.: Dover Publications.

Dowie, Jack. 1976. On the efficiency and equity of betting markets. *Economica* 43(170): 139–150.

Edelman, D., and L. Llorente Erviti. 2010. "A characterization of behavior on a betting exchange; Rank-dependent expected utility." Mimeo UPNA.

Edwards, A.W.F. 1982. Pascal and the problem of points. *International Statistical Review* 50:259–266.

Gonzalez, Richard, and George Wu. 1999. On the shape of the probability weighting function. *Cognitive Psychology* 38(1):129–166.

Güth, Werner, Loreto Llorente, and Anthony Ziegelmeyer. 2009. Inconsistent incomplete information: A betting experiment. *Jena Economic Research Papers* No. 2009–043. Jena, Germany: Friedrich-Schiller University and the Max Planck Institute of Economics.

Henery, Robert J. 1985. On the average probability of losing bets on horses with given starting price odds. *Journal of the Royal Statistical Society* 148(4):342–349.

Lane, Daniel, and William T. Ziemba. 2008. Arbitrage and risk arbitrage in team jai alai. In *Handbook of sports and lottery markets*, edited by Donald B. Hausch and William T. Ziemba. Amsterdam: Elsevier, 253–271.

Llorente, Loreto. 2007. A profitable strategy in the *pelota* betting market. In *Optimal play: Mathematical studies of games and gambling*, edited by Stewart N. Ethier and William R. Eadington. Reno: Institute for the Study of Gambling and Commercial Gaming, University of Nevada, 399–416.

Llorente, Loreto, and Josemari Aizpurua. 2008. A betting market: Description and a theoretical explanation of bets in pelota matches. *Theory and Decision* 64(2–3):421–446.

Marginson, David. 2010. Symposium—Market inefficiency in person-to-person betting: Examining "known loser" insider trading on the exchanges. *Southern Economic Journal* 76(4):932–952.

Quandt, Richard E. 1986. Betting and Equilibrium. *Quarterly Journal of Economics* 101(1): 201–207.

Sauer, Raymond D. 1998. The economics of wagering markets. *Journal of Economic Literature* 36(4):2021–2064.

Skiena, Stephen S. 1988. A fairer scoring system for jai-alai. *Interfaces* 18(6):35–41.

Smith, Michael A., David Paton, and Leighton Vaughan Williams. 2006. Market efficiency in person-to-person betting. *Economica* 73(292):673–689.

Thaler, Richard H., and William T. Ziemba. 1988. Parimutuel betting markets: Racetracks and lotteries. *Journal of Economic Perspectives* 2(2):161–174.

Vaughan Williams, Leighton, and David Paton. 1997. Why is there a favourite-longshot bias in British racetrack betting markets? *Economic Journal* 107(440):150–158.

Woodland, Linda M., and Bill M. Woodland. 1994. Market efficiency and the favorite-longshot bias: The baseball betting market. *Journal of Finance* 49(1):269–280.

CHAPTER 11

THE LURE OF THE PITCHER: HOW THE BASEBALL BETTING MARKET IS INFLUENCED BY ELITE STARTING PITCHERS

RODNEY J. PAUL, ANDREW P. WEINBACH, AND BRAD R. HUMPHREYS

1 Introduction

The starting pitcher plays a pivotal role in the sport of baseball. Pitchers are the key defensive player on the baseball diamond, and their abilities can drastically alter the outcome of games. Even poor teams can be bolstered by a great starting pitcher, and excellent offensive teams may lose games and championships due to weakness in starting pitching. Unlike many other team sports, a single player, the starting pitcher in a baseball game, can have a significant impact on the outcome of games. This may also affect outcomes in betting markets.

Although most starting position players on a team remain the same over the course of a 162-game baseball season, the starting pitcher differs from day to day; modern professional baseball teams generally use a five-man rotation, so a starting pitcher typically plays in every fifth game. Teams often play each other on consecutive days over the course of a three-game or four-game series. Although the same two teams may play each other several days in a row, the betting odds on these games vary considerably due to the starting pitcher matchups. Excellent starting pitchers playing against weaker opponents often command betting odds of -250, -300, or greater, implying that favorite bettors must lay out $250 or $300 to win $100. The favored team often differs over the course of a three- or four- game series between the same two teams due to the relative abilities of the starting pitchers in each game.

Starting pitchers receive their due respect on the betting boards in Las Vegas and on online betting outlets around the world. Unlike for other sports, when a baseball game is listed at a sportsbook, the starting pitcher for each team is identified along with the betting odds on the game's outcome, called *sides* in the Las Vegas sports betting jargon, and the over/under, a bet on the number of runs scored by both teams for the given game, called *totals* betting in Las Vegas. Even in other sports, such as hockey, where a key defensive player, the goalie, can have an important impact on game outcomes, the starting goalie in each game is not identified on the betting board. In baseball, the names of the starting pitchers appear clearly for all bettors to see. Typically, sports bookmakers allow bettors to place a bet on "listed pitchers" (or just "pitchers") compared to the "action" option on games. If a bettor chooses to place a wager on "listed pitchers" and either or both of the starting pitchers changes, the full bet is returned and there is no betting action on the game. In the "action" option, the bettor places a wager on the team no matter which starting pitchers appear in the game. Starting pitchers in baseball games occupy a unique position in betting markets.

Evidence shows that sports betting markets are weak-form efficient in that betting odds reflect all public information about game outcomes (Sauer 1998). In baseball betting markets, the posted odds should reflect the expected effect of the starting pitchers on game outcomes. However, recent research on the economics of betting markets has moved beyond the analysis of prices and game outcomes to examine variation in the volume of bets placed on games and betting on the total number of points scored in games, called over/under betting. The goals of this new line of research are to determine the extent to which gambling has both financial and consumption motives, as in the model developed by John Conlisk (1993), and to look for evidence of behavioral biases in the decisions of bettors and bookmakers. In this chapter we develop evidence that starting pitchers affect the volume of bets placed on games, the amount of money bet on favored teams, and the amount of money bet on the under, the proposition that a game will be relatively low-scoring. We find that the volume of bets placed on games, the fraction of bets placed on the favored team, and the fraction of bets placed on the under proposition can all be explained by the presence of high-profile starting pitchers in a game, other things being equal. These results indicate that betting on baseball games can be motivated by both financial and consumption benefits and that behavioral biases exist in this market.

2 Previous Research on Baseball Betting Markets

The focus of previous research on baseball betting markets has been the favorite-longshot bias or the reverse favorite-longshot bias. Evidence of a favorite-longshot bias, where favorites are underbet and longshots overbet by gamblers, has been found in some research on horse race betting. Linda Woodland and Bill Woodland (1994) showed that a reverse of this bias exists in baseball, where favorite odds are too high, resulting in underdogs winning more than implied by market efficiency. J. M. Gandar et al. (2002) showed that the original analysis by Woodland and Woodland (1994) did not properly account for a unit bet on favorites and underdogs and illustrated that the reverse favorite-longshot bias was not statistically significant for all underdogs in baseball, only for slight underdogs and home underdogs.

The contribution of this chapter is the inclusion of starting pitchers in econometric estimation of a bet volume equation in baseball, which allows for the analysis of bettor behavior in a more detailed way than simply observing the betting market odds on games. Our approach is important because of recent findings rejecting the long-held belief in the balanced book hypothesis. The balanced book hypothesis assumes that sportsbooks set point spreads to balance the betting action evenly on each side of a proposition, favorites and underdogs in sides betting, over and under in totals betting. With respect to fixed-odd betting, used in baseball, hockey, soccer, and other sports, the balanced book hypothesis predicts that odds will be set to proportionally balance the betting dollars across outcomes.

The balanced book hypothesis has recently been challenged by Steven Levitt (2004) in a study of the NFL betting market through a limited betting market tournament. He concluded that sportsbooks do not balance the betting action, but set prices to maximize profits by exploiting known bettor biases. The result concerning the rejection of the balanced book hypothesis was confirmed in the study of actual online sportsbooks in the NFL (Paul and Weinbach 2007), the NBA (Paul and Weinbach 2008), college football (Paul and Weinbach 2009), the NHL (Paul and Weinbach 2012b), and Major League Baseball (Paul and Weinbach 2012a). Although the balanced book hypothesis could clearly be rejected, with bettors consistently favoring favorites (road favorites in particular) in sides betting and overs in totals betting, the alternative hypothesis of Levitt (2004) was only supported in the NFL. In other words, only in the NFL did the sportsbook appear to price to exploit known bettor biases (which are clear across the sample of all sports) by setting prices where underdogs won more often than favorites. In all of the other sports studied, despite the consistent imbalance of bets, win percentages on favorites in sides betting were found to be statistically indistinguishable from 50percent. This result implies that sportsbooks price as a forecast of game outcomes despite consistent and clearly predictable bettor biases. This behavior by the sports bookmaker likely prevents informed bettors from entering the market (as they cannot exploit simple profitable betting strategies) and still earns the sportsbook a profit from the commission collected on losing bets in the long run.

3 Data Description

This chapter uses detailed data on betting percentages and volume to examine the impact of elite starting pitchers on the baseball betting market in the 2009 MLB season. The data were purchased from Sports Insights, a company that collects detailed betting market data from four online bookmakers located in the Caribbean. These online bookmakers have been in operation for a number of years and are well established in the online betting industry. They take bets from a large number of customers and can be considered representative of the online betting market. Sports Insights collects a large amount of game-specific data on the baseball betting market, including odds, game outcomes, and bet volume, on both sides and totals betting

Sports Insights collects betting market data from four online sports bookmakers: 5Dimes.com, a licensed, bonded online sports bookmaker based in Costa Rica in operation since 1998; BetUS.com, a licensed and bonded sports bookmaker recognized in both Costa Rica and Canada in operation since 1994; CaribSports.com, a licensed online sports bookmaker based in Belize in operation since 1997; and Sportsbook.com, a licensed online sports bookmaker and casino based in Costa Rica since 1996. These four sports bookmakers handle a large number of bets on a variety of professional and amateur sports worldwide.

Table 11.1 contains summary statistics for the three dependent variables, the volume of bets, the percentage bet on the favored team in the sides market, and the percentage bet on the under in the totals market, as well as the explanatory variable of interest, games in which one or both of the starting pitchers was one of the five top vote-getters in the 2008 and 2009 Cy Young Award voting. We have data on betting market outcomes for 2,372 MLB games played in the 2009 regular season. This does not include every game, since volume data are not available for every regular season game. The average number of bets on baseball games in the sample was almost 11,700. On average, more bets were placed on the favored team in these games, and more bets were placed on the over in the total market.

High-quality starting pitchers can be identified in many ways. The primary methods include performance-based identification methods based on common statistics, such as earned run average (ERA) and wins, among others, and on voting outcomes, such as the number of All Star votes players receive from fans. The simplest and most straightforward method to identify an elite pitcher, in our opinion, is to use pitchers who received votes for the Cy Young Award,[1] the annual award for the best pitcher in each league in MLB. Each year, the top five pitchers receiving votes for the Cy Young Award in each league are identified at the end of the season. Baseball writers, a group widely viewed as experts on baseball, vote for the Cy Young Award. The top five

Table 11.1 Summary Statistics

Variable	Mean	Std. Dev.	Min.	Max.
Number of bets placed on game	11692	5630	826	115409
Percentage bet on favorite	64	16	16	94
Percentage bet on under	43	13	7	87
Games started by Cy Young candidates	0.14	0.34	0	1

vote-getters can be objectively viewed as high-quality pitchers. Therefore, to keep the identification of elite pitchers as simple and objective as possible, we created a dummy variable identifying games started by these pitchers in the 2009 season; 14 percent of the games in our sample involved one or more of these starting pitchers. Note that rotations tend not to change once set, so there is little possibility that these starting pitchers would be assigned in a way to make them correlated with the equation error term. Once a starting rotation is set, the starting pitchers generally pitch every fourth or fifth game throughout the season.

We examine three key elements of the betting market for Major League Baseball. First, we investigate the role of the elite pitcher in the determination of betting volume. If many bettors enjoy wagering on the best pitchers, betting volume may reflect these preferences. If baseball betting has a major consumption element and fans follow the best pitchers, the games involving the elite pitchers may also be the games that attract the most interest from bettors.

Second, we examine the role of the elite pitchers in relation to the percentage bet on favorites and underdogs. Given that the baseball betting market has been shown to be imbalanced (not proportional to sportsbook set odds), there is the distinct possibility that elite starting pitchers play a key role in betting percentages. The best pitchers may attract the most betting action, which will be reflected in both betting volume and in the percentage bet on the team with the elite pitcher.

Third, we examine the role of the elite pitcher in the totals market. Typically, in totals betting, the over has been shown to be a much more popular proposition than the under, perhaps because fans find high-scoring games more entertaining. Across many different sports, bettors, like fans, prefer more scoring to less. This is reflected by a large percentage of the bets accumulating on the over in totals betting. Given that the starting pitcher plays a pivotal role in baseball and in the wagering market, the baseball betting market may provide an instance where the under is preferred to the over. If baseball bettors like to wager on the best pitchers, they may also wish to wager on the under, because when the best pitchers perform well they prevent their opponent from scoring runs, reducing the total number of runs scored in the game. If found, this result would not necessarily imply that baseball bettors have different utility functions and prefer less scoring, but it does show the role of consumption in betting markets, as marquee pitchers lead bettors to hope that their teams will win and that the pitchers will shut down their opponents, which (in the minds of bettors) will lead to more unders in these games.

All three of these empirical issues extend the existing research on gambling markets. The efficient markets hypothesis, the balanced book hypothesis, and the existence of behavioral biases in betting markets are all addressed in this study. In short, we further explore how popular baseball players, in this case elite starting pitchers, affect sports betting markets. We hope to further illustrate the role that consumption-based gambling motives play in this market.

4 Empirical Analysis

We examine the relationship between variation in three betting market outcomes, bet volume, the fraction of bets placed on the favored team, and the fraction of bets placed on the under, as well as variation in the quality of the starting pitchers in the game, controlling for other factors that affect betting market outcomes. We estimate regression models of the form

$$OUTCOME_{ijt} = b_1 GAME_{ijt} + b_2 TEAM_{ijt} + b_3 PITCHER_{ijt} + e_{ijt}, \qquad (11.1)$$

where $OUTCOME_{ij}$ is one of three betting market outcomes for a game involving home team i and visiting team j on date t; $GAME_{ijt}$ is a vector of game characteristics, is a vector of team characteristics, is a vector of variables that capture the quality of starting pitchers in the game between home team i and visiting team j on date t; and e_{ijt} is an unobservable random variable capturing all other factors that affect the betting market outcomes. We assume that e_{ijt} is an independent and identically distributed random variable with zero mean and potentially nonconstant variance. Because e_{ijt} may be heteroskedastic, we use the White-Huber "sandwich" correction to adjust the standard errors for potential heteroskedasticity. b_1, b_2, and b_3 are vectors of unknown parameters to be estimated. We estimated these parameters using the Ordinary Least Squares estimator.

The Impact of Starting Pitchers on Betting Volume

The first set of regression results uses the number of bets placed on each game as the dependent variable in the regression model described by equation (11.1). Again, this variable is the total number of bets placed at the four online sports bookmakers tracked by Sports Insights. Independent variables include an intercept, the months of the year, the days of the week, dummy variables for elite pitchers (both visitor and home), the absolute value of the odds on the game, a dummy for road favorites, and the win percentages of both the home and road teams going into the game.

The months of the baseball season are included to control for any seasonality in betting volume. This seasonality could manifest itself as bettor excitement early in the year, effects of the start of football betting on baseball betting volume at the end of the season, or other reasons. The omitted monthly category is the first month of the season, April.

Weekdays are also included as dummy variables to account for possible daily variation in betting volume. The dependent variable is betting volume per game, so there could be impact where bettors may place more wagers on days when there are fewer games (especially when we consider baseball gambling as purely an activity of consumption). Mondays and Thursdays typically have fewer baseball games, as these serve as travel days for teams. If bettors choose to wager on baseball as a form of consumption, due to fewer games on these days, individual game betting volume may increase on Mondays and Thursdays. Weekend effects also are possible, as more bettors may engage in gambling on the weekend as compared to weekdays. Tuesday is the omitted category of the days of the week.

The "Cy Young Voted Pitcher" variable (both home and road) is the proxy for the elite pitchers as mentioned in the introduction. The top five starting pitchers who received votes for the Cy Young Award in each league (NL and AL) are given a value of one, while all other pitchers are given a value of zero. The first regression specification uses pitchers from the previous season (2008) who met this criterion, while the second uses the current year to create the dummy variable, and the last specification combines the two years (if a pitcher appeared once or twice on the list he is given a value of one). We will use the previous season dummy as the main point of discussion of the results, but if one believes that elite pitchers can easily be identified during the season, new elite pitchers are likely to receive attention as the season progresses. Although this is a simple version of identifying an elite pitcher, we believe it will give a good representation of the importance of the pitcher in the minds of bettors of Major League Baseball.

The absolute value of the odds on the game is also included as an explanatory variable. This variable has been shown to affect the percentage bet on the favorite in Major League Baseball (Paul and Weinbach 2012a) in a positive and significant manner and is likely to affect betting volume as well. We use the absolute value of the betting odds due to the presence of home favorites and road favorites, which additionally is addressed by the inclusion of a road favorite dummy variable. Road favorites have been shown to be very popular bets across virtually all sports (due to a likely underestimation of the home field advantage—it appears that bettors are getting a "value" by betting on the good road team at a lower point spread or odds). With respect to betting volume, it is likely that road favorites attract a greater number of bets. The absolute value of the odds has a positive expected sign (bigger favorites are more popular betting propositions, as they likely are the best teams) and the road favorite also has a positive expected sign.

The winning percentages of both the home team and the road team also are included as independent variables in the regression model on betting volume. Assuming consumption plays a pivotal role in sports betting, good teams are likely to attract more bets, thereby increasing betting volume when they play. This may partially be captured by the absolute value of the point spread, but it is important to remember that low betting odds on the favorite could be a function of two similar teams of any quality playing in a game. We expect games between two good teams to attract more betting attention than games between two bad teams; therefore we expect the sign on both of the win percentage variables to be positive.

Table 11.2 presents the regression results for the betting volume model. *T*-statistics are shown in parentheses below the parameter estimates. The month of the year dummy variables are relative to the omitted month, April. The month of May saw a positive and significant jump in volume, while the later months of the year saw a significant

Table 11.2 Baseball Bet Volume Regression Results, 2009 Season Dependent Variable: Betting Volume per Game

Variable	Pitching Dummies—Previous Season (2008)	Pitching Dummies—within Season (2009)	Pitching Dummies—Both Seasons (2008–2009)
Intercept	−1744*	−1782*	−1694*
	(−1.74)	(−1.77)	(−1.69)
May	1847***	1839***	1827***
	(4.74)	(4.71)	(4.70)
June	230	209	196
	(0.669)	(0.60)	(0.57)
July	−693**	−690**	−693**
	(−2.07)	(−2.05)	(−2.07)
August	−1728***	−1746***	−1741***
	(−4.00)	(−4.03)	(−4.03)
September	−4358***	−4387***	−4390***
	(−10.53)	(−10.56)	(−10.63)
October	−7971***	−8006***	−8000***
	(−15.59)	(−15.74)	(−15.68)
Sunday	293	288	298
	(0.664)	(0.65)	(0.67)
Monday	863*	860*	871*
	(1.86)	(1.86)	(1.88)
Wednesday	−20.74	−55.46	−29.12
	(−0.04)	(−0.10)	(−0.05)
Thursday	923**	937**	917**
	(2.05)	(2.09)	(2.05)
Friday	419	403	416
	(0.98)	(0.94)	(0.97)
Saturday	796*	792*	793*
	(1.75)	(1.74)	(1.75)
Cy Young voted	1275***	1327***	1342***
pitcher—Visitor	(2.77)	(3.60)	(3.99)
Cy Young voted	1172.8**	770**	900***
pitcher—Home	(2.55)	(2.04)	(2.67)
Absolute value—Betting	41.38***	41.57***	41.21***
odds	(12.14)	(11.98)	(11.91)
Road favorite dummy	1198***	1129***	1116***
variable	(4.98)	(4.64)	(4.60)
Home win percentage	7832***	7804***	7742***
going into game	(7.08)	(7.01)	(6.98)
Road win percentage	7636***	7723***	7576***
going into game	(7.12)	(7.18)	(7.07)
R^2	0.222	0.222	0.225

decrease in the volume of bets on baseball. Through August, September, and October baseball betting volume dropped considerably as the regular season came to a close. Likely reasons for this are the start of football season and some teams being eliminated from playoff contention, which could lessen interest in those games (assuming betting as a consumption activity).

In terms of daily betting patterns, Monday, Thursday, and Saturday all saw significant increases in betting volume per game compared to the omitted day, Tuesday. Monday and Thursday likely saw significant increases due to there being fewer games per day, as these days often serve as travel days for teams going from one city to another. The increase on Saturday is likely due to the opportunity cost of time on the weekend with more people being willing to place wagers (and likely watch) baseball games on Saturday afternoons and evenings.

The elite pitchers in the league had a positive and significant effect on betting volume. Both the home and road pitcher were shown to positively affect volume. Elite pitchers at home increased betting volume by over 1,100 bets per game, while elite pitchers on the road increased betting volume by over 1,200 bets per game. Both results were found to be significant at the 1 percent level. This implies that many bettors are wagering on the best pitchers, leading to considerable increases in volume when these athletes take the mound.

Similar to what was shown previously for baseball in Paul and Weinbach (2012a), the absolute value of the odds on the game was shown to have a positive and significant effect on betting volume. Bigger favorites were shown to generate more bettor interest in the game. In addition, road favorites were shown to be more popular with bettors across the board, as games involving a road favorite led to well over a thousand additional bets on the game. Records of the teams also played a significant role, as bettors placed more wagers on games between better quality opponents. Both the home and road win percentages of the teams (heading into that day's game) were shown to have a positive and significant effect on betting volume.

Overall, these results underscore the major role that consumption plays when betting on baseball. Quality of the teams, days of the week, months of the year, and elite pitchers all attract more bettors to wager on baseball games.

The Impact of Starting Pitchers on Percentage Bet on the Favorite

To illustrate the effect of elite pitchers on the sides (betting on a particular team) market, we use the percentage bet on the favorite as the dependent variable in this regression model. To distinguish between home and road favorites, we run two separate regressions, one consisting entirely of home favorites, the other of road favorites. Results appear on table 11.3.

The independent variables are closely related to those included in the regression model on betting volume, with month of the year dummies, day of the week dummies,

Table 11.3 Sides Regressions—Home Favorites and Road Favorites—2009 Season Dependent Variable: Percentage Bet on the Favorite

Variable	Home Favorites	Road Favorites
Intercept	−4.045	14.60*
	(−0.92)	(1.79)
May	0.565	−0.201
	(0.49)	(−0.10)
June	0.477	−1.225
	(0.365)	(−0.64)
July	2.947	1.835
	(2.43)	(0.94)
August	1.384	−1.772
	(1.16)	(−0.99)
September	1.507	0.616
	(1.30)	(0.35)
October	−0.559	−10.19***
	(−0.32)	(−3.37)
Sunday	−0.036	1.623
	(−0.03)	(1.10)
Monday	0.835	−0.298
	(0.669)	(−0.171)
Wednesday	−1.142	1.616
	(−0.95)	(1.055)
Thursday	−1.058	1.817
	(−0.85)	(1.18)
Friday	1.348	0.813
	(1.21)	(0.53)
Saturday	0.168	1.433
	(0.15)	(0.82)
Cy Young voted pitcher—Visitor	−6.537***	1.975
	(−2.96)	(1.42)
Cy Young voted pitcher—Home	0.214	−11.55***
	(0.18)	(−2.68)
Percentage bet on favorite implied by closing odds	1.250***	1.091***
	(20.71)	(8.29)
Home record	17.88***	−26.47***
	(3.65)	(−4.97)
Road record	−35.17***	12.34**
	(−8.81)	(1.99)
R^2	0.402	0.285

Note: Pitcher dummy variables are based on Cy Young voting from previous season (2008) only. Other results are available upon request from the authors.

Cy Young vote-getting pitcher dummies, and team win percentage variables included for most of the same reasons as described above. We include the days and months to account for fluctuations that occur over time during the week and the season; team win percentage variables are likely to influence which teams are the most popular in the betting market, as good teams are more likely than poor teams to attract betting action (based on consumption value and purpose in the minds of bettors), and we expect the elite pitcher dummy variables to illustrate their impact on the percentage bet on the favored team.

The main difference between this regression model and the one in the previous section is that an independent variable is constructed to represent the percentage bet on the favorite that is implied by the closing betting market odds. Since the baseball market operates in an odds-based format, rather than through a point spread, we expect the favorite (and underdog) to attract betting percentages reflective of the odds on the game rather than assume a 50/50 split under the balanced book hypothesis in point spread–based markets. Therefore, based on the midpoint of favorite and underdog odds, we can determine the percentage that would be expected to be bet on the favorite (and underdog) if the sportsbook was behaving in a way to avoid being an active participant in the wager. This variable, constructed from the closing game odds, is then included as an independent variable in the regression. A value of one would indicate that the sportsbook is pricing to clear the market, while a value greater than one would suggest that the sportsbook is willing to accept a higher percentage of bets on the more popular side of the betting proposition, the favorite.

In relation to the sides regressions for home favorites and road favorites, the months of the year and the days of the week had little to do with the percentage bet on favorites (except for the few games in October with respect to road favorites). Not surprisingly, winning percentages of the teams played a statistically significant role in the percentage bet on favorites, with the signs on the variable being as expected (positive for the favored team and negative for the underdog). The percentage that should have been bet on the favorite, derived from the closing odds on the game, was found to have a positive and significant effect on the percentage bet on the favorite, as expected. The actual percentage bet, however, was found to be greater than the expected percentage bet (based on the odds), as the coefficient on these variables was found to be greater than one in both regressions.

Elite starting pitchers also played an important role in the percentage bet on the favorite. The significant results, however, were found with respect to the underdog pitcher. If an elite pitcher was an underdog in the game, bettors wagered a considerably greater amount in these contests. The percentage bet on home favorites dropped by 6.5 percent when an elite visiting pitcher was on the mound. In the rarer case where the elite home pitcher was the underdog, the percentage bet on the favorite fell by more than 11 percent. Bettors appear to be quite sensitive to situations where elite pitchers are underdogs. These situations appear to entice bettors, who are given the opportunity to wager on an excellent pitcher with the added bonus of receiving odds back in their favor if the underdog team wins the game.

The Role of the Pitcher in the Totals (Over/Under) Market

As mentioned above, the presence of an elite starting pitcher in a game likely affects the totals betting market. In the totals market, commonly referred to as over/under betting, bettors wager on the total number of runs that will be scored in a game. Bettors can place a wager of either over (more runs scored by both teams) or under (fewer runs scored by both teams) relative to the posted total.

To investigate this, we used the percentage wagered on the under as the dependent variable in the regression model defined by equation (11.1). The results are shown in table 11.4. Again, we used monthly dummies, weekday dummies, win percentages of both teams, and the elite pitchers dummy variables in the regression. The independent variable that differs in this regression model is the inclusion of the total (as opposed to the odds on the game). We used only the total itself (not the odds adjustment—which at most is -130 in either direction) in the regression to attempt to determine whether bettors are more willing to wager on the over in expected higher-scoring games (higher totals). This tendency has been shown to exist in the NFL, college football, and the NBA in point spread markets as well as in the NHL and previous studies of Major League Baseball.

The dummies to account for the pitchers who received Cy Young votes are expected to have a positive and significant effect on the percentage wagered on the under if fans like to bet on the best pitchers to prevent the other team from scoring runs. The pitcher is one of the few cases in North American sports where a notable player's key role is defensive (in the case of pitchers—to prevent runs scored). Therefore, we believe that many bettors who wager on specific good pitchers are also likely to wager on the under in these games, as they are cheering for the pitcher to win the game and keep the other team from scoring. If this premise is true, the elite pitchers will have a positive and significant effect on the percentage bet on the under.

From table 11.4, the percentage wagered on the under was shown to respond to the months of the season, with fewer bets placed at the beginning of the season (the omitted month of April). Days of the week did not see much of an effect, except for a statistically significant result on Sunday, where more wagers were placed on unders. Team records were shown to not have a significant effect on the percentage bet on the under.

The total itself and the presence of an elite pitcher significantly affect the percentage of bets placed on the under. Higher totals attract fewer wagers on the under, as bettors appear more willing to bet the over when the total is high (when the expected amount of scoring is already high). For each additional point of the total, the percentage bet on the under dropped by around 4 percent. Elite pitchers also played an important role in the totals market, as the best pitchers were shown to increase the percentage bet on the under. Whether on the road or at home, an elite starting pitcher led to a 4 percent increase in the percentage bet on the under (8 percent if two elite pitchers were starting).

Table 11.4 Percentage Bet on the Under Regression—2009 Season Dependent Variable: Percentage Bet on the Under

Variable	Pitching Dummies—Previous Season (2008)	Pitching Dummies—within Season (2009)	Pitching Dummies—Both Seasons (2008–2009)
Intercept	73.44***	69.49***	68.65***
	(21.19)	(19.18)	(19.36)
May	3.792***	3.780***	3.702***
	(4.08)	(4.08)	(4.019)
June	10.82***	10.88***	10.84***
	(11.31)	(11.39)	(11.39)
July	13.39***	13.49***	13.48***
	(13.39)	(14.79)	(14.86)
August	4.771***	4.831***	4.838***
	(5.49)	(5.59)	(5.63)
September	10.81***	10.86***	10.84***
	(11.77)	(11.91)	(11.92)
October	7.589***	7.699***	7.668***
	(3.77)	(3.85)	(3.91)
Sunday	2.653**	2.045**	2.098**
	(2.41)	(2.39)	(2.46)
Monday	1.155	1.143	1.175
	(1.21)	(1.20)	(1.23)
Wednesday	1.251	1.118	1.217
	(1.40)	(1.26)	(1.37)
Thursday	−0.361	−0.317	−0.416
	(−0.38)	(−0.34)	(−0.44)
Friday	0.621	0.480	0.554
	(0.72)	(0.55)	(0.646)
Saturday	0.054	0.018	−0.003
	(0.06)	(0.02)	(−0.01)
Cy Young voted	4.28***	4.430***	4.678***
pitcher—Visitor	(3.42)	(4.12)	(5.09)
Cy Young voted	3.699***	4.246***	4.359***
pitcher—Home	(3.17)	(4.14)	(4.99)
Total	−4.589***	−4.158***	−4.050***
	(−15.23)	(−12.78)	(−12.80)
Home record going into	1.608	1.343	1.068
game	(0.66)	(0.55)	(0.44)
Road record going into	2.234	2.279	1.782
game	(0.93)	(2.35)	(0.75)
R^2	0.252	0.256	0.262

The result of bettors favoring the under appears to be solely a function of the elite pitchers. Although the percentage bet on the over usually exceeds 50 percent by a considerable margin in this sample, the best pitchers attract a higher percentage of bets on the under when they are starting. This likely again reflects consumption value to bettors, as many fans of baseball tend to follow the best pitchers. Bettors may not only wager on these best pitchers when they pitch, but they also appear to have the tendency to wager on the under.

This result appears to contradict findings observed in other sports betting markets, where bettors appear to clearly favor the over to the under. This likely stems from bettor preferences for more points scored compared to fewer points scored. That factor is still present in the baseball totals market (as evidenced by the negative and significant effect of the total on the percentage bet on the under), but the pitchers offer an interesting exception to this rule. When fans/bettors are given an opportunity to watch or follow a great pitcher, they likely hope this pitcher performs well. When he does, this leads to very few runs for the opposing team. Bettors express their desire to see this by placing wagers on the under in hopes of seeing excellent pitching from these stars. When pitching stars are not present, however, baseball bettors simply revert back to hoping to see more scoring, as evidenced by the percentage bet on the over rising with each point of the posted total.

5 Conclusions

We investigated the role of elite starting pitchers in the Major League Baseball betting market. Although there are numerous ways to identify an elite starting pitcher we used a simple explanatory variable to analyze the effect of elite starting pitchers on betting market outcome: pitchers who received votes for the Cy Young Award, given to the best pitcher in each league in MLB, were defined as elite pitchers.

Detailed data on outcomes in the baseball betting market, including data on the volume of bets placed on games, the percentage of bets wagered on favorites, and the percentage of bets wagered on the under (in the totals market) were analyzed using regression models. In terms of bet volume, elite pitchers generate a significant increase in the number of bets on a baseball game. Games involving elite pitchers led to an increase of around one thousand bets, on average, on a game. In addition to the elite pitchers, favorites (in particular road favorites) and the win percentages of the teams attract a statistically significant increase in bets. Overall, baseball bettors appear to bet heavily on games involving the best teams and the best pitchers. Profit-maximizing bettors would wager on games with the highest expected return; utility-maximizing bettors would bet on the games with the highest consumption value. Unless the betting odds offered by sportsbooks on games involving the best teams, and the best pitchers, systematically have a higher expected return to bettors, this pattern suggests that consumption-based gambling is a major factor in sports wagering markets.

Appendix Cy Young Award Voting—AL and NL—2008 and 2009

2008 AL Cy Young	2008 NL Cy Young	2009 AL Cy Young	2009 NL Cy Young
Cliff Lee (Cleveland)	Tim Lincecum (San Francisco)	Zack Greinke (Kansas City)	Tim Lincecum (San Francisco)
Roy Halladay (Toronto)	Brandon Webb (Arizona)	Felix Hernandez (Seattle)	Chris Carpenter (St. Louis)
Francisco Rodriguez (LA Angels)	Johan Santana (Mets)	Justin Verlander (Detroit)	Adam Wainwright (St. Louis)
Daisuke Matsuzaka (Boston)	Brad Lidge (Philadelphia)	CC Sabathia (Yankees)	Javier Vasquez (Atlanta)
Mariano Rivera (Yankees)	CC Sabathia (Milwaukee)	Roy Halladay (Toronto)	Dan Haren (Arizona)

Based on our analysis of variation in the percentage bet on the favorite in baseball games, elite pitchers were also shown to have a significant effect on this wagering. Although bettors prefer the best teams, elite pitchers on the underdog team attracted a significant number of bets. When an elite pitcher appears as an underdog, many bettors jump at the opportunity to wager on the best pitchers with the possibility of earning odds back in their favor if their wager is successful. The balanced book model suggests that sports bookmakers set odds to equalize the faction of bets on either side of a game, given all information about that game. Since bettors prefer to bet on elite pitchers, even if they pitch for an underdog team, this unbalanced betting could indicate the presence of bettors who want to bet on specific pitchers no matter what odds are set on the game. This would represent a type of behavioral bias in this market.

Elite pitchers also played a key role in the totals market in baseball. Although Major League Baseball bettors enjoy betting the over in far greater numbers than the under, likely due to a preference for scoring compared to a lack of scoring, the elite pitchers serve as a key exception. In games involving elite pitchers, a statistically significant increase in bets on the under was observed. The efficient market hypothesis suggests that the total set by sportsbooks should reflect all available information about the game, including the quality of the starting pitchers. When the best pitchers are pitching, bettors appear to want to see these pitchers perform well. If these pitchers are successful, few runs are scored for the opposing team. This leads them to wager on the under, revealing a preference for the lack of scoring compared to the normal preference for scoring. This result is likely not due to a desire to see defensive ability to the star power of the elite pitchers and the under being associated with a winning outcome on the part of their performance. Again, this could indicate the presence of bettors who would bet the under when elite starting pitchers appear in a game, no matter what total is set by the sportsbook. This could also be explained by behavioral biases in this market. It would be useful to explore these possible explanations.

Notes

1. The pitchers in the top five in votes for the Cy Young Award are listed in the appendix.

References

Conlisk, John. 1993. The utility of gambling. *Journal of Risk and Uncertainty* 6:255–275.

Gandar, J. M., Richard A. Zuber, R. S. Johnson, and W. Dare. 2002. Re-examining the betting market on Major League Baseball Games: Is there a reverse favorite-longshot bias? *Applied Economics* 34(10):1309–1317.

Levitt, Steven D. 2004. Why are gambling markets organised so differently from financial markets? *The Economic Journal* 114:223–246.

Paul, Rodney J., and Andrew P. Weinbach. 2007. Does Sportsbook.com set pointspreads to maximize profits? Tests of the Levitt Model of sportsbook behavior. *Journal of Prediction Markets* 1(3):209–218.

———. 2008. Price setting in the NBA gambling market: Tests of the Levitt Model of sportsbook behavior. *International Journal of Sport Finance* 3(3):137–145.

———. 2009. Sportsbook behavior in the NCAA football betting market: Tests of the traditional and Levitt models of sportsbook behavior. *Journal of Prediction Markets* 3(2):21–37.

———. 2012a. Behavioral biases and sportsbook pricing in Major League Baseball. In *The Oxford handbook of sports economics: Volume 2, Economics through sports*, edited by Stephen Shmanske and Leo H. Kahane. New York: Oxford University Press, 302–320.

———. 2012b. Sportsbook pricing and the behavioral biases of bettors in the NHL. *Journal of Economics and Finance* 36(1):123–135.

Sauer, Raymond D. 1998. The economics of wagering markets. *Journal of Economic Literature* 36:2021–2064.

Woodland, Linda M., and Bill M. Woodland. 1994. Market efficiency and the favorite-longshot bias: The baseball betting market. *Journal of Finance* 49(1):269–279.

CHAPTER 12

INFORMATION EFFICIENCY IN HIGH-FREQUENCY BETTING MARKETS

J. JAMES READE AND JOHN GODDARD

1 Introduction

The betting industry has been transformed by the Internet. Growth of person-to-person betting, mediated through online betting exchanges, has been a key element of this transformation (for a review see Smith and Vaughan Williams 2008). Betting exchanges enable traders to either back (buy) or lay (sell) bets on a wide range of sporting events. Betfair, launched in June 2000 and floated on the London Stock Exchange in October 2010 with a market valuation of £1.4 billion, has established its position as the dominant online betting exchange.[1] Customers deposit funds with Betfair via bank transfer or credit card, and Betfair adjusts the customer's account each time a transaction is initiated or completed. Betfair generates revenue by charging a commission of between 2 and 5 percent on its customers' net winnings. Customers with a positive balance can withdraw funds at any time. Betfair publishes the three best prices currently being offered by customers wishing to back any particular bet, the three best prices offered by customers wishing to lay that bet, and the total stakes available at those prices.[2]

Given its capability to update instantaneously as individual customers enter and leave the market, it is natural that Betfair has been at the forefront of the development of in-play betting markets, in which bets can be backed or laid continuously while the sporting event in question is under way. Traditionally bookmakers offered bets only until an event began but more recently have adapted to provide in-play betting.[3] As such, these continuously operating online betting markets have ensured the transition of the use of high-frequency data from the financial setting into the betting market context. A high-frequency dataset records prices that are traded continuously at as high a frequency as possible, usually at the minute-by-minute level and often even second-by-second.

Betting markets have traditionally attracted academic interest, particularly from economists interested in the efficiency of markets: Leighton Vaughan Williams (1999) provided a survey of research on information efficiency in betting markets, mostly before the growth in online betting and the exchange-led revolution. As opposed to financial markets, where insider information can make timing of information breakage unclear and the true value of an asset is rarely ever revealed, sports betting markets offer nonexperimental opportunities to observe cleanly breaking information on assets whose value is revealed when the sporting event in question finishes, which can often be just minutes or seconds away. It is to these studies using these kinds of datasets to which we refer in this review of high-frequency investigations of information efficiency.

This chapter is organized as follows: section 2 provides an introduction to betting exchanges such as *Betfair* and high frequency data on in-play betting markets, section 3 considers how the value of bets is determined in-play, and then the concept of information efficiency is briefly introduced in Section IV. We will then consider a number of papers that have used high frequency data from betting markets to investigate information efficiency in Section V, splitting our review into weak-form efficiency (Section V.1) and semi-strong and strong form efficiency (Section V.2). Finally, Section VI concludes.

2 Betfair and In-Play Betting

In the 2010–2011 financial year revenue from football accounted for 42.3 percent of Betfair's sports gambling net revenues, narrowly overtaking for the first time horse racing, with its share of 42.1 percent. The United Kingdom accounted for 53 percent of Betfair's total "core" net revenue. The rest of Europe accounted for 42 percent of the total, and the rest of the world accounted for 5 percent. Betfair's ability to trade in any country is heavily dependent on the regulatory and taxation system applicable to the betting industry. The U.K. government adopted a relatively benign regulatory stance at an early stage of Betfair's historical development. Betting exchanges in the United Kingdom are subject to the conditions of a regular betting license with several additional restrictions imposed: for example, betting exchanges are not permitted to participate in bets; monies belonging to customers must be ring-fenced from the exchange's own resources; and the parties to any bet must not be identified to one another. Importantly, however, the licensing of individual customers is not a requirement. At the time of this writing, a number of other European countries were still developing a regulatory framework for online gambling. Regulatory and taxation arrangements, which can vary markedly from country to country, will have a major bearing on Berfair's prospects for future growth outside the United Kingdom. During the 2010–2011 financial year Betfair encountered severe regulatory difficulties in France and Italy, and the company ceased trading altogether in France in May 2010.

In addition to sports betting, Betfair offers a number of non–sports betting products, including poker, online casino games, and proprietary exchange games. Revenue from these core Betfair products in the 2010–2011 financial year was £330 million. This revenue figure reflects growth of 36 percent on the corresponding 2007–2008 figure of £242.3 million. Betfair claims to have had 949,000 active customers in 2011 (increased from 522,000 in 2008), accounting for 7 million betting transactions per day in 2011 (increased from 3.9 million in 2008). The Betfair Group's total revenue for the 2010–2011 financial year was £393 million. In addition to its core products, the group generates revenues from its other investments. These are: TVG (Betfair US), an online racetrack betting service that owns the TVH racing channel shown in 35 million U.S. homes, acquired by Betfair in January 2009; and LMAX, a platform for online retail financial trading, launched by Betfair in October 2010.

Betfair's growth in market share relative to the traditional High Street bookmakers has been driven by effective competition on price, facilitated by the lower transaction, information, and infrastructure costs associated with the online betting exchange business model. By exerting downward pressure on the bookmaker's overround, Betfair claims to be able to offer its customers more competitive prices than traditional bookmakers. In particular, the phenomenon of the favorite-longshot bias (with bets at long odds offering a lower average return than bets at short odds) is less apparent in Betfair prices than in those of the traditional bookmakers (Smith, Paton, and Vaughan Williams 2006). Mark Davies et al. (2005) cited the growth of Betfair as a manifestation of Moore's Law, that is, that the number of transistors on a microchip doubles every 18–24 months, resulting in a doubling of the speed of microprocessors. Before the 2000s the computing power needed to process the millions of transactions handled by Betfair daily would not have been available, and the company's subsequent growth has paralleled the growth in its technological capacity to process data.

Relative to its competitors, Betfair enjoys a positive network externality deriving, in part, from its first-mover advantage (Davies et al. 2005; Koning and van Velzen 2010). By virtue of the fact that Betfair has more customers than any other betting exchange, a trader is more likely to see his or her position matched on Betfair than on any other exchange. Accordingly, there is an incentive for all customers to trade using Betfair. This inducement may be decisive, even if other exchanges charge lower commissions than Betfair. Consequently the short history of the online betting industry has been punctuated by a series of high-profile closures, including Cantor Index Limited's SpreadFair online spread-betting exchange in December 2008 and Bet Bull Holdings' Betbull.com exchange in February 2009.

Through the transmission of new forms of betting market data, and by offering greater flexibility, including the facility to develop hedged positions by backing and laying, Betfair has been particularly successful in attracting business from sophisticated bettors. It is widely documented that from the mid-2000s onward, the traditional bookmaker firms have themselves made increasing use of betting exchanges, rather than racecourse bookmakers, for hedging their exposures. Unsurprisingly, however, the traditional bookmakers have been among the most vocal critics of Betfair and betting exchanges in general. Criticism has focused on the negative effect of the erosion of the traditional bookmakers' overround on the level of subsidy that flows from traditional bookmaker firms to the horse racing industry as well as on the danger that the facility to lay bets, in particular, creates opportunities for illegal trading on the part of sports industry insiders (especially in the case of horse racing) with privileged access to nonpublic information. Betfair has argued, on the contrary, that its unique technology places it in in a better position than that of the traditional bookmakers to identify suspicious betting patterns and take appropriate steps when the integrity of a betting market appears to be in jeopardy.

In-play, high-frequency, datasets embody several distinctive characteristics that we introduce by providing a graphical depiction of a particular high-frequency dataset before providing a more technical review. In this section, we consider prices from bookmakers and Betfair for a particularly interesting Euro 2008 match between Turkey and the Czech Republic. We specifically focus on the match outcome market rather than any of the other in-play betting opportunities now offered; later on in this section we discuss a number of alternative types of in-play markets. Before considering the evolution of prices across different online gambling markets, we first introduce the user interface for each market, beginning with Betfair. Figure 12.1 shows a screenshot from Betfair's match outcome market, which is the focus of all Betfair studies described in this chapter, for a match between Spartak Moscow and Chelsea in October 2010. A trader on Betfair can choose to back (buy) or lay (sell) contract i for event j at decimal odds DO_{ij}. If event j occurs, then the seller (layer) of the contract pays the backer (buyer) DO_{ij}. On Betfair, backers and layers make it clear how much money they wish to back or lay, and this is revealed underneath each decimal odd on Betfair such that

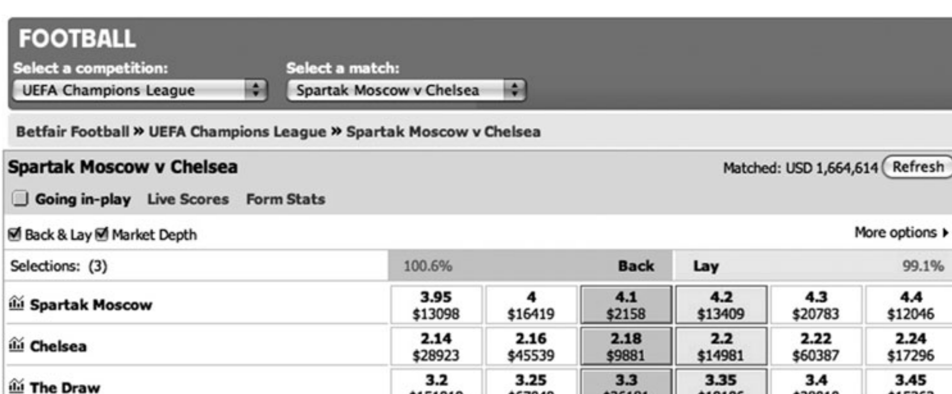


FIGURE 12.1 Screenshot from www.betfair.com illustrating how the betting exchange appears to a consumer (accessed Oct. 19, 2010).

any other bettor could choose to back or lay bets up to that particular amount. If $\$x$ is matched between a backer and a layer at decimal odds DO_{ij} then if event j happens the layer pays the backer xDO_{ij}. Thus in the context of the clash between Spartak Moscow and Chelsea, a user could have backed Chelsea at decimal odds of 2.18, meaning that for each dollar staked at these odds the user would have received $2.18 and that up to $9,881 was available to stake at these odds. A user exhausting all this available liquidity would stand to win $21,540.58 if (as happened) Chelsea won.

An alternative market structure is provided at Intrade.com, where contracts that pay out $10 if event j happens are traded at a price $\$P$.[4] Assuming the Law of One Price and an Absence of Arbitrage, the price $\$P$ that the contract trades at can be viewed as the probability of the event occurring. Figure 12.2 provides a screenshot of the Intrade market for the reelection of President Barack Obama; the price of the most recently traded contract at the time of the screenshot was $5.08, implying a probability of 50.8 percent that Obama would be reelected in 2012, as happened, taking 51.02 percent of the popular vote.[5] Users of Intrade could either buy or sell such contracts, as indicated by the green and red buttons, and when the event expired post-election, the buyer of a contract received $10 from the seller. The equivalence between market structures like TradeSports and Betfair is the reciprocal of the price; hence $P = 1/OD$; on both markets, the availability of contracts at particular prices is listed. The jargon differs somewhat, but the structures are essentially identical; bettors can go long (buying a contract or backing the event) or short (selling a contract or laying an event) and hence can cover their initial positions by taking opposing trades at a later stage.

Figure 12.3 returns to the previously mentioned Euro 2008 match between Turkey and the Czech Republic, plotting the implied probability, or the inverse of the decimal

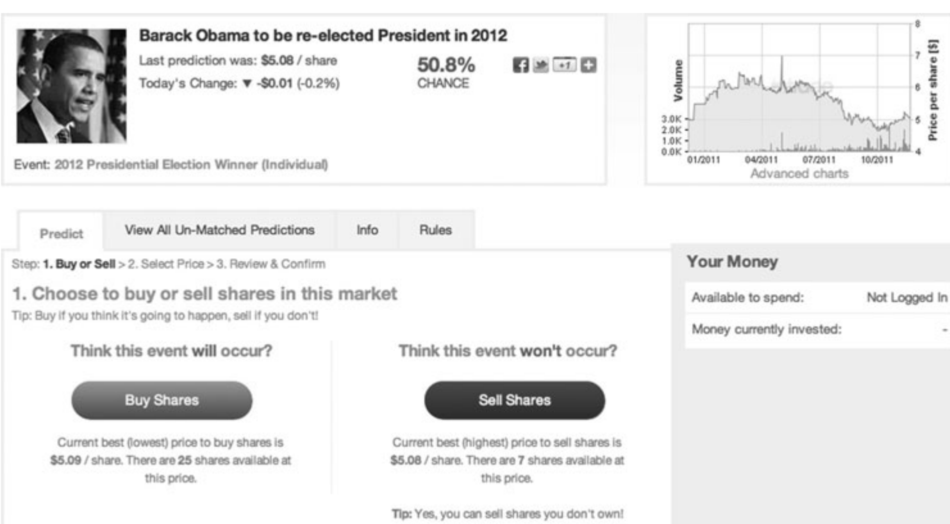


FIGURE 12.2 Screenshot from www.intrade.com market on the reelection of President Barack Obama (accessed Nov. 21, 2012).

odds, for a Turkey win. Prices from Betfair as well as the two largest traditional bookmakers, William Hill and Ladbrokes, are plotted. In this particular game, the Czech Republic scored the first two goals, which are noted on the plot as "1–0" and "2–0" and correspond to discrete shifts in the implied probability of a Turkey win. After the second goal, the decimal odds were such that the implied probability of a Turkey win was essentially zero. However, Turkey scored three goals in the final 15 minutes to win 3–2, and again these goals are marked on the plot. The third Turkey goal came so late that the Ladbrokes price never actually adjusted to it, remaining with an implied probability of around 15 percent until the end of the match (though punters would have been unable to back this event at this price). Betfair and William Hill did adjust, and the Betfair implied probability converged essentially to unity, implying a certain event, as would be expected in an efficient market.

Apparent in figure 12.3, relative to bookmakers, is the volatility of the Betfair series. The best back price available on Betfair is plotted, and naturally as bets are matched, and more offered for backing or laying, the price that is the best back or lay price will change; Karen Croxson and J. James Reade (2011b) found that, on average, \$527 of bets are matched *per second* on Betfair during football matches, showing just how quickly the best back price can change.[6] Also apparent from the inset in figure 12.3 is the rapidity of the adjustment to goals on Betfair relative to the bookmakers. In the inset, Betfair's market is suspended for 7 seconds, and within another 10 seconds of the

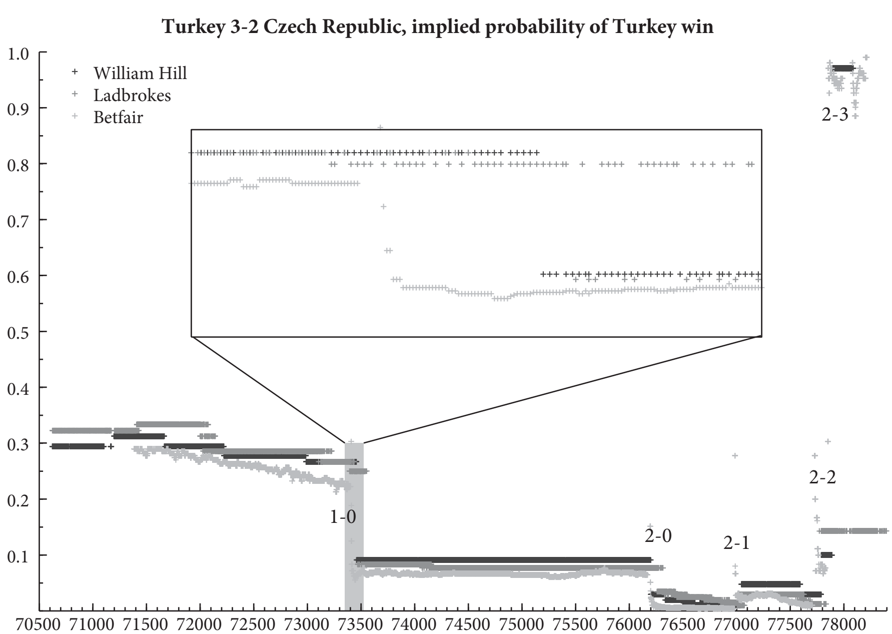


FIGURE 12.3 Graphical representation of information arrival in the market for a win by Turkey in the Euro 2008 match between Turkey and the Czech Republic.

market reopening, the price has settled at its new level, toward which the bookmakers move around half a minute later.

3 Valuation of In-Play Betting Odds

The valuation of in-play betting opportunities, conditional on the state of the match at the time a particular betting opportunity is available, requires the enumeration of probabilities for the occurrence of the outcome that is the subject of the bet in question. This section reviews two approaches that have been suggested in the academic literature for the computation of the required probabilities. Stephen Dobson and John Goddard (2011) employed Monte Carlo simulations to derive probabilities for final match result outcomes (in home win/draw/away win format) conditioned on the state of the match at any particular point within the regulation 90-minute match duration. An empirical model for the in-play arrival rates of player dismissals and goals was estimated based on prior data. Following a modeling approach similar to that of Mark Dixon and Michael Robinson (1998), it was assumed that the arrival rates of goals and player dismissals can be represented as Poisson processes, such that the probability that a new arrival occurs is independent of the time that has elapsed since the previous arrival. A competing risks model comprising hazard functions for four events (a goal being scored by the home team or the away team and a home team or an away team player being dismissed) was estimated. These hazards are linear in a set of covariates including team-quality measures (which do not vary while the match is under way) and a set of dummy variables that reflect the state of the match at any point, defined as the difference between the two teams in goals already scored (if any) and the difference between the two teams in players already dismissed (if any). Additional dummy variables controlled for variations in the goal-scoring and player-dismissal hazards during the first minute after kickoff at the start of each half and after a goal has been scored, and during the final recorded minute of each half, which usually extends for several minutes to include time added on to compensate for stoppages in play during the half.

The in-play home win, draw, and away win probabilities conditional on the current state of the match (again defined by the relative team-quality measure and the current differences between the teams in goals scored and players dismissed) are calculated from stochastic simulations of player dismissals and goals, which start from the current minute and continue over the remaining duration of the match, and are assumed to be generated in accordance with the estimated hazard functions. Simulated occurrences of the relevant events (goals and dismissals) are simulated by firing random numbers at the hazard functions.

The in-play probabilities are the proportions of home wins, draws and away wins obtained at the end of each simulated match, over 10,000 replications of this procedure.

For example, consider a match between two teams that are equally balanced in terms of quality, such that the home win/draw/away win probabilities (allowing for home field advantage) are of the order 0.44/0.28/0.28. If the match has remained level until the 15th minute, the simulations suggest that these probabilities should have drifted to around 0.42/0.31/0.27. An opening goal scored by the home team in the 15th minute then tilts the probabilities in the home team's favor, to 0.70/0.19/0.11. Conversely, an opening goal scored by the away team tilts the probabilities to 0.20/0.27/0.53.

Naturally the later in the match that the opening goal (or any goal that establishes a lead) is scored the bigger the value (expressed in terms of the shift in the probabilities) of the goal to the scoring team. An opening goal scored by the home team in the 75th minute tilts the probabilities rather dramatically in the home team's favor, from 0.22/0.63/0.15 to 0.81/0.17/0.02. The corresponding probabilities immediately following a 75th minute opening goal for the away team are 0.03/0.21/0.76. Although the Dobson and Goddard Monte Carlo simulation method is computationally demanding, it does provide a scientific basis, grounded empirically in the observation of past in-play match data, for the evaluation of the probabilities required for an objective assessment of the "correct" betting odds at any stage of the match.

As A. D. Fitt, C. J. Howls, and M. Kabelka (2006) pointed out, the range of available in-play betting opportunities extends to various spread betting markets. In contrast to fixed-odds betting or person-to-person betting via an online exchange, the possible payoffs to the bettor (both positive and negative) from a spread bet may be unbounded. For spread betting via an online bookmaker, the bookmaker quotes a spread (B, T), and the bettor has a choice of either "buying the spread at T" or "selling the spread at B." If the realized value of the index at the end of the event is S, the payoffs to the bettor for a unit stake are $S - T$ if the bettor bought the spread or $B - S$ if the bettor sold the spread. If the payoff is positive the bookmaker pays the bettor; if the payoff is negative the bettor pays the bookmaker.

Fitt, Howls, and Kabelka examined the valuation of spread bets on various outcomes of football matches, assuming that the arrival rates of goals, corners, bookings and sendings off can be modeled as Poisson processes. Commonly traded spread bets include total goals, total corners, multicorners (the product of the number of corners taken in each half of the match), and four flags (the time in minutes when corners have been taken from all four corner positions). They illustrate the calculation of payoffs to the bettor and bookmaker and assess the fair value of a range of bets. The assumption that the arrivals of goals and corners can be modeled as Poisson processes was found to provide a good approximation to the reality, but the arrivals of bookings and sendings off appear to be at variance with this assumption.

Below we describe in some detail the methodology employed by Fitt, Howls, and Kabelka to value two specific in-play spread bets: "total goals" and "multicorners". "Total goals" is the total number of goals scored by the home and away teams; "multicorners" is the product of the numbers of corners taken in each half of the match.

For any event, the probability of observing $N(a, b)$ occurrences during the time interval (a, b) can be written as follows:

$$P\left(N\left(a,b\right)=n\right)=\frac{\mu(b-a)^{n}e^{-\mu(b-a)}}{n!}, \tag{12.1}$$

where μ is the expected number of occurrences over the entire match. Applying this formula to a "total goals" spread bet, the expectation at duration t (where $t=0$ denotes the start of the match and $t=1$ denotes full-time) of the total number of goals scored, conditioned on r goals having been scored between duration 0 and duration t, is as follows:

$$C\left(t\right)=E_{t}\left(N|N\left(0,t\right)=r\right)\ =\sum_{n=0}^{\infty}\left(r+n\right)\frac{\left(\mu(1-t)\right)^{n}e^{-\mu(1-t)}}{n!}=r+\mu(1+t). \tag{12.2}$$

$C(t)$ is interpreted as the center spread for the spread bet on the total number of goals scored in the entire match. The center spread takes the value μ at the start of the match. As the match progresses the center spread decreases linearly by $\mu/90$ in each minute in which no goal is scored and increases by a jump of one on each occasion a goal is scored.

For a "multicorners" spread bet, let X denote the product of the first-half and second-half corner counts, and let $N(a, b)$ denote the number of corners taken during the period (a, b). We can write $X = N(0, 1/2)N(1/2, 1)$, and $\mathrm{E}_0(X) = \mathrm{E}(N(0, 1/2)N(1/2, 1))$. Let μ denote the expected number of corners taken (by both teams) in the entire match. $\mathrm{E}(N(0, 1/2)) = \mathrm{E}(N(1/2, 1)) = \mu/2$, and $\mathrm{E}_0(X) = \mu^2/4$. The expectation at duration t in the first half of X, conditional on r corners having already been taken between duration 0 and duration t, is as follows:

$$\begin{aligned}C\left(t\right)&=\left(\sum_{n=0}^{\infty}\left(r+n\right)\frac{\left(\mu(1-t)\right)^{n}e^{-\mu(1-t)}}{n!}\right)\times \mathrm{E}(N(1/2,1))\\&=\left[r+\mu(1/2-t)\right]\left(\frac{\mu}{2}\right)=\frac{\mu^{2}}{4}-\frac{\mu^{2}t}{2}+\frac{\mu r}{2}.\end{aligned} \tag{12.3}$$

The expectation at duration t in the second half of X, conditional r corners having already been taken between duration 0 and duration t, R_1 of which were taken in the first half, is as follows:

As before, $C(t)$ is interpreted as the center spread for the multicorners spread bet. The center spread takes the value $\mu^2/4$ at the start of the match. During the first half the center spread decreases linearly by $\mu^2/180$ in each minute in which no corner is taken and increases by a jump of $\mu/2$ each time a corner is taken. During the second half the center spread decreases linearly by $\mu R_1/90$ in each minute in which no corner is taken and increases by a jump of R_1 each time a corner is taken.

Similar procedures are used by Fitt, Howells, and Kabelka to calculate fair prices for a wide range of in-play spread bets. The time paths of the valuations over the duration of particular football matches are compared with the time paths of the quoted spreads in order to evaluate whether the spread betting prices are fair and whether the betting markets in question are informationally efficient. Clearly the assumption that the arrival rates follow Poisson processes with constant means over the entire match duration is a simplification: it is widely documented, for example, that more goals are scored during the later stages than during the early stages of matches and that the rates at which bookings and sendings off accrue likewise increase over the duration of the match. Nevertheless, the use of mathematical modeling grounded in probability theory provides a useful starting point for the valuation of a wide range of in-play spread bets.

Fitt (2009) applied Markowitz portfolio theory to the selection of spread bets. In formulating a spread-betting strategy in a situation where multiple betting opportunities are available, Fitt suggested that the bettor faces a choice analogous to the choice faced by the risk-averse investor in portfolio theory. The investor (bettor) seeks to create a portfolio of assets (bets) that either maximizes the expected return for any given target level of risk or minimizes the risk for any given target expected return. Risk is measured by the variance of the return on the portfolio of assets (bets). In the case of betting on football, all of the risk is concentrated into 90 minutes of play, and no alternative risk-free asset exists that would offer a meaningful return over such a short period. The bettor's sole decision is the selection of the unique optimum portfolio of risky bets. For any array of bets the computation of the optimal portfolio is a mechanical task. Theoretically there is only one unique combination out of any available selection of bets that any bettor need ever consider.

4 Information Efficiency

The informational efficiency of markets is a fundamental concept in economics; the contention that the market price is the most effective distributor of information to the greatest number of people is a central tenet of economic theory.[7] Friedrich Hayek (1945, p. 526) captured the idea as follows: "The mere fact that there is one price for any commodity ... brings about the solution which ... might have been arrived at by one single mind possessing all the information which is, in fact, dispersed among all the people involved in the process." The idea is highly controversial, not least because of the risk thus in financial markets that market participants with greater amounts of information may profit at the expense of those with inferior information. However, if a market is efficient, it is argued, its price will instantaneously and fully adjust to any such new information, thereby ruling out the possibility of well-informed traders benefitting at the expense of those with less information. The concept is perhaps most commonly associated with Eugene Fama 1965, 1970, 1998) though also further back to Louis Bachelier (1900).

Fama delineated three levels of information efficiency that vary in stringency. The first level, *weak-form efficiency*, stipulates that current prices embody all information contained in historical prices. Thus no trading strategy is possible based on past price levels and movements that can yield a positive expected return. A particularly important implication of this is that betting at particular prices, or ranges of prices, should not yield different expected returns to betting at any other price or range of prices. A violation of this would be what is referred to in betting markets as the favorite-longshot bias (FLB), where the odds offered on favorites and outsiders are systematically biased. Various studies have investigated FLB, and they will be discussed in section 5.1.

The second level outlined by Fama is *semi-strong form efficiency* (SSFE), which requires that market prices reflect all publicly available information; hence profitable trading strategies based on publicly available information cannot be possible. With SSFE, Vaughan Williams (1999) noted the implication that across different providers of bets (bookmakers, betting exchanges) it should not be that expected returns differ and also pointed out that the response of markets to information is crucial. If market prices are known to over- or underreact to news events then this information could be used to construct a profitable betting strategy.

The third and strictest notion of information efficiency is *strong-form efficiency*, which requires that market prices reflect all information whether publicly or privately held. Considering strong-form efficiency where private information is factored in, it should not be possible for any participant trading on superior information to make abnormal returns. Furthermore, prices set later in a market, after trading has taken place based on private information, should not incorporate any more information than that set earlier in the market. In particular, insider trading constitutes a central focus of investigations of strong-form efficiency.

We shape our review of the literature investigating information efficiency in high-frequency betting markets around these three levels of efficiency.

5 Information Efficiency in High-Frequency Betting Markets

We now review investigations into information efficiency in high-frequency markets by considering the three levels of efficiency outlined in the previous section; in section 5.1 we review investigations into weak-form efficiency before, in section 5.2, we look at studies of semi-strong and strong form efficiency.

5.1 Weak-Form Information Efficiency

As mentioned in section 4, weak-form information efficiency has often been assessed by investigating the favorite-longshot bias, which states that bets placed at different prices yield different expected returns. Testing for FLB in betting markets has been carried out by many authors in the context of explaining the phenomena; Linda Woodland and Bill Woodland (1994) tested in the context of U.S. baseball markets, finding a reverse FLB; Erik Snowberg and Justin Wolfers (2010) attempted to explain FLB as a misperception on the part of bettors using pre-event data from U.S. horse races; Michael Smith, David Paton, and Leighton Vaughan Williams (2006) considered U.K. horse racing pre-event and concluded in favor of an information explanation for FLB; bettors face real or perceived costs of information accumulation, meaning that they accept the prices offered by bookmakers that embody FLB as a profit maximizing strategy (Levitt 2004). Smith, Paton, and Vaughan Williams (2006) found in particular that the person-to-person betting exchange prices exhibit substantially less FLB than do bookmaker prices.

A common calibration test (employed by, for example, Woodland and Woodland 1994 and Franck, Verbeek, and Nüesch 2010) asks whether a contract priced such that the implied probability of the event occurring is $x \in [0, 1]$ pays out $100x$ percent of the time and is usually conducted via some form of limited dependent variable regression involving the observed outcome for event i, $y_i \in \{0, 1\}$regressed on the implied probability from market m for event i, $p_{m,i}$:

$$y_i = \beta_0 + \beta_1 p_{m,i} + \varepsilon_i. \tag{12.4}$$

Because $\partial y_t / \partial p_{m,i} = \beta_1$ it would be expected that an efficient market is such that $\beta_1 = 1$: the actual (or frequentist) probability of an event, and the market's subjective probability of it, move one-for-one. In addition, no wedge would be expected between the actual probability and the market's subjective probability such that $\beta_0 = 0$ also. Hence the absence of any bias implying that $\beta_0 = 0$ and $\beta_1 = 1$ so that the event is accurately predicted by the implied probability from market m. It is perhaps useful to note that $\beta_0 = 0$ is thus, in essence, simply a paired difference in means test between the actual event outcome probability and the subjective one, and the $\beta_1 = 1$ test is one of proportional movements in these two entities. While it might be asserted that other information surely exists that is correlated with both y_i and $p_{m,i}$, hence likely causing omitted variable bias in the calculation of β_1, it ought to be the case that this information in an efficient market has already been factored into $p_{m,i}$, hence that additional information should not actually affect coefficient estimates.

The Franck, Verbeek, and Nüesch study (2010) compared bookmakers and Betfair pre-event and found evidence that Betfair exhibits less FLB than do bookmakers and thus is more efficient, though the study also provided evidence against semi-strong form efficiency in the betting market more generally.

Croxson and Reade (2011a) considered high-frequency data for bookmakers and Betfair during the Euro 2008 football tournament. They thus extended the calibration testing from the cross-section-style setting of pre-event prices into a more panelesque environment in which numerous events were considered (N) but also many observations through time (T) during these matches. Due to the nature of matches lasting slightly different lengths (injury time), but also due to technical difficulties of collecting the data, the authors derived different T's for each N and hence have an unbalanced panel. However, because, as pointed out by Woodland and Woodland (1994), the aim with equation (12.4) is not so much to provide a model with high explanatory power as to test how well calibrated is the market, the authors did not treat equation (12.4) as a panel model. Not least, each panel would have a dependent variable without any variation, but additionally this would involve restricting the dataset over which calibration is assessed; the entire premise of calibration testing is to test over as large a number of observations as possible whether contracts priced at a particular implied probability pay out that often. Croxson and Reade (2011a) upheld the results found by both Smith, Paton, and Vaughan Williams (2006) and Franck, Verbeek, and Nüesch (2010) in non-high-frequency data settings: Betfair is more accurate, exhibits lower FLB, but nonetheless still exhibits some bias.

Croxson and Reade (2011b) also considered weak-form efficiency in a considerably larger dataset graphically; their dataset contains second-by-second information from 1,206 football matches across nine different competitions at both domestic club level and international team level, including the Euro 2008 matches referred to above.[8] Figure 5 in Croxson and Reade (2011b), reproduced here as figure 12.4, displays a graphic representation of the calibration test described above. On the horizontal axis is the win probability implied by the Betfair price of a contract, while on the vertical axis is the proportion of contracts priced at this level that pay out. A calibrated and hence weak-form efficient market would have all such proportions equal to the implied probability of the price, hence plotted along the 45-degree line; in this situation the expected returns from consistently placing bets at particular ranges of prices would be the same. From figure 12.4, by and large this is the case for this high-frequency market; the implied slope of the plotted points is 1.04, implying a slight, and visible, FLB, as deviations from the 45-degree line are slightly below the line when the probability is less than 0.5 and generally above the 45-degree line when the probability is greater than 0.5.

Samuel Hartzmark and David Solomon (2008) investigated NFL matches via betting contracts traded on the U.S.-focused betting portal TradeSports.com and considered one specific departure from SSFE known within the behavioral finance literature as the *disposition effect.* TradeSports.com operates somewhat differently from Betfair, as described in section 2. Hartzmark and Solomon produced a plot similar to figure 12.4 in their figure 2, plotting the frequency with which contracts priced at particular implied probabilities pay out. Hartzmark and Solomon noted not that the points appear to suggest presence of the FLB, as the implied slope of the points is greater than unity, but that they actually appear to follow an S-shaped pattern. At the lower end of the range of implied contract probabilities points are generally below the 45-degree line

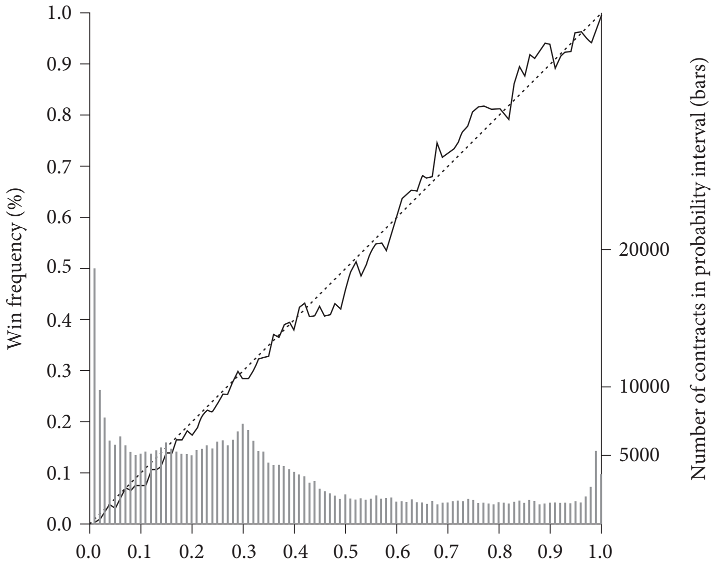


FIGURE 12.4 In-play calibration test.

Source: Figure 15 from Croxson and Reade (2011b).

but converge back to the line as the probability approaches zero, and likewise at the top end of the range, again as the event converges on certainty, so does the actual frequency of payouts, and hence rather than being a linear relationship, what is observe is an S-shaped curve.

Hartzmark and Solomon suggest that this S-shaped pattern is supportive of a disposition effect interpretation rather than the FLB interpretation discussed thus far. The disposition effect states that traders close out winning positions earlier than they ought to and hold on to losing positions longer. This may manifest itself in the post-news period in price movements; if a favorite has scored, those that hold positions with the favorite (they bought contracts that pay out in the event the favorite wins) may seek to sell contracts at this point to exploit the now substantially higher price that such a contract will secure. Such selling pressure will exert a downward pressure on the implied probability (or TradeSports.com price) of such a contract, and indeed this is exhibited in the markets Hartzmark and Solomon study, as for a fixed probability of winning (vertical axis) prices are pushed to the left for favorites (contracts implying probability above about 60%). Equivalently if a news event occurs that makes a team less likely to win than it was previously, many traders who hold contracts on that team face a losing position and should seek to sell such contracts to cover any losses. Thus if the disposition effect exists we might expect to see an absence of selling pressure after a negative news event for that team, pushing its price up, as people hold on to these losing positions in the hope they recover. Again, this can be observed in both Hartzmark and Solomon's figure 2 and figure 12.4 in this chapter; for a contract that has a low probability of paying out, the price is pushed down (hence rightward on the horizontal axis). Hence we would find that when a price (or implied probability) has reached the upper end of the scale, very near to unity, then prices are pushed back down relative to what they would otherwise be (upward biased due to the FLB) and vice versa for very unlikely events where the implied probability has reached near zero. It is important to point out that these price distortions are argued to be short term, and as such, over a longer time period it would be expected that prices would return to the level at which they should be. It is this aspect of the disposition effect that the authors use to identify it from other potential explanations (such as the standard explanations for the FLB outlined in Snowberg and Wolfers 2010 and Smith, Paton, and Vaughan Williams 2006).

As well as estimating the S-shaped curve using nonlinear least squares and bootstrapping to get standard errors to investigate the significant of the S-shape, they ran a number of regressions to examine whether price movements do indeed correspond to their hypothesized explanation. They established that there is significant price reversal for favorites after positive news shocks and for outsiders after negative news shocks. Although their attempt to rule out alternative explanations of the S-shape is quite weak, they nonetheless established yet further the absence of weak-form efficiency in high-frequency betting markets both via the S-shaped curve and by identifying serial correlation in post-news price movements.

5.2 Semi-Strong and Strong Form Information Efficiency

Turning to semi-strong and strong-form efficiency, these two concepts explore the reaction of markets to new information, with the former considering only public information and the latter public *and* private information. At the high-frequency level, Ricard Gil and Steven Levitt (2007) considered betting prices from Intrade for the 2002 FIFA World Cup, while Croxson and Reade (2011b) used prices for more than one thousand football matches from many competitions worldwide and Buraimo, Peel, and Simmons (2008) considered a significant sample of English Premiership football matches between 2006 and 2008.

Gil and Levitt (2007) tested information efficiency using the 2002 FIFA World Cup and Intrade markets, attempting to test SSFE with what is essentially a dynamic calibration test; they tested the hypothesis of rapid updating of prices to new information, which they define to be goals in football matches. Table 1 from Croxson and Reade (2011b) makes it clear, as does figure 12.3 graphically, that goals are *the* big news events in football matches. All of these studies add that the football setting makes the testing of SSFE that much cleaner; even though match outcomes can be anticipated (legally or otherwise), the exact timing of goals is essentially random such that even in the seconds before a goal no market participant can be certain that a goal will occur. Furthermore, once a goal has happened, the news is immediately spread; all online betting portals now have facilities through which events can either be viewed visually or followed via audio feeds, and in addition, Betfair suspends its market the moment a goal is scored such that it is obvious to all market participants that a material event has occurred.

These three papers each individually propose different methods to counter the joint hypothesis problem because despite the fixed endpoint and value for betting contracts, the problem remains. Specifically, while a match is in-play, each second and minute that pass provide new information independent even of events that took place in the preceding minute; the information is that less time exists in which either team can influence the outcome of the match. As such, in-match offered and implied probabilities always drift toward either zero or unity depending on whether at that point that event was what the match scoreline reflected (i.e., if the score was a draw or one team was favored over the other). This is displayed graphically in figure 12.3; while Turkey is not winning, its implied probability drifts toward zero. The problem thus is ascertaining whether the drift post-goal is caused by the information embodied in the goal or in the subsequent seconds and minutes that pass. The method that Gil and Levitt (2007) employed is calibration testing; by averaging over all goals they arrived at the objective (frequentist) and subjective (Intrade) probabilities of the contract outcome. They essentially repeated this test for 15 minutes before a goal happened and 15 minutes after; the hypothesis of information efficiency was thus that post-goal there should be no drift in the subjective probability (and neither pre-goal for efficiency). Hence this method of testing removes the drift problem using the information embodied in the fixed endpoint value of each contract, the property of sports betting markets noted in section 1.

Gil and Levitt included dummy variables for each of the 30 minutes in their sample and tested the significance of these dummies both pre- and post-goal. Specifically, they ran the regression model:

$$p_{m,c,w,t} = \beta_0 + \sum_{g=-14,g\neq 0}^{15} \beta_g Goal_{m,c,w,t+g} + \varepsilon_{m,c,w,t} \sim IID(0,\sigma^2), \tag{12.5}$$

where p denotes the volume-weighted average price and the subscripts m, c, w, and t denote the particular match, contract traded (we consider home win and away win contracts), goal "window," and minute of time considered.[9] *Goal* is an indicator variable corresponding to a particular minute which is equal to unity when a goal is scored in favor of the contract and zero otherwise. Hence the null hypothesis of their test for information efficiency is that $\beta_g = \beta_h$, $g \neq h$, g, $h > 0$; that post-goal there is no drift up or down in prices.

The result of their test is perhaps best displayed graphically; the estimates for all β coefficients in equation (12.5) are displayed in Gil and Levitt's figure 1. Post-goal (average) prices drift upward and are much more volatile than pre-goal. This suggests that the news of a goal has a substantial and long-lasting impact on the market, indicative of an *absence* of SSFE.

It should be emphasized again that this drift should not be confused with the drift that would be observed in any individual match price as the event neared its conclusion.

Nonetheless, there are problems with the Gil-Levitt test, notably an absence of liquidity. In their markets there were few participants (on average just 75 individuals per match), trades were relatively infrequent (on average 100–200 per match), and the overall volume of money traded was very low, at about \$1.5 million across all matches in the World Cup, while their sample of matches at 50 is rather small in terms of a calibration test. Buraimo, Peel, and Simmons (2008) went on to criticize the econometric specification of equation (12.5), noting that it does not accommodate the likely heteroskedasticity caused by the bounded nature of subjective and objective probabilities. Model misspecification need not be a moot point here; as Gil and Levitt considered market prices, if they are efficient then they all satisfy a random walk, meaning that the first lag of the price would be highly significant and its omission could cause spurious significance. Buraimo, Peel, and Simmons (2008) raised the additional concern that Gil and Levitt did not calibrate the time of trading on Intrade (GMT) and the time of goals scored (minutes in match), meaning that there may be discrepancies between the two, discrepancies that could explain the prolonged upward shift around minute zero (when the goal happens) found by Gil and Levitt.

Croxson and Reade (2011b) and Buraimo, Peel, and Simmons (2008) studied SSFE in the context of Betfair betting markets on match outcome, and despite different methodologies appear to have arrived at similar results. Croxson and Reade considered more than one thousand football matches from various competitions (see their table A.3), while Buraimo, Peel, and Simmons considered 296 Premiership matches between 2006 and 2008.

Croxson and Reade adopted a dual strategy for investigating efficiency, both using unit-root testing and investigating the existence of feasible and profitable trading strategies.

They attempted to circumvent the joint-hypothesis problem by considering only the half time interval rather than general in-play prices. This removed information related to the passage of time from the observations they considered because during halftime, or at least for the first few minutes of halftime, the clock stops and little if any newsworthy events that are publicly known occur; teams are in their dressing rooms, and TV stations are on advertising breaks. Hence the information causing the drift during play ought not to exist at this point, and in-match plots of prices/implied probabilities, such as figure 12.3, generally reveal this visually (see the observations around 74500). In particular Croxson and Reade considered whether the event of a goal on the cusp of halftime (within a minute of the end of play) has any impact on the evolution of prices during halftime relative to matches where no goal events happen. They adopted a panel unit root testing approach and in order to cope with fat-tailed error distributions simulated critical values. They reached somewhat ambiguous results, with some panel test variants suggesting a rejection at the 5 percent level but not the 1 percent level (see their table 3). Responding to this ambiguity they investigated potentially profitable strategies: either laying one of the contracts at the start of halftime and backing it five to ten minutes later to exploit any upward drift or, vice versa, to exploit downward drift. They found, using a paired difference in means t-test, that neither strategy would have yielded a positive return on average (see their tables 4 and 5) and hence determined that while the statistical evidence may be mixed, there were no feasible trading strategies possible to exploit any drift or inefficiency during the halftime interval and hence concluded in favor of SSFE.

Interestingly, on their substantially larger dataset Croxson and Reade ran the model of Gil and Levitt and found instead of an upward drift an insignificant hint of negative drift (see their figure 9). In addition, with such a larger dataset, the reduced volatility of the different β estimates is notable. These results further emphasize their conclusions *in favor of* SSFE.

Buraimo, Peel, and Simmons alternatively considered, a la Gil and Levitt, the immediate time after a news event has been occurred and asked how soon the market would respond. As the title of their study suggests, they found that within about 60 seconds all information has been impounded into the market price. They also asserted that the nature of information contained in each goal can be dramatically different if the goal is scored early or late, or happens when the scores are level or one team is several goals ahead, and suggested that Gil and Levitt did not account for this (and neither did Croxson and Reade). However, arguably Gil and Levitt did via their averaging over all goals in their dataset to consider the objective and subjective outcome probabilities surrounding these goals, and equally did Croxson and Reade in the sense that they allowed for different-sized jumps in their autoregressive testing via the error/residual term.

Buraimo, Peel, and Simmons developed a model that accounts for the different magnitude of jumps in price for goals of differing importance and, as such, also attempted to control for the heteroskedasticity problem. Specifically they proposed that the true probability of an outcome, p, evolves according to:

$$dp = p\left(1 - p\right)u, \tag{12.6}$$

where u is news. Hence if $p = 0.5$ then the impact on the true probability is largest and decreases as the probability nears zero or unity in order that a news event cannot push the probability outside the unit interval, on which it is bounded. The authors then used a discretized version of (12.6) as their dependent variable:

$$Y_t = \frac{p_t - p_{t-1}}{p_{t-1}\left(1 - p_{t-1}\right)}. \tag{12.7}$$

They identified news events by considering the extremes of the distribution of residuals from an AR(1) model of Y_t from (12.7):

$$Y_t = \alpha_0 + \alpha_1 Y_{t-1} + \mu_t. \tag{12.8}$$

Hence Buraimo, Peel, and Simmons did not simply consider goals to be the only news events but instead attempted to identify all news events empirically via u_t in (12.8), considering residuals in the very tails of the distribution (top and bottom 0.025%) to be news events. This led them to identify considerably more than just goals: they uncovered 1,445 "newsbreaks," of which only 454 corresponded to goals. While goals are infrequent events, Croxson and Reade's sample (more than one thousand matches from multiple competitions, domestic, European, club level and international level) contains on average 2.55 goals per game; this would imply about 755 goals in Buraimo's sample, and hence it would appear that their strategy missed a number of goals, most likely those that occurred with a team already significantly ahead and late in the game, therefore not registering as a news event.

Specifically Buraimo, Peel, and Simmons tested for weak-form efficiency by testing that in the aftermath of a goal Y_t is not serially correlated; that is, news isn't serially correlated and instead its impact happens in just one period. This would appear to be more of a test of SSFE than of weak-form efficiency, and hence we treat it as that. The test is a simple t-test of the significance of α_1 in (12.8). Buraimo, Peel, and Simmons considered six time frequencies to test SSFE: 2 seconds, 5 seconds, 10 seconds, 30 seconds, 1 minute, and 2 minutes and hence aggregated their data for each case. They measured the time from the news event, and hence if the Betfair market was suspended, then the 2- and 5-second regressions will miss some news events since at 2 or 5 seconds later the market is still suspended. Nonetheless, their table 3 reports the results of the t-test on α_1 for these varying time frequencies. For 2 and 5 seconds the t-statistics are greater than or equal to 1.96 and hence significant, indicating that at such short time horizons the market is not efficient, not absorbing news instantaneously. However the t-statistics decrease with the time interval, and for 10 seconds the t-statistic is already below 1.96; hence it might be concluded that for time intervals of 10 seconds and longer the market fully processes any new information.

Buraimo, Peel, and Simmons did also uncover evidence against weak-form efficiency, pre-match, by noting that betting strategies involving backing home teams that were outsiders generated significantly higher (positive) returns (16.95%) than did backing home teams that were favorites (4.87%), while backing away team favorites incurred smaller losses (−11.7%) than did backing away team outsiders (−31.6%). In-play, however, after news events, they found that all strategies resulted in losses, though again they found that returns differed for different types of bets, arguing against weak-form efficiency, though in favor of strong-form efficiency since no information, public or private, could be used to generate positive returns.

Continuing with strong-form efficiency, as mentioned in section 4, insider trading is a particular aspect of this; were a market strong-form efficient, insider trading would not be profitable, as the market would reflect any private information already held. Betfair has attracted significant controversy by allowing market participants to lay bets; although those associated with a particular horse are not allowed to lay (to bet against or to act as the bookmaker for a bet on) a particular horse, they can provide information to others regarding horses. Anecdotally Betfair is much more transparent than traditional bookmakers regarding user accounts, and a number of successful investigations have been carried out against suspected suspicious activity noticed by Betfair.[10] That Betfair has a team of employees working specifically on detecting suspicious activity suggests that it does exist and hence that substantially positive expected returns must exist, arguing against strong-form efficiency (again, only anecdotally).

5.2.1 *Information Efficiency across Betting Market Formats*

As mentioned in section 4, Vaughan Williams (1999) noted that different expected returns found across different providers of bets can provide evidence against SSFE in the betting market as a whole. The high-frequency comparison of bookmakers and betting exchanges in (Croxson and Reade 2011a) thus provides insight into whether SSFE exists in betting markets more generally than just in Betfair. Via their tables 8–11 and figure 12 they clearly showed that the expected returns on Betfair for winning bets are statistically significantly higher than those available from traditional bookmakers; in addition, they investigated the size of the bet for which this ceases to be the case on betting exchanges by calculating the average return possible once one wishes to bet more than is available at any given price on Betfair. It is only for bets of over \$450 in size that traditional bookmakers begin to become competitive on average with the returns provided by Betfair. As such, this would appear to provide evidence against SSFE.

Croxson and Reade (2011a) also investigated the process of information arrival on betting exchanges, making use of the cointegration methodology applied by Joel Hasbrouck (1995) to stocks listed on multiple exchanges in financial markets. It would be expected that in terms of information discovery, and hence efficiency, Betfair would lead due to its disaggregated nature; anyone with information can trade, and because there are sufficient numbers of traders, if the new information is initially mispriced this inefficiency will quickly be arbitraged away. Conversely, bookmaker odds are set and altered by some small group of experts operating within the bookmaker, and hence it would seem plausible that these would be unable to react with the speed and accuracy of Betfair. Figure 12.3 would appear to graphically make the case that it must be Betfair leading information discovery: after a goal happens in the zoom inset, the Betfair market is suspended for about seven seconds before reopening, and it appears that within ten seconds of restarting the price at Betfair has reached its new level, around which it then fluctuates. The two bookmaker prices, conversely, do not appear to adjust to their new level for around a minute (William Hill) and almost three minutes (Ladbrokes). This plot would suggest that Betfair displays quite notable SSFE, adjusting almost instantaneously to new information (the goal), while the bookmakers do not; furthermore, it suggests that information discovery is clearly led by Betfair rather than the bookmakers.

Of course this is but one example of one goal in one match, yet it provides quite a clear picture of the process of information arrival in the betting market: Betfair appears to adjust much more quickly than the bookmakers to the news of a goal. Croxson and Reade then attempted to test this formally using the cointegrated VAR methodology of Søren Johansen (1995), a method that allows the identification of stationary steady-state relationships between nonstationary variables (such as between prices from the different betting companies) and also, importantly, the adjustments to these relationships. If a firm's price *does not adjust*, it implies that that firm leads, or drives, the relationship between the prices and hence *leads information discovery*. It would be expected then that the remaining prices adjust to the relationship, meaning that the information is then transmitted from the leading price to the following ones.

However, figure 12.3 does hint at some of the difficulties with attempting to statistically model raw data series like these; the bookmaker prices are constant for long periods with discrete jumps, and furthermore the odds are exponential in their behavior, converging slowly to unity (certain event) but also exploding to very high values as an event becomes increasingly unlikely, while if probabilities are used then they are constrained to lie on the unit interval hence inducing heteroskedasticity, as Buraimo, Peel, and Simmons point out. In addition, as noted earlier by Croxson and Reade, Betfair returns (hence odds) are statistically significantly better than bookmakers, and this also can be observed in figure 12.3; although all three implied probabilities adjust to the news, they appear to converge to slightly different values. Reflecting these difficulties, the cointegration results provided by Croxson and Reade are somewhat ambiguous in the support they offer for the Betfair leadership hypothesis; although the median values over all 22 matches (they estimate each match separately as opposed to attempting a panel cointegrated VAR model) are supportive of Betfair leading information accumulation, all test results reject this idea, finding that all firm prices adjust to the cointegrating vectors (tables A.6–A.11).

6 Conclusions

In this chapter, we provide a review of recent academic research on the topic of information efficiency in high-frequency in-play football betting markets. Several studies have reported evidence that is contrary to the weak-form information efficiency hypothesis, in the form of a favorite-longshot bias in in-play betting prices. However, there is evidence in the literature in favor of the semi-strong form of the information efficiency hypothesis, with in-play betting prices shown to respond rapidly to the arrival of large news events in football matches, especially goals being scored. One study we reviewed also reports interesting evidence in support of the strong-form information efficiency hypothesis. As in-play betting markets continue to develop and gain in popularity, driven in part by further improvements in computing power, we anticipate parallel growth in research on information transfer and price formation in financial markets, an exciting new area of academic study.

Notes

1. See Betfair (http://betfair.com)
2. See figure 12.1 for a screen grab from Betfair's website.

3. In-play betting is also often described as in-running; in this chapter we refer to betting while an event is occurring as in-play betting.
4. Intrade's sports betting portal, TradeSports, which Hartzmark and Solomon (2008) used for their empirical investigation of the disposition effect, was one of the victims of the intense competition in the online sports betting industry, closing in November 2008. See Chris F. Masse, "TradeSports ceases its operations," MidasOracle.org, Nov. 16; http://www.midasoracle.org/2008/11/18/tradesports-end (last accessed Nov. 21, 2011).
5. Wrade suspended all tradiry in March 2013.
6. This naturally varies depending on the type of match under consideration, and raises to as much as $2,534 per second during Euro 2008 matches, of which the match plotted in figure 12.3 is one (see table A.3 in Croxson and Reade 2011b).
7. Information efficiency in markets is often synonymously described in the literature as *market efficiency*; in this chapter we use the term *information efficiency*.
8. This corresponds to a number of observations well in excess of one million.
9. Gil and Levitt (2007) also demean their price data; since this is a linear transformation its effect is simply captured here in the constant coefficient, β_0. Gil and Levitt omit this term.
10. See "Eddie Freemantle, "How Betfair beat the bookies," *The Observer*, March 29, 2009 (last accessed Nov. 10, 2011) for a number of such examples.

References

Bachelier, Louis. [1900] 1964. Theorie de la speculation. In *The random character of stock market prices*, edited by Paul Cootner. Reprint, Cambridge, Mass.: MIT Press, 17–78.

Buraimo, Babatunde, David Peel, and Rob Simmons. 2008. Gone in 60 seconds: The absorption of news in a high-frequency betting market. Selected works of Babatunde Buraimo, January; http://works.bepress.com/babatunde_buraimo/17.

Croxson, Karen, and J. James Reade. 2011a. *Exchange vs. dealers: A high-frequency analysis of in-play betting prices.* Discussion Paper No. 11-19. Birmingham: Department of Economics, University of Birmingham.

———. 2011b. *Information and efficiency: Goal arrival in soccer betting.* Discussion Paper No. 11-01. Birmingham: Department of Economics, University of Birmingham.

Davies, Mark, Leyland Pitt, Daniel Shapiro, and Richard Watson. 2005. Betfair.com: Five technology forces revolutionize worldwide wagering. *European Management Journal* 23(5):533–541.

Dixon, Mark J., and Michael E. Robinson. 1998. A birth process model for association football matches. *The Statistician* 47(3):523–538.

Dobson, Stephen, and John Goddard. 2011. *The Economics of Football.* 2nd ed. Cambridge: Cambridge University Press.

Fama, Eugene F. 1965. The behaviour of stock market prices. *Journal of Business* 38(1):34–105.

———. 1970, Efficient capital markets: A review of theory and empirical work. *Journal of Finance* 25(2):383–417.

———. 1998. "Market efficiency, long-term returns and behavioral finance. *Journal of Financial Economics* 49(3):283–306.

Fitt, Alistair D. 2009. Markowitz portfolio theory for soccer spread betting. *IMA Journal of Management Mathematics* 20(2):167–184.

Fitt, A. D., C. J. Howls, and M. Kabelka. 2006. Valuation of soccer spread bets. *Journal of the Operational Research Society* 57(8):975–985.

Franck, Egon, Erwin Verbeek, and Stephan Nüesch. 2010. Prediction accuracy of different market structures—Bookmakers versus a betting exchange. *International Journal of Forecasting* 26(3):448–459.

Gil, Ricard, and Steven D. Levitt. 2007. Testing the efficiency of markets in the 2002 World Cup. *Journal of Prediction Markets* 1(3):255–270.

Hartzmark, Samuel M., and David H. Solomon. 2008. Efficiency and the Disposition Effect in NFL Prediction Markets. Working Paper.

http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1540313

Hasbrouck, Joel. 1995. One security, many markets: Determining the contributions to price discovery. *Journal of Finance* 50(4):1175–1199.

Hayek, Friedrich A. 1945. The use of knowledge in society. *American Economic Review* 35(4):519–530.

Johansen, Søren. 1995. *Likelihood-based inference in cointegrated vector autoregressive models.* Oxford: Oxford University Press.

Koning, R. H., and B. van Velzen. 2010, Betting exchanges: The future of sports betting? *International Journal of Sports Finance* 4:42–62.

Levitt, Steven D. 2004. Why are gambling markets organised so differently from financial markets? *Economic Journal* 114:223–246.

Smith, Michael A., and Leighton Vaughan Williams. 2008. Betting exchanges: A technological revolution in sports betting. In *Handbook of Sports and Lottery Markets*, edited by Donald B. Hausch and William T. Ziemba. Amsterdam: Elsevier, 403–417.

Smith, Michael A., David Paton, and Leighton Vaughan Williams. 2006. Market efficiency in person-to-person betting. *Economica* 73(292):673–689.

Snowberg, Erik, and Justin Wolfers. 2010. Explaining the favorite-longshot bias: Is it risk-love or misperceptions?" *Journal of Political Economy* 118(4):723–746.

Vaughan Williams, Leighton. 1999. Information efficiency in betting markets: A survey. *Bulletin of Economic Research* 51(1):1–30.

Woodland, Linda M., and Bill M. Woodland. 1994. Market efficiency and the favorite-longshot bias: The baseball betting market. *Journal of Finance* 49(1):269–279.

SECTION III

HORSE RACE BETTING

This page intentionally left blank

CHAPTER 13

ON THE LONG-RUN SUSTAINABILITY OF TOTE BETTING MARKETS

DAVID EDELMAN

1 Background—Pari-Mutuel or Tote Betting

Pari-mutuel wagering, an ingenious method of organizing the betting on horse races and other predominantly sportsbetting markets, was first invented in the late 1800s by a Catalan entrepreneur named Joseph Oller Roca and has since come to be the mainstay of state-sponsored wagering worldwide. The primary historical reason for this is that, as a system, it is inherently equitable and democratic in the sense that people are rewarded for their financial votes or shares in the winning outcomes, without subjectivity, prejudice, or favoritism, and do not require counterparties with special views or interests, such as bookmakers; neither do they require any risk to be underwritten.

The fact that pari-mutuel markets are particularly well designed to aggregate, reflect, and reveal (Eisenberg and Gale 1959) as well as possibly to distort (Hurley and McDonough 1995) information has led to increasing attention to it in the academic literature. In addition, the simplicity of pari-mutuel markets makes them particularly well suited to such matters as the examination of the role of asymmmetric or "private" information in markets (Koessler, Noussair, and Zielgelmeyer 2008; Ottaviani and Sørensen 2009) and the microstructure of trading in markets generally (Lange and Economides 2005). Of these aspects, the role of market asymmetry will be emphasized here, though (unlike as in most of the literature) not from a behavioral or particularly analytical way, but rather a pragmatic way, exploring the potential consequences of worst cases to the functioning and future of real-world pari-mutuel markets.

Mathematically, a pari-mutuel system may be described as follows: bettors allocate amounts to their various choices, and when the result is known, all bettors having made the correct choice receive a share of the pool in proportion to their wagers on the winning combination. For instance, if a pool is 100 units, divided (for simplicity) among two choices by A:60 and B:40, each unit wagered on A would return $(100/60) = 1.67$, if A wins, whereas if B wins the unit bet on B would return $(100/40) = 2.50$.

Of course, pari-mutuel operators do not pay out all of the money wagered, so that in this example, if the take is 20 percent the dividend for a unit bet on A would be $.8(1.67) = 1.33$ and for a unit bet on B would be 2.00.

In totalizator or tote markets throughout the world, the amounts returned per unit bet are displayed continuously throughout betting until a race begins, and the clever player will compare his or her estimated probabilities to the prices on offer and decide which, if any, of the possibilities merits a wager. More sophisticated players will, of course, consider the impact of their individual wagers on the dividend so as to not over bet and reduce the dividend unduly. (It is this fact that we shall refer to later as being a key stabilizing mechanism of pari-mutuel betting.)

More specifically, the rational player should take risk into account when deciding the amount to wager. However, in his paper on optimal staking, game theory expert Rufus Isaacs (1953) evaluated the case in which the primary consideration in deciding how much to bet is "pool impact" (this corresponds to the case in which the size of the pool is of a much smaller order of magnitude than the bettor's wealth).

With a significant takeout, one may wonder why anyone would want to participate in these markets. An explanation of that is beyond the scope of this discussion, but clearly if the takeout were increased to too large a level turnover would drop quickly. It is the firmly held belief (myth?) of many (most?) tote operators and governments that the primary threat to their long-term sustainability would be if expert players were to become too plentiful, thereby raising the effective takeout rate for the average player further and further until the experts were taking most of the money and the more casual players were increasingly driven away.

It is the primary purpose here to dispel this impression and to prove the limited potential impact of expert players using simple economic and mathematical arguments. In fact, it will be shown that under broad assumptions the opposite of this hypothesis is true; namely, the more expert players there are in a tote market, the less able they will be to profit collectively from it, given a fixed number of others.

2 A Brief Note on Information Asymmetry and Sophisticated Play

For reasons primarily of simplicity and tractability, much of the published literature in finance and economics historically has focused on the abstraction of "homogeneous agents" (or, by extension, "representative agents"), where all participants in a market are assumed to have (or may be considered to have) similar risk-return preferences and (more importantly) similar access to relevant information and information processing. Generally the assumption of "efficient markets" has pervaded, with a strong burden of proof placed on any attempt to controvert it. We do not attempt here to prove inefficiency in any market; we merely ask the reader to imagine what would happen if the usual tenets and ramifications of efficiency failed to hold in a pari-mutuel market.

More recently increasing attention has been paid to hypothetical situations such as this, where customary notions of market efficiency fail to hold. In this case, we will suppose that information asymmetry exists. The simplest formulation of this might be expressed in terms of information distance (see, e.g., Edelman 2000). Suppose a "true" (experts') distribution of outcome $\mathbf{p}$ and a "casual" distribution of outcome $\mathbf{q}$ exist. Then information asymmetry is tantamount to the condition that

$$D(\mathbf{p}\|\mathbf{q}) \sum_i p_i \log\left(\frac{p_i}{q_i}\right) > 0.$$

(Note that the coincidence of "true" and "expert" is not necessary. One may imagine an "expert" distribution $\mathbf{r}$ as separate from the "true" distribution $\mathbf{p}$, in which case, the asymmetry condition would be that $D(\mathbf{p}\|\mathbf{q}) - D(\mathbf{p}\|\mathbf{r}) > 0$. However, for simplicity, we assume here that $\mathbf{r} \equiv \mathbf{p}$.)

For horse racing markets, the primary source of inefficiency is arguably the processing of publicly available information, including the collection of data in electronic form and the mathematical and statistical modeling which are foreign to the lion's share of casual bettors.

3 Limited Capacity for Sophisticated Play

In general, casual bettors for a horse race bet in the 20 minutes to half hour preceding a race, whereas expert players (or sophisticated players as they are often called) will, either by electronic or other means, have the capability of betting just prior to race start and later than most of the casual players.

We begin by imagining a "best-case" scenario for the sophisticated player in which all other wagers occur previous to his or her wagers and in which he or she is the lone sophisticated player in the market. Then, following Isaacs (1953), we imagine the example discussed earlier but where the expert has an opportunity to send one last wager, supposing that his or her probability for horse (outcome) B is 0.60. We shall, as discussed previously, suppose the expert to be a risk-neutral (expectation-seeking) investor who would like to maximize the total expected amount of return from wagering an amount b, which is

$$b\left\{.6\frac{(.8)(100+b)}{40+b}-1\right\},$$

which is found by simple calculus to be maximum for $b \simeq 7.07$, achieving a value of approximately 0.65 units (i.e., nearly a 10% margin on 7 units). Crucially, if b were increased beyond this, the expectation would decrease, to 0.64 if 8 units were bet, down to 0.56 for 10 units, down to a loss of 0.09 if 16 units were bet.

On average, then, the other (here, casual) players can expect to receive back in aggregate an amount of

$$80 - 0.65 - (.2)(7.07) = 77.94$$

out of their 100 units of wagers instead of 80 without the expert betting. In effect, then, for similar races the expert has the effect of increasing the effective takeout rate for the casual player from 20 to 22.06 percent.

If the takeout rate were small, this best case for the sophisticated player would correspond to the worst case for the casual player. In this case 20 percent is a significant percentage; hence this equivalence will not necessarily hold, for if the sophisticated player bet 20 units, then (according to the 0.6 probability assumed here) the return to the public would be 76.80 (or a 23.2% takeout), a somewhat worse outcome for the casual player than the case of optimal betting. From the casual players' point of view, this latter outcome represents a worst-case scenario irrespective of how many experts have access to the same information advantage and late-wagering capability.

It seems unreasonable to suppose that, collectively, sophisticated players will tend to behave in a manner that is close to their collective optimum; it is rather more likely that the reverse will be true, that the more players there are, the more chance there will be of their incorrectly anticipating the play of others like themselves, leading to overbetting (or, conceivably, underbetting) in relation to the optimum which would be bet in a concerted effort. Thus for a given level of information asymmetry the more (uncoordinated) expert players there are, the better it is for the casual player.

To fix ideas, suppose a pari-mutuel has a 10 percent takeout rate and a group of sophisticated players has an information advantage suggesting to them that in a particular case the probability of a betable event is 25 percent higher than the public proportion. Then for these values it is possible to gauge the impact of sophisticated play for single (or, approximately, collectively for many) players as a function of market proportion. The resulting values can be seen in table 13.1, which presents computed values of a, the optimal proportion of the pool which would be bet, as a function of p the market proportion and the resulting total expectation $E(R)$ (equivalently, total pool Impact) as a proportion of the pool (where basis points "b.p.," or hundredths of a percent, are used for the smaller percentages).

For instance, if 10 percent (1 in 9) of the casual bettors favor a certain outcome, and the sophisticated player(s) is (are) correct in the assumption of the truer proportion being 12.2 percent (1.25 in 9), the most that may collectively be extracted from the

Table 13.1 Maximum Impact as a Function of Market Proportion

p	a	$E(R)$
.01	.0007	0.5 b.p.
.05	.0035	2.4 b.p.
.1	.0067	4.0 b.p.
.2	.012	5.3 b.p.
.3	.014	4.4 b.p.
.4	.013	2.4 b.p
.5	.006	<1 b.p.

pool, on average, based on this information is 4 units in 10,000. It appears, then, that possession of information increasing odds by 25 percent can only result in (at most) 5 basis points (the worst case here occurring when the market proportion is 20%) or one-twentieth of one percent of the total pool. (This may be comforting to tote operators who worry about the influence of sophisticated play in their markets.) The key point here, following from consideration of pool impact, is that for a given level of information asymmetry the downside effect to the casual player is limited in aggregate (and likewise the collective upside to the sophisticated players).

4 Conclusions

Using a simple mathematical example and a theoretical best-/worse-case argument, we have established that the discouraging effect of sophisticated play to the casual bettor must be limited, even in a worst-case scenario, from the point of view of the casual player. Further, even this theoretical worst case is unlikely to be achieved due to the friction between expert players.

While the abstraction of sophisticated players being homogenous (with regard to the nature and quantity of their information advantage), used here for simplicity, could be challenged, it is clearly also a worst case, since any disagreement between experts would lead to a fixed level of aggregate losses among the experts in virtually every race, even if their respective information advantage levels (as measured by $D(\mathbf{p}\|\mathbf{q}) - D(\mathbf{p}\|\mathbf{r})$) were all positive. Although we have established the above in a conceptual sense, the actual behavior of a system with many experts attempting to profit from a pari-mutuel pool over time would be interesting to attempt to model, as lower profit levels would lead to experts leaving the market but then perhaps returning periodically whenever the level of expert betting was perceived to be lower.

It is hoped that the arguments presented here, however simplistic, may prove useful in the understanding of pari-mutuel markets (even to the extent of preventing panic in some quarters) and help industry decision makers refocus their energies toward addressing other factors that may be affecting their business. For instance, might not the fact that the entertainment industry (including the sporting events on which these tote markets are based) has undergone a digital revolution over the past 10 years be the more likely dramatic source of disruption than the manner in which the role of expertise in tote betting markets may have been playing out over the past 100 years or so?

Acknowledgment

This chapter has emanated from research conducted with the financial support of the Science Foundation Ireland under Grant No. 08/SRC/FM1389.

References

Edelman, David. 2000. On the financial value of information. *Annals of Operations Research* 100(1–4):123–132.

Eisenberg, Edmund, and David Gale. 1959. Consensus of subjective probabilities: The pari-mutuel method. *Annals of Mathematical Statistics* 30(1):165–168.

Hurley, William, and Lawrence McDonough. 1995. A note on the Hayek hypothesis and the favorite-longshot bias in parimutuel betting. *American Economic Review* 85(4):949–955.

Isaacs, Rufus. 1953. Optimal horse race bets. *American Mathematical Monthly* 60(5):310–315.

Koessler, Frédéric, Charles Noussair, and Anthony Zielgelmeyer. 2008. Parimutuel betting under asymmetric information. *Journal of Mathematical Economics* 44(7–8):733–744.

Lange, Jeffrey, and Nicholas Economides. 2005. A parimutuel market microstructure for contingent claims trading. *European Financial Management* 11(1):25–49.

Ottaviani, Marco, and Peter Norman Sørensen. 2009. Surprised by the parimutuel odds?" *American Economic Review* 99(5):2129–2134.

CHAPTER 14

THE ECONOMICS OF RACETRACK–CASINO (RACINO) GAMBLING

RICHARD THALHEIMER

1 Introduction

The inclusion of casino-style gambling (henceforth gaming) at pari-mutuel racetracks, beginning in 1990, has changed the landscape of the pari-mutuel racing and wagering industry in the United States.[1] To investigate the implications of racetrack–casino (henceforth racino) betting on the pari-mutuel racing and wagering industry, this chapter is organized as follows. Section 2 examines long-run trends in pari-mutuel horse race wagering. Section 3 reports results of economic studies of the effects of competition from state lotteries and casino gaming on the long-run trends in pari-mutuel wagering. Section 4 considers the rationale for legislation enabling casino-style gaming at racetracks and on various characteristics of racinos in the United States. Section 5 reports key findings from economic studies of the demand for racino wagering. Section 6 presents a summary and conclusions.

2 U.S. Pari-Mutuel Horse Race Wagering—Historical Trends

A review of trends in horse race wagering (handle, turnover), the predominant form of pari-mutuel wagering in the United States,[2] is helpful in understanding the background against which pari-mutuel racetracks have been permitted by law to have casino-style gaming. As of 2010, some form of pari-mutuel horse race wagering was conducted in 38 states. Live horse racing was conducted in 34 of these states.[3] Betting on live races simulcast from racetracks conducting live races to other pari-mutuel wagering facilities was first permitted in 1983. Simulcasting is the process by which live races held at one racetrack, the host track, are transmitted simultaneously to other locations that allow patrons at the receiving racetrack or off-track betting (OTB) facility, the guest location, to place wagers on the races transmitted by the host track. The host track charges the guest facility a fee based on the amount wagered on its races at that location.

U.S. horse race wagering on all racehorse breeds (thoroughbred, harness, quarter horse) was $3.4 billion in 1960 and had increased to $10.8 billion in 2010 (Association of Racing Commissioners International 2010). Adjusted for inflation, a very different picture of the historical trend in horse race betting in the United States emerges. Horse race wagering from all sources (live and simulcast) peaked in 1977 and declined 69 percent over the next three decades through 2010. Over this same period, wagering at racetracks on live horse racing as a percentage of live and simulcast wagering declined from 100 percent of the total prior to 1983, when simulcast wagering was first introduced in New Jersey, to only 12 percent in 2010 (Association of Racing Commissioners International 2010). The decline in total all-source horse race wagering has occurred despite the introduction of market expanding measures such as simulcast wagering between racetracks or to OTB facilities, and telephone and internet account wagering.

Why then, in light of the market-expanding measures instituted over the past several decades, has inflation-adjusted horse race wagering fallen 69 percent? A large part of the answer lies in the introduction and growth of competition from other types of gaming, specifically casinos and state lotteries. The decrease in wagering from increased gaming competition has more than offset any increase due to the market-expanding measures undertaken by the racehorse industry. Competition from casino gaming has been found to reduce pari-mutuel horse race wagering on the order of 31 to 39 percent. Similarly, the competition from state lotteries has been found to reduce pari-mutuel horse race wagering on the order of 10 to 36 percent. For a review of the economic literature on the effect of competition on horse race wagering see Thalheimer and Ali (2008a).

3 Background of Racino Gaming

Concern for the decline in handle and revenue of the pari-mutuel horse racing industry, brought about by competition from other forms of gaming, has prompted a number of states to permit casino-style gaming at pari-mutuel racetracks. A consistent theme of the legislation in most states permitting casino-style gaming at racetracks is preservation of horse and greyhound racing and breeding industries in light of competition from other forms of gaming, such as state lotteries and casinos. Also mentioned in many of the state racino statutes is the preservation of live racing, economic development, tourism, job creation, and the agricultural nature of the racehorse industry. An important aspect of enabling legislation in racino states is that in addition to the allocation of a percentage of gaming revenue to state and local governments and the racino operator, a percentage of gaming revenue is allocated also to purses for horsemen with horses at the racino.[4]

Following is a discussion of important aspects of racinos in the United States.

Racino types. In a number of racino states gaming devices at racinos are operated under the auspices of the state lottery and are placed under the lottery's regulatory authority. In those instances the gaming devices are referred to as video lottery terminals (VLTs). In other racino states, the privately owned gaming devices at racinos are placed under the regulatory authority of a state racing or gaming commission and are referred to as electronic gaming devices (EGDs) or slot machines. In either case VLTs or slot machines are equivalent casino-style gaming devices and are transparent to the customer. The number of slot machines (VLTs) at racinos ranges from around 250 to 5,000, depending on state regulations and market demand (McQueen 2009).

Initially all racinos were limited by legislation to slot machines (VLTs) and were not permitted to have casino-style table games, such as blackjack, roulette, and craps. This restriction was lifted beginning with the Iowa racinos in 2004. Following Iowa's lead, table games were permitted at racinos in West Virginia (2007), Delaware (2009), and Pennsylvania (2010). In addition to table games, the Delaware racinos were also permitted to offer limited sports betting in 2009. Delaware is the only state outside of Nevada to permit sports betting at gaming facilities.

Racino launch dates and locations. As shown in table 14.1, the number of states that permitted casino-style (racino) gaming at pari-mutuel racetracks in chronological order of launch date has grown from 1 in 1990 to 15 in 2011. There were 51 racinos in those 15 states.

4 The Economics of Racino Gaming—Empirical Studies

Racino wagering is a special case of live race and casino-style wagering where the two products are offered at the same location. There have been a number of non-racino studies of the demand for pari-mutuel horse race wagering. These studies include: Ali and Thalheimer (1997, 2002), Church and Bohara (1992), Coate and Ross (1974), Degenarro (1989, 2009), Gramm et al. (2007), Gruen (1976), Morgan and Vasche (1979, 1982), Simmons and Sharp (1987), Suits (1979), and Thalheimer and Ali (1992, 1995a, 1995b, 1995c). In addition to studies of the demand for pari-mutuel horse race wagering, David Forrest, O. David Gulley, and Robert Simmons (2010) examined the determinants of the demand for non-pari-mutuel (fixed-odds) bookmaker betting on horse and dog racing in Britain. Finally, Paton, Siegel, and Vaughan Williams (2004) examined the demand for bookmaker betting on the aggregate of horse and dog

Table 14.1 U.S. Racino States, 2011

State	Type	Under State Lottery?	Launch Date	Number of Racinos	Breed (H, G)*
West Virginia[1]	VLT's/Tables	Yes	1990/2007	4	H,G
Rhode Island[2]	VLT's	Yes	1992	2	J,G
Iowa	Slots/Tables	No	1995/2004	3	H,G
Delaware[3]	VLTs/Tables/Sports	Yes	1995/2009/2009	3	H
New Mexico	Slots	No	1998	5	H
Louisiana	Video Poker/Slots	No	1992/2002	4	H
Alabama[4]	Slots	No	1993	2	G
New York	VLTs	Yes	2004	9	H,G
Maine	Slots	No	2005	1	H
Oklahoma	Slots	No	2005	2	H
Florida[5]	Slots	No	2006	5	H,G,J
Arkansas[6]	Slots	No	2006	2	H,G
Pennsylvania	Slots/Tables	No	2006/2010	6	H
Indiana	Slots	No	2008	2	H
Maryland	VLTs	Yes	2011	1	H

Note: Ohio permitted VLT gaming at its seven racetracks in 2011. One racino was launched in 2012. Table games were permitted at the Maine racino in 2012. Florida has card rooms at all pari-mutuel racetracks. Minnesota has card rooms but no slot machines at racetracks. Kansas permits racino gaming but had not received a license application through 2011. South Dakota, Oregon, and Montana each have limited gaming at many locations (racetracks and other). Canada has extensive racino gaming.

Racino gaming revenue, which was nonexistent prior to 1990, had grown to $6.4 billion by 2009 (American Gaming Association 2010).

[1] Limited number of VLTs (video lottery terminals) at Mountaineer racetrack on an experimental basis in 1990. In 1994 statutes were enacted permitting VLTs at all four West Virginia racetracks.
[2] Simulcast betting only in 2011. There was no live greyhound racing or jai alai at the two pari-mutuel facilities in 2011.
[3] Table games and sports wagering both were launched in 2009. Delaware racinos are the only ones to also have sports wagering.
[4] Slot machines (Class 2 Indian electronic gaming machines) at racetracks under local authority.
[5] Slot machine racinos in Broward and Dade counties only. Card rooms also permitted at all pari-mutuel racetracks and jai alai *frontóns* in the state.
[6] Arkansas has Electronic Games of Skill (EGS).

Source: Thalheimer Research Associates (TRA, Inc.) and state racing and gaming commissions.
* (H = Horse, G = Greyhound, J = Jai Alai)

racing, sports, and slot machines (fixed-odds betting terminals) at betting shops in the United Kingdom. There has been only one non-racino study of the demand for casino wagering (Thalheimer and Ali 2003). The demand variable in this study was aggregate slot machine wagering for riverboat casinos and racinos in the Midwestern states of Illinois, Iowa, and Missouri.[5]

While there have been many papers on the demand for pari-mutuel horse race wagering, only three studies have examined the demand for pari-mutuel wagering when slot machine (VLT) wagering is offered at the same location. A discussion of these papers follows.

Racino study 1. The subject of the earliest racino study (Thalheimer 1998) was Mountaineer Racetrack and Gaming Resort (Mountaineer) in West Virginia. Mountaineer was the first racetrack in the United States to offer VLT (slot machine) gaming. The VLTs were placed under the auspices of the West Virginia Lottery with Mountaineer acting as agent for the lottery.

The daily sample period for this study was 1989 through 1991, during which live racing was also offered at the track. A limited number of VLTs were placed at the thoroughbred racetrack in June 1990 as an experimental pilot project under the auspices of the state lottery. The number of VLT's was increased from 70 to 150 by 1991. Wagering at the racetrack on simulcasts of an entire day's program of races (whole card simulcasts) imported from a few racetracks around the country was launched at Mountaineer in September 1990.

Three wagering demand models were estimated, one for live horse race wagering, one for import simulcast horse race wagering, and one for VLT wagering. Each of the three demand models was specified as a function of its own-demand variables as well as those of the other two wagering products. The same set of live race, import simulcast race, and VLT variables was used in each of the three wagering demand models. This setup resulted not only in estimates of relationships of each of the three wagering product demands to its own-demand variables but also to the variables associated with the demands of the other two wagering products. The price of wagering variables, takeout rate for pari-mutuel wagering, and win percentage for VLT wagering were constant over the study period and so were not included in the demand models.

An important finding of the study involved the interaction between the racing and gaming sides of the racino. On the racing side pari-mutuel wagering from live and import simulcast racing combined was found to have declined 24 percent as a result of VLT gaming at Mountaineer. On the gaming side customers who wagered on horse racing were also found to wager on the VLTs. Therefore, while existing horse race customers were found to reduce their wagers on horse racing with the introduction of the VLTs, new customers attracted by the VLTs, on net, did not bet on the horse races. Wagering on the VLTs from existing and new customers was enough to more than offset the estimated reduction of pari-mutuel wagering and to add an additional 21 percent to total wagering.

The change in revenue tells a different story, however. Revenue from pari-mutuel wagering is computed as the amount wagered multiplied by the price of wagering, the takeout rate. The takeout rate is set by state statute as a percentage of total wagering and is deducted from each bet before any winnings are returned to the patron at the conclusion of each race. Revenues from the takeout rate are distributed to the state (pari-mutuel tax), racetrack (commissions), and horsemen with horses at the racetrack (purses). The effective takeout rate for live and simulcast wagering at Mountaineer was approximately 20 percent. The price of VLT (slot machine) gaming is the win percentage. While the takeout rate is known and set by state statute, the win percent for VLT (slot machine) wagering is not a fixed deduction from each wager. Rather it is based on the long-run probability of winning and the payoffs associated with that probability. The expected payoff for lottery-run video gaming terminals is set by state statute and determined by a random number generator. The effective win percentage is computed as total wagering less the amount paid out in winnings divided by total wagering. The win percentage for the VLTs at Mountaineer Park was 12 percent over the study period.

The revenue per dollar wagered on the VLTs at 12 percent was much lower than the 20 percent from pari-mutuel wagering, which it replaced. As a result, total revenue increased only 3 percent relative to the 21 percent increase in total wagering from betting on the VLTs. The findings of the study suggest that if the number of VLTs is too small VLT revenue may not be large enough to offset the decline in pari-mutuel revenue. As the number of VLTs was increased, VLT wagering was found to increase more than the decrease in pari-mutuel wagering.

Racino study 2. Mountaineer was also the subject of a second racino study (Thalheimer 2008). The pilot project begun in June 1990 with a limited number of VLTs at the racetrack was deemed a success, and in March 1994 the West Virginia Video Lottery Act (the Act) was passed. The Act gave statutory authority to the West Virginia Lottery to conduct VLT gaming at all four (two thoroughbred and two greyhound) West Virginia pari-mutuel racetracks, subject to local referendum. The temporary pilot VLT project at Mountaineer was ended on September 2, 1994. At that time VLT gaming was launched at the racetrack on a permanent basis, subject to the requirements of the newly enacted state statutes. The Act permitted more and different types of VLTs than the limited number and types permitted during the pilot project. The study period included 443 weeks from July 1994 through December 2002. Market area population was 8.4 million over the study period.

Annualized, total VLT wagering increased from $131 million to $2.6 billion over the first through the last year of the study period. At the same time, total on-track pari-mutuel wagering increased minimally from $38.2 million to $39.4 million. Adjusted for inflation, VLT wagering increased 1,559 percent compared to a 13 percent decrease in pari-mutuel wagering. Over this period the relative importance of pari-mutuel wagering fell from 23 to 2 percent of total pari-mutuel plus VLT wagering. In 2002 pari-mutuel and VLT revenues were $16 million and $221 million, respectively. Pari-mutuel revenue was 7 percent of total revenue.[6]

Two weekly wagering demand models were estimated, one for on-track (live plus import simulcast) pari-mutuel wagering and one for VLT wagering.[7] VLT gaming was conducted daily year-round. Live racing was conducted every month of the year and varied from 0 to 7 days per week, averaging 4.4 days per week or about 220 days per year. Import simulcast wagering was conducted every week over the year and varied from 5 to 7 days per week and from 2 to 21 racetrack programs per day. Following the earlier study of Mountaineer (Thalheimer 1998), each of the wagering demand models was specified as a function of its own-demand variables as well as those of the other wagering product. The same set of pari-mutuel and VLT variables was used in each of the two wagering demand models. Variables for the price of pari-mutuel wagering (takeout rate) and the price of VLT wagering (win percentage) were constant over the study period and so were not included in the demand models.

As in the earlier study of Mountaineer (Thalheimer 1998), this study examined the interaction of the pari-mutuel and gaming sides of the racino. In making this determination, VLT demand variables were constructed to measure the effect of government regulations (restrictions) on VLT and pari-mutuel wagering. This was made possible by the provision by racino management of a unique dataset on the racino floor layout of the number and types of VLTs at Mountaineer at various times over the study period. Restrictions imposed on the VLTs under the statutes included: the number of VLTs, the type of game and machine, the maximum bet-per-play, and the location of the VLTs between the racino's hotel and the racetrack. The 1994 statutes that permitted a maximum of 400 VLTs were subsequently amended seven times, each time increasing the maximum number of VLTs allowed at the racino. As a result, the number of VLTs was increased from 165 before termination of the pilot project to 3,000 by the end of the study period.

In addition to the number of VLTs, the relationship of VLT wagering demand to the type of VLT was examined in the study. The provisions of the pilot project prior to the 1994 Act stipulated that each of the first generation VLTs be restricted to offering a single game (card, keno, or slot). The 1994 Act provided for technologically advanced multi-game VLTs without a video slot machine option. The Act was amended in 1996 to permit a video slot machine option. There were subsequent amendments to the Act permitting other game types, such as Las Vegas–type slot machines. Several other restrictions imposed on the VLTs by the 1994 Act were examined in the study, including a bet limit of $2 per play, which subsequently was raised to $5, and a restriction on the placement of the VLTs between the racino's hotel and its nearby racetrack, which was changed several times during the study period.

The number of VLTs, the distribution of existing and new VLT types relative to the pilot period first-generation single-game VLTs, the placement of the VLTs, and restrictions on VLT maximum bet-per-play all were found to be significant with respect to their effect on VLT wagering. The relaxation of restrictions on the number of VLTs leading to their increase to 3,000 over the study period resulted in a 107 percent increase in VLT wagering. The relaxation of restrictions on the type of machines that could be included in the product mix increased VLT wagering by 178 percent. Relaxing restrictions on VLT placement between the hotel and racetrack increased VLT wagering by 23 percent. Finally, increasing the maximum bet-per-play from $2 to $5 increased VLT wagering by 11 percent. The joint or combined increase in VLT wagering resulting from relaxing all of these VLT restrictions was 687 percent over the study period. These results demonstrate that allowing management to more freely react to changes in consumer demand by lifting government restrictions on each of the factors limiting VLT play resulted in large increases in VLT wagering.

The interaction of the horse racing product with VLT wagering demand also was found to be significant, as in the earlier study of Mountaineer (Thalheimer 1998). VLT wagering was found to be positively related to live race days and the number of import simulcast racetracks per day. VLT wagering was not found to be significantly related to average purse or stakes races of national or regional importance.

On the racing side of the racino pari-mutuel wagering demand was found to be positively related to all of its own-demand variables, including the number of live race days per week, the number of import simulcasts per day, and the average daily purses and high-quality stakes races. The live race average purse elasticity was 0.1, indicating that total (import simulcast plus live) wagering demand is highly inelastic with respect to a change in live race purse. The finding of inelastic demand with respect to purses has also been found in Ali and Thalheimer (2002) and Gramm et al.(2007). The presence of stakes races of national and regional importance in a given week was found to have large positive impacts on pari-mutuel wagering. As in the earlier study of Mountaineer (Thalheimer 1998), pari-mutuel wagering was found to decrease with the introduction and growth of VLTs. The growth in the number of VLTs to 3,000 by the end of the study period resulted in a 39 percent reduction in pari-mutuel wagering. Lifting restrictions on the placement of the VLTs at the hotel versus the racetrack resulted in a 15 percent reduction in pari-mutuel wagering. Finally, raising the maximum VLT bet from $2 to $5 resulted in a 3 percent reduction in pari-mutuel wagering. The joint effect of relaxing government restrictions on the number of VLTs, their location, and the maximum bet-per-play was to reduce on-track pari-mutuel wagering by 49 percent.

The study also addressed the important policy issue of the effectiveness of subsidizing the racehorse industry by dedicating a statutory percentage of VLT revenues to purses. The intent of the state legislature to preserve the racing industry is clearly stated in the West Virginia Video Lottery Act of 1994 as follows (Chapter 29, Article 22 A-2):

> The Legislature finds and declares that the existing pari-mutuel racing facilities in West Virginia (horse and greyhound) provide a valuable tourism resource for this state and provide significant economic benefits to the citizens of this state through the provision of jobs and the generation of state revenues; that this valuable tourism resource is threatened because of a general decline in the racing industry and because of increasing competition from racing facilities and lottery products offered by neighboring states; and that the survival of West Virginia's pari-mutuel racing industry is in jeopardy unless modern lottery games are authorized at the racetracks.

It should be noted that while purses funded by VLT revenues are an expense to the racino operator, they constitute revenue to racehorse owners and breeders. VLT revenues to the racetrack operator fund jobs at the racetrack–casino. Purses from VLT revenues provide revenue to racehorse owners and to racehorse breeders to fund their operations.[8] These racehorse industry businesses create jobs associated with the care and maintenance of horses that compete for purses at the racetrack and reside on breeding farms in the state.

The analysis compares the cost of purses to the revenues they generate. Purses were $39 million in 2002, $31 million from VLT revenues and $8 million from pari-mutuel wagering. The average daily live race purse, which measures the quality of racing, increased from $23,500 to $163,100 over the study period as a result of revenues from increased VLT wagering. Adjusted for inflation, average daily purses increased 481 percent. As a result of the highly inelastic purse elasticity, pari-mutuel revenue increased only $1 million. The study also found that VLT wagering did not increase significantly with an increase in purses. For these reasons, the large increase in average purses would not generate sufficient additional on-track pari-mutuel or VLT revenue to offset their cost.

There is, however, another effect of the VLT purse subsidy that cannot be overlooked. Competitive purses are necessary to attract horsemen in order to conduct live racing at the racetrack. Without the VLT purse subsidy, the level of purses from pari-mutuel wagering alone would not have been sufficient to fund horsemen's operations at Mountaineer, especially on a year-round basis as required under the statutes. The absence of live racing at the racetrack would result in a loss of $12 million in live race pari-mutuel wagering revenue, including revenue from wagering on export simulcasts of live races at Mountaineer to other in-state and out-of-state locations.[9] An important finding of this study was that VLT wagering increases on days when there is live racing. If live racing were to cease there would be an estimated 18 percent loss in VLT wagering with a corresponding $40 million VLT revenue loss on total VLT revenues of $221 million.[10] Assuming that live racing is preserved by the VLT purse subsidy of $31 million, VLT and pari-mutuel revenues it generates would more than cover the cost of purses and contribute to racino operating costs.

Racino study 3. Prairie Meadows Racetrack and Casino (Prairie Meadows), a horse race racino in Iowa, was the subject of the third racino study (Thalheimer 2012). Thoroughbred, quarter horse, and harness horse wagering were all conducted at the racetrack. The study period for analysis of the demand for pari-mutuel horse race wagering at Prairie Meadows included the 168 months from 1993 through 2006. The market area population was 1.6 million in 2006.

Slot machine gaming was introduced at the racetrack in April 1995. The study period for analysis of the demand for slot machine wagering at Prairie Meadows included the period from April 1995, with the introduction of slot machines, through December 2006. The number of slot machines was increased from 1,100 to 1,600 over the first and last year of the study period. A major focus of this study was the effect on pari-mutuel and slot machine wagering demands due to the expansion of the gaming product at Prairie Meadows to include casino-style table games such as poker, blackjack, craps, and roulette in December 2004. Iowa was the first state to permit table games at racinos.

Slot machine wagering increased 72 percent, from $1.5 billion to $2.6 billion from the first year to the last year of study period. Adjusted for inflation, slot machine wagering increased 31 percent over the study period. Adjusted for inflation, on-track pari-mutuel wagering, which was $17.8 million in 2006, fell 47 percent from its 1993 level. The number of live race days varied from 54 to 118 and was 95 in 2006. Import simulcasting was year-round at the racetrack, and the number of import simulcast racetrack programs per day varied from 2 to 16 over the study period. Due to the statutory allocation of slot machine revenues to purses, average daily purses increased from $23,500 in 1993 to $163,100 in 2006.

Following Thalheimer (2008), two wagering demand models were specified, one for slot machine wagering and one for on-track (live plus simulcast) pari-mutuel horse race wagering. Each of the wagering demand equations was specified as a function of its own-variables as well as those of the other wagering product. The same set of pari-mutuel and slot machine variables was used in each of the two wagering demand models. The price of pari-mutuel wagering (takeout rate) was constant at 21 percent over the study period and so was not included in the demand models. The price of slot machine wagering (win percentage) varied from 6 to 8 percent over the study period.

The interaction of racing and slot machine gaming at the racetrack was consistent with the findings of the two earlier racino studies (Thalheimer 1998, 2008). On the racing side of the racino operation, pari-mutuel wagering was found to decrease 21 percent after slot machines were introduced. For the first time in the literature, the effect of table games on pari-mutuel wagering at a racino was determined. Pari-mutuel wagering was found to decrease 16 percent after the introduction of table games. In total, pari-mutuel wagering decreased 34 percent as a result of the combined effect of slot machines and table games at the racino.

On the gaming side of the racino operation, slot machine wagering declined 8 percent following the introduction of table games at the racino. This result was consistent with a study of casino gaming in the Midwest (Thalheimer and Ali 2003) in which slot machine wagering at the casinos was found to decrease 13 percent as a result of table gaming. As in the earlier racino studies (Thalheimer 1998, 2008), slot machine wagering demand was found to be positively related to pari-mutuel wagering demand variables. In particular, slot machine wagering increased 13 percent in the presence of live horse racing and 14 percent in the presence of import simulcast horse racing. For comparison, a study by the Pennsylvania Gaming Control Board (2010) concluded that on days when there was live racing at the state's racinos slot machine revenue was 15 percent greater than on non-racing days. Slot machine wagering was also found to be positively related to stakes races but not to average daily purses.

Among other findings of the study, pari-mutuel wagering was found to be positively related to the number of live horse race days, the number of import simulcast racetrack programs per day, and average daily purses and stakes races. Pari-mutuel wagering was found to decrease with competition from nearby casinos. On the gaming side of the racino operation wagering demand was found to be positively related to the number of slot machines and negatively related to the price of wagering (win percentage). Slot machine wagering demand was found to be price inelastic (-0.8) at a win percentage of 8 percent and was found to decrease with competition from a nearby casino.

Finally, the study addressed the important policy issue of the allocation of stakeholder shares from table game and slot machine revenues. The Iowa gaming statutes make no distinction in the allocation of shares of slot machine or table game revenues to the stakeholders: government, racino operator, and horsemen (purses). This is not the case for table game revenues at the West Virginia, Delaware, and Pennsylvania racinos. In each of these states the shares of table game revenue to government and horsemen have been reduced and the share to the racino operator increased relative to their corresponding shares from slot machine revenue. Stakeholders with lower relative table game revenue shares may lose gaming revenue (slot machine plus table game) if table games do not produce enough revenue to offset the expected loss of slot machine revenue.

Racino study—Gaming side only. A fourth racino study is mentioned here in the interest of completeness (Thalheimer and Ali 2008b). This study examined the determinants of demand for slot machine wagering at each of the three Delaware racinos. The study was limited to the slot machine gaming side of the racino operation and so was in effect a casino wagering demand study. A major finding of this study was that slot machine wagering at each of the three racinos was reduced 16 percent as a result of the introduction of a statewide smoking ban.

5 Summary and Conclusions

Adjusted for inflation, pari-mutuel horse race wagering declined 69 percent from its peak in 1978 through 2010. Increasing levels of competition from casinos and state lotteries have been shown to be major causes of this decline. Studies have reported that casino gaming has reduced pari-mutuel horse race wagering in the range of 31 to 39 percent. Similarly, state lotteries have reduced pari-mutuel horse race wagering from 10 to 36 percent. Recognizing the effect of increased competition on the pari-mutuel racing industry, a number of states have enacted laws permitting racetracks to offer casino-style gaming at their facilities. These racetrack-casinos are often referred to as racinos.

A key theme of the enabling legislation in most states permitting casino-style gaming at racetracks is preservation of the racehorse and greyhound racing and breeding industries in light of competition from other forms of gaming, such as state lotteries and casinos. Also mentioned in many of the statutes is the preservation of live racing, economic development, tourism, job creation, and the agricultural nature of the racehorse industry.

Racino gaming was first launched in 1990 at a racetrack in West Virginia. As of 2011, there were 51 racinos in 15 states following passage of legislation permitting casino-style gaming at racetracks, and this number is expected to grow. In a number of racino states, racino gaming devices are operated under the auspices of the state lottery and are placed under the lottery's regulatory authority. In those instances, the gaming devices are referred to as video lottery terminals (VLTs). In other racino states privately owned gaming devices at racinos are placed under the regulatory authority of a state racing or gaming commission and are referred to as electronic gaming devices (EGDs) or slot machines. In either case VLTs or slot machines are equivalent casino-style gaming devices and are transparent to the customer. Initially all racinos were legislatively limited to VLTs or slot machines and were not permitted to have such table games as blackjack, roulette, and craps. This restriction was lifted beginning with the Iowa racinos in 2004. Following Iowa's lead, three additional states have been permitted table games at racinos, and the trend toward permitting racinos to have table games is expected to continue.

Many studies have been conducted on the demand for pari-mutuel horse race wagering, but only three such studies have examined the demand for pari-mutuel and slot machine (VLT) wagering at a single location, that is, a racino (Thalheimer 1998, 2008, 2012). Of particular interest to policy makers and racehorse industry stakeholders is the interrelationship between the pari-mutuel and gaming sides of the racino operation. All of the racino studies found that placing slot machines (VLTs) at horse racing racetracks resulted in significant reductions in pari-mutuel wagering at the racetrack, on the order of 21 to 39 percent. On the other hand, these studies also found that slot machine (VLT) wagering increases significantly on days when there is wagering on live, and/or import, simulcast races at the racetrack. In particular, since enabling racino legislation focuses on the preservation of live racing, it is of interest that slot machine (VLT) wagering was found to increase from 13 to 18 percent on live race days. The conclusion that can be drawn from these results is that the crossover between the racing and gaming sides of the racino is one way, from racing to casino gaming. Existing horse race customers place fewer bets on the races and play the slot machines (VLTs) when they become available. New customers, attracted by the slot machines (VLTs), bet, on net, on the gaming machines and not on the races. It should be noted that pari-mutuel revenue is a very small percentage of total gaming and pari-mutuel racino revenue. For example, consider that in 2002 VLT and pari-mutuel revenues at the Mountaineer Racetrack and Gaming Resort (Mountaineer) racino in West Virginia were $221 million and $16 million, respectively. Therefore, the relatively large percentage reduction in pari-mutuel revenue from slot machines (VLTs) at racetracks results in a small revenue loss relative to the very large gain in slot machine (VLT) revenue from a smaller percentage increase due to the presence of live racing at the racetrack.

Unique to racino legislation is the allocation of a statutorily set percentage of gaming revenue to purses to support racing and breeding operations in the state. With a few exceptions for casinos in two racino states (Maryland and Pennsylvania), this is not the case for casinos in all other states. Purse allocations from racino gaming revenues have resulted in large increases in purses as intended by the enabling legislation even as pari-mutuel wagering revenue has continued to decline. In one racino study (Thalheimer 2008), an analysis was made of revenues generated by purses relative to their cost at Mountaineer. It was shown in this study and others that demand for pari-mutuel wagering was highly inelastic with respect to purses. As a result, even with the large increase in purses from VLT revenue, the revenue from pari-mutuel wagering did not increase sufficiently to cover the cost of purses. On the gaming side of the racino operation, the relationship of VLT wagering to purses was found to be insignificant. As a result of the statistically weak relationship between purses and wagering on both sides of the racino operation, the change in racino revenue with respect to a very large increase in purses was not sufficient to cover a meaningful portion of the cost of purses.

There is, however, another effect of the purse subsidy from gaming revenue that cannot be overlooked. Competitive purses are necessary to attract horsemen in order to conduct live racing at the racetrack. Without the purse subsidy from gaming revenue, the level of purses from pari-mutuel wagering alone would not be sufficient to fund horsemen's operations at Mountaineer, especially on a year-round basis as required under the statutes. If year-round live racing were to have ceased at the racetrack, there would have been an 18 percent loss in VLT wagering and revenue. This loss would have been greater than the cost of purses. This finding was based on the estimated reduction in VLT wagering if live racing were to have ceased. Stated another way, the operation of live racing at the racetrack was found to generate VLT revenue which would more than cover the cost of purses. A smaller amount of pari-mutuel revenue is also generated from wagering on the live races. Total revenue from both sides of the racino operation, therefore, was sufficient to more than cover the cost of purses from VLT revenue and to contribute to the operating cost of the racino operation.

Other findings of racino studies are worth noting. Thalheimer (1998) found that if the number of slot machines (VLTs) is too small gaming revenue may not be large enough to offset the decline in pari-mutuel revenue. Thalheimer (2008) later found that lifting government restrictions on the number and type of gaming machines, the placement of those machines, and bet limits resulted in significant gains in slot machine (VLT) wagering and corresponding reductions in pari-mutuel wagering. These results demonstrate that allowing management to more freely react to changes in consumer demand by lifting government restrictions on slot machine (VLT) play may result in large increases in gaming revenue. Finally Thalheimer (2012) found that adding table games to a slot machine racino reduced pari-mutuel wagering by 16 percent and slot machine wagering by 8 percent.

Based on findings in the literature the outlook for the pari-mutuel racing industry is mixed. On a positive note, the presence of live racing has been found to generate significant racino gaming revenues. However, pari-mutuel wagering has continued to decline, even with large increases in purses funded by racino gaming revenues. Pari-mutuel revenue at racinos is much less than gaming revenue. The long-run viability of the pari-mutuel racing industry will depend on the ability of racing to generate increased revenue from racing customers and thus, cover the cost of racing.

Notes

1. Even though this chapter is restricted to racino gambling in the United States, it should be noted that racino gambling is also permitted throughout Canada. At this writing no academic studies on the economics of racino gaming in Canada have been published.
2. In 2010 greyhound (dog) racing was conducted in seven states and accounted for only 6 percent of total pari-mutuel dog and horse racing (Association of Racing Commissioners

International 2010). Jai alai is conducted only in the state of Florida, and pari-mutuel wagering on this product is minimal.

3. Live greyhound racing was conducted in 6 of the 34 states that had live horse racing and in Alabama where there is no live horse racing but where simulcast horse race wagering does take place.
4. In some racino states a percentage of gaming revenue is also allocated to racehorse (greyhound) breeder awards.
5. A number of other studies on casino gaming have been published, but in each case the dependent variable is casino revenue, which is not a demand variable since it is the product of gaming price (win percentage) and the demand variable, wagering. The price variable is also a demand determinant (independent variable), which results in biased estimates of the revenue model coefficients. The interested reader is referred to Thalheimer and Ali (2003) for a review of the casino revenue studies.
6. Pari-mutuel revenue includes both on-track wagering and revenue from export simulcasts of Mountaineer's live races to other in-state and out-of-state locations.
7. Import simulcast races at the racino included thoroughbred, harness, and greyhound races exported from other in-state and out-of-state racetracks.
8. In an economic analysis of the supply and demand for racehorse breeding stock (yearlings), Neibergs and Thalheimer (1997) found that an increase in expected purses results in an increase in yearling prices (i.e., breeder revenue) and an increase in the supply of horses.
9. Live race revenue equals \$16.2 million on-track wagering times 23 percent of the live race takeout rate plus \$268 million in wagering on export simulcasts of live races at Mountaineer at other in-state and out-of-state betting locations times 3 percent of the wagering on the host racetrack (Mountaineer) races remitted by those locations.
10. VLT revenue equals \$2.6 billion VLT wagering times the 8.5 percent win percentage.

References

Ali, Mukhtar M., and Richard Thalheimer. 1997. Transportation costs and product demand: Wagering on pari-mutuel horse racing. *Applied Economics* 29(4):529–542.

——. 2002. Product choice for a firm selling related products: A pari-mutuel application. *Applied Economics* 34(10):1251–1271

American Gaming Association (AGA). 2010. *2010 State of the states: The AGA survey of casino entertainment*. Washington, D.C.: AGA; http://www.americangaming.org/sites/default/files/uploads/docs/sos/aga-sos-2010.pdf.

Association of Racing Commissioners International. 2010. *Pari-Mutuel wagering 2010: A statistical summary*. Lexington, Ky.: Association of Racing Commissioners International.

Church, Albert M., and Alok K. Bohara. 1992. Incomplete regulation and the supply of horse racing. *Southern Economic Journal* 58(3):732–742.

Coate, Douglas, and Gary Ross. 1974. The Effect of off-track betting in New York City on revenues to the city and state governments. *National Tax Journal* 27(1):63–69.

DeGenarro, Ramon P. 1989. The determinants of wagering behavior. *Managerial and Decision Economics* 10(3):221–228.

——. 2009. New evidence on the link between government subsidies and wagering. *Journal of Gambling Business and Economics* 3(2):37–62.

Forrest, David, O. David Gulley, and Robert Simmons. 2010. The relationship between betting and lottery play. *Economic Inquiry* 48(1):26–38.

Gramm, Marshall, C. Nicholas McKinney, Douglas H. Owens, and Matt E. Ryan. 2007. What do bettors want? Determinants of pari-mutuel betting preference. *American Journal of Economics and Sociology* 66(3):465–491.

Gruen, Arthur. 1976. An inquiry into the economics of race-track gambling. *Journal of Political Economy* 84(1):169–178.

McQueen, Patricia A. 2009. "Challenging times for mature markets. *Casino Journal*, Dec.; http://www.casinojournal.com.

Morgan, W. Douglas, and Jon David Vasche. 1979. Horseracing demand, pari-mutuel taxation and state revenue potential. *National Tax Journal* 32(2):185–194.

———. 1982. A note on the elasticity of demand for wagering. *Applied Economics* 14(5):469–474.

Neibergs, J. Shannon, and Richard Thalheimer. 1997. Price expectations and supply response in the thoroughbred yearling market. *Journal of Agricultural and Applied Economics*, 29(2):419–435.

Paton, David, Donald S. Siegel, and Leighton Vaughan Williams. 2004. Taxation and the demand for gambling: New evidence from the United Kingdom. *National Tax Journal*. 57(4):847–861.

Pennsylvania Gaming Control Board (PGCB). 2010. The economic impact of slot machines on Pennsylvania's pari-mutuel wagering industry—Benchmarking the industry: 2006–2009. Harrisburg, Pa.: PGCB; http://gamingcontrolboard.pa.gov/files/reports/2009_Pari-Mutuel_Benchmark_Report.pdf.

Simmons, Susan A., and Robert Sharp. 1987. State lotteries' effects on thoroughbred horse racing. *Journal of Policy Analysis and Management* 6(3):446–448.

Suits, Daniel B. 1979. The elasticity of demand for gambling. *Quarterly Journal of Economics* 93(1):155–162.

Thalheimer, Richard. 1998. Parimutuel wagering and video gaming: A racetrack portfolio. *Applied Economics* 30(4):531–544.

———. 2008. Government restrictions and the demand for casino and parimutuel wagering. *Applied Economics* 40:773–791.

———. 2012. The demands for slot Machine and pari-mutuel horse race wagering at a racetrack-casino. *Applied Economics* 44(9):1177–1191.

Thalheimer, Richard, and Mukhtar M. Ali. 1992. Demand for parimutuel horse race wagering with specific reference to telephone betting. *Applied Economics* 24(1):137–142.

———. 1995a. Exotic betting opportunities, pricing policies and the demand for parimutuel horse race wagering. *Applied Economics* 27(8):689–703.

———. 1995b. Intertrack wagering and the demand for parimutuel horse racing. *Journal of Economics and Business* 47(4):369–383.

———. 1995c. The demand for parimutuel horse racing and attendance. *Management Science* 41(1):129–143.

———. 2003. The demand for casino gaming. *Applied Economics* 35(8):907–918.

———. 2008a. Pari-mutuel horse race wagering—Competition from within and outside the industry. In *Handbook of Sports and Lottery Markets*, edited by Donald B. Hausch and William T. Ziemba. Amsterdam: Elsevier, 3–15.

———. 2008b. The demand for casino gaming with special reference to the effect of a smoking ban. *Economic Inquiry* 46(2):273–282.

CHAPTER 15

THE MODERN RACING LANDSCAPE AND THE RACETRACK WAGERING MARKET: COMPONENTS OF DEMAND, SUBSIDIES, AND EFFICIENCY

RAMON P. DEGENNARO AND ANN B. GILLETTE

Introduction

In 2012 $11.4 billion was wagered in the United States on horse racing. These dollars affected state revenues, the equine industry, and racetrack income. Competition for the wagering dollar has become increasingly dynamic and intense with the introduction of new wagering and gaming venues. On any given day, a U.S. bettor has the opportunity, via simulcasting and online betting, to place bets on more than one hundred races across different tracks. In addition, interstate horse race wagering has become more competitive and complex across racetracks as state legislatures have approved new betting venues. Legislative prohibitions against most other forms of gambling in the United States protected the industry until the advent of state lotteries (New Hampshire established the first in 1964) and, later, casinos and other gaming. By the end of 2013 it is likely horse racings monopoly on online betting will be eroded away as three states have passed legislation for online gambling by in-state customers and ten others have passed online poker. Now lotteries, casino and sports gambling, and slot machines also compete for that wagering dollar.

Over the years the horse racing industry has benefited from several types of local, state, and federal government subsidies. In recent years, with more and more states having trouble balancing their budgets, these subsidies, funded by varied sources of state tax revenues, casinos taxes, and gaming, are coming under increased scrutiny. Racetrack stakeholders often cite viable returns from such subsidies in terms of higher wagering handles (the total amount of money wagered) and increased employment in the horse and agriculture sectors that eventually funnel increased revenues into state coffers.

As of 2010 statistics compiled unofficially from private data by the Daily Racing Form (DRF) found that racetracks receiving subsidies from slot machines and casinos generated on average only two-fifths of the betting handle that nonsubsidized tracks generated, even after accounting for similar average daily purse distributions (Paulick Report 2011). They also indicated that the off-track simulcast marketplace is becoming more of a driver of handle, which allows small tracks that are strategically positioned in the racing schedule to do relatively well in attracting the wagering dollar. Furthermore they quote Chris Scherf, executive vice president of the Thoroughbred Racing Association, a racetrack trade group, remarked, "The weakness of the subsidized tracks in the pari-mutuel market could lead legislators to cut off the subsidies" using arguments that tracks are not supported by the public. Recently Pennsylvania and Indiana have reallocated some of their horse racing subsidies to compensate for budgetary shortfalls in areas such as education. Other states are considering following suit.

Walker and Jackson (2011), however, found support for horse racing's contribution to state government revenues, relative to other forms of gambling. Using panel data analyses from 1985–2000 across all 50 states and five types of gambling industries, the authors examined the volume of gambling on total state government receipts' net of federal government transfer funding. Accounting for competition across states and both gambling and non-gambling entertainment alternatives for the consumer dollar, they found that for the typical state horse racing and lotteries tended to increase state revenues while casinos and greyhound racing tended to have a negative effect. Walker and Jackson emphasized that their results were for the "average" state since they did not find consistency in the complementary (substitute) relationship among different gaming options across states. These differences across states arise because of varieties in the types of gambling industries within a state, types of non-gambling entertainment industries available, regulations, and relative tax rates on various types of spending.

Another recent study by Forrest, Gulley, and Simmons (2010) using bookmaker data for horse wagering in the United Kingdom also concludeed that bettors view other gambling options, such as dog races, soccer games, and numbers betting, as substitutes. In particular, they found evidence that horse racing bettors are value-sensitive and respond to changes in relative wagering prices and taxes, although the cross-price elasticity effects are hard to measure. Earlier studies looked at the effect of a single gambling industry on horse racing handle and state tax revenues. For example, using data from 1960–1987 of racetracks on the Kentucky–Ohio border, Thalheimer and Ali (1995) found that although the lottery reduces the handle at racetracks it increases overall state tax revenues. The competitive interaction of gambling industries and their consequent effect on state revenues appears to be a fruitful area for further study.

Complicating estimates of the return on racing subsidies for state taxpayers is the fact that the ownership of racetracks has been changing. Private investors are increasingly leasing state-owned tracks and acquiring control of a portfolio of tracks across state lines. This transfer away from state ownership of racetracks has prompted gaming venues, whose state surcharges help subsidize racing, to promote the argument that the horse racing industry should be self-sufficient. For example, in Maryland, this self-sufficiency argument has resonated with the state legislature, which is angry that out-of-state corporations currently own the Maryland Jockey Club. The Maryland legislature (and other states) is legislation that would make it more difficult to provide support for certain players put forth in the industry. After two years of negotiations between track owners, horsemen, and the legislature the horsemen have ceded some of their slot revenue to the track management. These events highlight the political complexities affecting the strategic competition for the racetrack wagering dollar.

The Modern Racing Landscape

The total economic impact of the equine industry in the United States is $101.5 billion. The industry contributes $1.9 billion to government taxes and fees, according to the last commissioned study performed by Deloitte Consulting LLP (2005). The horse racing industry, which includes racetracks and off-track betting operations, horses in training and breeding, contributed $10.697 billion directly to 2005 GDP, with an additional economic impact of $26.1 billion; employed 146,625 individuals in full-time jobs; and indirectly employed 383,826. Currently 32 states allow horse racing. The dominant racing breeds and their relative dollar contributions to horse racing are: Thoroughbred (78%), Quarter horse (6%), and others, including Standardbred or harness racing (16%).

The United States is the world leader in the number of races run and the number of registered foals bred for racing but fourth in overall betting handle behind Japan, France, and Australia. The total betting handle in the United States has declined steadily in recent years, from a high of $15.180 billion in 2003 to $11.410 billion in 2010. Today the host track of a race is just one intermediary for betting on the race, as simulcast and online bets increasingly contribute to the overall amount wagered. Significantly, since 1996, when the Jockey Club began collecting data, the on-track handle has steadily declined to $1.199 billion from $2.944 billion while the off-track handle rose steadily until 2004, to $13.239 billion, before dropping steadily to 1998 levels. Off-track wagering still dominates the share of total wagering at 89.5 percent.[1].

U.S. race tracks use a pari-mutuel wagering system. In pari-mutuel betting the payoffs are set collectively. In essence, bettors are betting against each other instead of against designated odds set by a bookmaker. The takeout is the amount taken by the track operators and government and typically runs 15 to 30 percent of handle depending on the type of wager. At most tracks the takeout is higher for more complex wagers. Studies have examined the effect of takeout rates on trading volume and, as expected, have found an elastic price elasticity of demand (for example, Thalheimer and Ali 1995). Interestingly, Gramm et al. (2007) found that the elasticity of track takeout differs across wager types. This is not surprising since the takeout determines the cost and the risk-reward tradeoff of placing a given type of bet.

Recent evidence from changes in California takeout rates highlights how online betting and simulcasting have heightened this elasticity effect. At the beginning of 2012 the California Horse Racing Board implemented a higher takeout on certain wagers. The response from the racing community was immediate. Players Boycott, an online group, quickly formed and advocated a boycott on California thoroughbred racing. The boycott became official when the Horseplayers Association of North America, whose fifteen hundred members bet $65 million annually, promoted the boycott. By August racetrack executives had proposed reducing the takeout for these bets.

The variety of bets offered and the allocation of total betting volume across different types of bets have changed over the past few decades. Intra-race exotic bets and multi-race exotics at and across racetracks are now offered alongside the standard bets of win (which requires the bettor to select the race winner), place (which typically returns a smaller amount if the chosen horse finishes first or second), and show (which typically pays a still smaller amount if the selected horse finishes in the top three). Among exotic bets, the exacta requires the bettor's two horses to finish in the exact win-place order specified. Trifecta and superfecta wagers extend their selection order to third and fourth positions, respectively. A quinella bet is an attractive alternative for less self-assured bettors because it allows the selected horses to finish in either order. The tradeoff is a smaller payoff if the bettor wins.

Multi-race exotic bets require the bettor to choose the winner in each of a specified number of consecutive races, typically called the daily double, Pick 3, Pick 4, and Pick 6. These multi-race exotics attract even greater wagering pools if there are carryovers. Carryovers occur if no one wins a given wager for that day. This happens more frequently for wagers with a low probability of success, such as a Pick 6 wager. If no one correctly selects all six winners, then a portion is paid to the bettors who selected the most consecutive winners without a loss and the majority of the pool (typically 75%) is carried over to the next race day. This continues until there is a winner. This means that the previous day's track patrons subsidize the next day's track patrons. If the dollar amount of wagers on the exotic bet is constant through time, carryovers increase the expected return on bets when the carryover is large. Naturally patrons find this attractive.

Ramon DeGennaro (2009) found that in standardbred racing a dollar of Pick 6 carryover is associated with almost a $2 increase in total betting. In particular, he found that the ratio of the Pick 6 carryover to the total amount bet averaged 0.4 percent of the amount that gamblers bet on the entire program but ranged as high as 11.1 percent. He also found that the mean of the amount bet on straight wagers was barely a third of the mean handle, even though the number of straight wagering opportunities on a race card is greater than the number of exotic bets available. In thoroughbred racing, Gramm et al. found that large consecutive carryovers increase handle on all wagers and that a large Pick 6 carryover tends to increase total betting volume by 25 to 43 percent.

State governments and racetracks are looking for new types of wagering and gaming venues to attract new gaming dollars as well as ways to keep wagering dollars in state. Purse levels and distribution can affect the quantity and quality of horses racing at a particular track and thereby affect total handle. The purse distribution varies across states, and prior to 1970 most thoroughbred tracks paid only the first four finishers in a race. In 1975 Florida implemented a novel distribution scheme that pays all finishers in the race. Today several states have adopted a similar distribution format and others have it under consideration. This trend is also evident in standardbred racing, as exemplified by Woodbine Entertainment Group's announcement that starting December 1, 2011, they would pay all starters in the race a nominal amount. Using experiments, DeGennaro and Ann Gillette (2013) examined the track choices made by individuals and racehorse syndicate groups under different racetrack purse distributions. Syndicate ownership is becoming increasingly prominent in the industry. DeGennaro and Gillette found that in general both individuals and syndicate groups prefer tracks that pay something to all finishers something unless there is a significant probability that they have a superior horse. These findings have implications for the competition for quality horses across racetracks holding congruent racing meets and, therefore, for a track's share of the overall racing handle.

In addition, direct competition for pari-mutuel horse race betting dollars in the United States has become more intense. For example, pari-mutuel futures markets have in recent years become common for notable races such as for the Kentucky Derby and Breeders' Cup races. Instant Racing, also known as historical racing, is also becoming more common. Instant racing is a pari-mutuel betting system that allows players to wager on previously run races at a video lottery terminal (VLT). There are more than 21,000 digitized videos of historic races. Instant Racing was developed in 2000 by Oaklawn Park in association with AmTote, Inc., and has generated millions of dollars in purses for the track (*Blood-Horse*, Aug. 22, 2011). Several states have recently allowed it, most notably Kentucky. For each historic race played, Instant Racing players are given handicapping information that does not include the date or location of the historic race or the names of the jockey or horses. Players use this information to select three horses they think will finish in the top three positions. As in live racing players wager against each other, and winning payoffs are distributed after the takeout from the pool. Winners receive graduated payoffs for selecting the top three finishers, the first three horses in any order, the top two finishers, the winner, or any two of the top three finishers.

In 2011 New Jersey became the first U.S. state to introduce Exchange Wagering (EW), a form of Internet betting that is prominent in the United Kingdom and Australia. EW differs from pari-mutuel betting in that an individual can make a peer-to-peer bet with just one or more individuals who each agree to the odds before the exchange takes place.

This type of betting also introduces a "betting-to-lose" category that is not available in pari-mutuel betting: a lay bet, as it is called. This is controversial, since a jockey can produce the loss by manipulating the outcome. This integrity issue has raised concerns with the Jockey Guild, thoroughbred owners, breeders, and track associations.

The Components of Demand for Racetrack Wagering

An important question facing the racing industry is how to increase demand for racetrack wagering in today's dynamic competitive landscape. Classification and quality of races differ and affect betting volume, with higher quality races tending to attract a larger dollar amount of wagers. Quality of a race is often measured by purse size or the grade classification of the race. Thoroughbred races are classified into six levels of descending quality and prestige: stakes, allowances, maiden special weight, starter allowance, claiming and maiden claiming.

Stakes races attract the best horses and pay out the highest purses. Allowance races are often run with conditions for entry. For example, they may be restricted to horses that have yet to win a race other than a maiden race or to certain age groups or sex. Any horse in a claiming race is eligible to be claimed (bought). In order to claim a horse owners or trainers must notify, in writing, the racing office no later than ten minutes before the scheduled post time and have the appropriate funds in their racing account. The transfer of ownership occurs at the beginning of the race, but all purse and prize monies are retained by the prior owner (in the event of the death of a claimed horse in a race, the new ownership is fiscally responsible, though many racing organizations provide insurance for such causalities and as of 2013 tracks are reevaluating this liability transfer.). In the event that two or more owner/trainers claim the same horse a draw is held to determine the new ownership. Gramm et al. (2007) found that wagering volume increases with the quality of the race. In particular, stake races attract roughly 15–35 percent greater betting volume relative to their control allowance race group, whereas mid-level quality races attract 22–26 percent less volume and low-level claiming races even less, 37–48 percent less.

Furthermore Chou, DeGennaro, and Sauer (2000) used data from races at Santa Anita Park, a major track in California (USA), in 1987 to demonstrate that the dispersion of the race odds within a race is smaller using claiming races—a price system—than it is using a command system dictated by a highly skilled planner, the racing secretary. This approach is the allowance (or conditioned) approach. Chou, DeGennaro, and Sauer also studied the behavior of horse owners by comparing their self-selection into these two common methods of controlling the ability of horses entered in races. Under the claiming approach, the racing secretary sets the claiming price, letting the owners determine if their horses belong in the race. Horses that enter are explicitly being offered for sale at the predetermined claiming price. Under the allowance approach the racing secretary writes a list of restrictions that exclude some potential entrants which are likely to make for an uncompetitive race because they are either too good or too bad in comparison to the other entrants. The simplest restrictions typically come in the form of age and previous racing success, such as earnings or number of wins. Restrictions may be more complex, though.

A quarter of the races are conditioned races. Chou, DeGennaro, and Sauer speculated on why they remain common despite producing races with a wider dispersion of betting odds, which they interpret as being less competitive. They argue that the reason is probably the wedge between a horse's current racing ability and its total value. The value of horses with the potential to produce quality offspring is often disproportionately large in comparison to their value as racehorses. They give the example of two horses, each worth $300,000. The value of one of the two is distributed equally between its current racing value and its future breeding value. The other horse derives its entire value from its current racing ability. The owner of the first horse will not enter a race with a claiming price of $300, 000 because the horse has little or no chance of being competitive. But the owner cannot enter a race with a claiming price of $150,000, either, because someone will claim the horse at a below-market price. Similarly, claiming races are a poor way to match competitive fields of older horses against very young horses that have yet to reach their full with potential with a large portion of their racing earnings to come in the future. Conditioned or allowance races are better suited to match horses of comparable ability when the efficiency lost to the command or non-price system is smaller than the wedge between a horse's value in excess of his or her current racing ability.

Gramm et al. formally investigated the demand for racetrack wagering assuming that there are two primary types of bettors: an informed bettor who is risk-averse and whose demand for racetrack wagering is dependent on both race-specific returns and information quality and an uninformed bettor who is either risk-neutral or risk-loving with high costs to gathering and processing information. Gramm et al. focuses on whether bettors are deterred by lower returns or by noisier information about a betting interest.

Gramm et al. note that from 1985 to 2002 the total amount wagered on thoroughbred races in the United States almost doubled despite a concurrent 20 percent drop in the number of races run. The authors suggest that the rise in handle is primarily attributable to the ascent of off-track wagering venues. However, they acknowledge that new exotic bets play a significant part. They found that during this time frame, adjusting for inflation, total thoroughbred horse wagering increased 21 percent while per-race wagering more than doubled, increasing by 53 percent.

Using 2002 fall thoroughbred race data from twelve major racetracks, Gramm et al. examined the attributes of a given race that may affect a bettor's preference to wager. They hypothesized that both types of bettors prefer higher returns but that informed bettors reduce wagering when information is noisier, whereas uninformed bettors are unaffected. In their analyses the return on a given wager was related to the track takeout and whether or not it was a carryover wager. The larger the track takeout, the higher the price of wagering (or, alternatively, the lower the expected return), and one expects the wagering volume to be less. A carryover raises the expected return to that specific wager, and one would expect greater interest in wagering.

Gramm et al. hypothesized that risk is increased for a bettor when either there is more noise in a wager or less information available on a betting interest. The variables they used to classify as increasing the noise of information to an informed bettor are: larger field size, lower race quality, competitiveness of the race, and poor track conditions. The effect of race length is uncertain. On the one hand, the longer the race, the more likely that racing luck can be overcome so that quality dominates. For example, a bad start that cannot be overcome in six furlongs might be surmountable over 10 furlongs. On the other hand, a longer race also increases uncertainty because there is more time for random events to occur during the race. Gramm et al. used ordinary least squares regressions with heteroskedasticity robust standard errors to regress the log of race handle on 38 independent variables. Two different regression models using purse size and race classification measures were used to account for the race quality.

They found that for a given race, higher quality races (as measured by purse size or grade classification of race, which is consistent with DeGennaro 1989), more competitive races (as measured by the dispersion of the odds of the race, which justifies the conjecture in Chou, DeGennaro, and Sauer 2000), and turf races increased the volume of betting while higher pari-mutuel takeout, poorer track conditions, and the occurrence of concurrent simulcast races lowered wagering volume. Most of these results are intuitive. A higher takeout lowers the expected return, poorer track conditions reduce the benefit of handicapping skill, and competition for the bettors' wagering dollar is sure to siphon off bets.

Interestingly, these race attributes can vary for different types of wagers in a given race. Specifically, betting volume by quality of race increases in the show pool but has little effect on exacta and trifecta pools. We can think of two reasons why the show pool might increase with race quality more than it does for exotic wagers. First, casual bettors are probably more likely to bet on high-class horses; to them, the sporting aspect of the race is relatively more attractive than gambling. If these casual bettors are more risk-averse, then they are more likely to choose safer wagers. Second, high-class races, drawn from the upper tail of the distribution of quality, are more likely to feature a horse that dominates the betting. Coupled with rules mandating a minimum payoff (usually either 5 or 10% of the amount bet), a dominant horse allows bettors to place large wagers with higher expected returns than other bets. A horse with a 98 percent chance of finishing in the top three actually offers an expected profit, despite the takeout.[2]

Furthermore, Gramm et al. found that each additional betting interest in a race (another entrant) had a nonlinear effect on betting volume. For relatively small numbers of entrants they found that betting volume tended to increase with an additional entrant but that this began to diminish with around 10 or 12 entrants. Thus the optimal field size for maximizing the size of the wagering pool is 10 to 12. This varies across betting wagers, although the estimates are not significantly different: 10 is the optimal field size to maximize the amount bet on trifectas, 11 for win and place bets, and 12 for show and exacta wagers. Gramm et al. suggested that this nonlinear relationship, along with evidence of lower wagering volume for maiden races, which offer bettors less information, support the conjecture that informed bettors reduce betting volume in response to increased noise. They also found evidence that there is less betting volume in races with race restrictions, such as a restriction to just fillies and mares as well as to state-bred horses.

To examine the effects that differential returns to wagering have on the volume wagered, Gramm et al. analyzed track takeout, carryovers, and competition. In their data, the effects of track takeout differed across wagers. Specifically, the demand for wagering was price elastic for straight wagers and exactas but inelastic for trifecta wagers. Carryovers increased betting pools, as expected. In particular, large consecutive carryovers increased handle on all wagers, and a large Pick 6 carryover tended to increase total betting volume by 25 to 43 percent. Betting venues that face competition for race wagers, such as through simulcasting, tend to experience lower betting volume. For each additional race run in the same hour, betting volume declined by 3 to 5 percent for each form of wager. As expected, volume wagered was related to the day of the week, with Saturdays being the most popular followed by Fridays and Wednesdays.

Moreover, they found that track patrons also bet more on races with higher purses, typically stakes races, with a purse elasticity of 0.212. This means that on average for a 1 percent increase in the purse the volume wagered increased by 0.212 percent. Competitiveness of race (as measured by a modified Herfindahl index of the spread of odds) did not affect total handle, though it did affect different wagering types for a given race. To protect the turf courses over extended race day meets, thoroughbreds race on turf less frequently than on dirt or synthetic surfaces. Interestingly, turf races typically are associated with increased wagering volume of 6–11 percent. This may be due to the relative freshness appeal of races run on the surface at a given meet or simply because there is less handicapping noise because in the United States the pool of horses typically racing on turf is smaller.

DeGennaro (1989, 2009) explored the determinants of wagering in the standardbred industry.[3] Unique among racehorses, standardbreds race on either the pacing gait or the trotting gait. Pacers move both legs on one side of their bodies simultaneously. Trotters move their diagonal legs simultaneously. Trotters and pacers almost never race against each other. In DeGennaro's 2009 study, the number of races scheduled for trotters ranged from zero to six. Weather impairs the track 8.5 percent of the time, and 2.8 percent of race cards are contested on state holidays. The number of races per day ranges from 8 to 12, with a mean of almost 11. Races for very young horses appear on 11.5 percent of the race cards; for three-year-olds the figure is 22.7 percent.

DeGennaro estimated the following model:

$$\begin{aligned} H_t = a_0 &+ a_1\, \mathit{TrackCond}_t + a_2\, \mathit{Subsidy}_t + a_3\, \mathit{Temp}_t + a_4\, \mathit{Races}_t + a_5\, \mathit{Trots}_t \\ &+ a_6\, \mathit{OtherFeature}_t + a_7\, \mathit{Tuesday}_t + a_8\, \mathit{Wednesday}_t + a_9\, \mathit{Thursday}_t \\ &+ a_{10}\, \mathit{Friday}_t + a_{11}\, \mathit{Holiday}_t + a_{12}\, \mathit{Pick6}_t + e_t \end{aligned} \tag{15.1}$$

The variable *Trackcond* is a binary variable that equals one if the track surface is impaired for any race on the card and zero otherwise. *Subsidy* is a binary variable that equals one if a sire stakes is contested on the race card (zero otherwise). *Temp* is air temperature. *Races* is the number of live races held that day. *Trots* is the number of live trotting races held that day. *OtherFeature* is one if a feature race other than a sire stakes is contested (zero otherwise). The days of the week are unity on the corresponding days (zero otherwise). *Holiday* equals unity for weekends and holidays (zero otherwise). In this specification $a_{12} = 0$ for the observations for which *Pick6* is unavailable.

Economic intuition and DeGennaro's 1989 results predicted the signs of the coefficients. The coefficient on *Trackcond* should be negative because an impaired racing surface almost always traces to bad weather; bad weather keeps some patrons home, and some of them probably do not wager off-track instead. We defer discussion of *Subsidy* until later, noting only that its coefficient should be positive here. High temperatures are almost never a problem, but cold weather sometimes keeps people at home. Therefore, the sign on the coefficient of *Temp* should be negative. The more races contested, the higher the mutuel handle should be. A feature event probably attracts more betting, just as the Super Bowl or World Cup final attracts more betting than regular season games. Conventional wisdom suggests that the coefficient on *Trots* should be negative. Because potential patrons have more time on holidays and on weekends, the signs on the weekday dummy variables are likely to be negative, while those on Fridays and holidays are likely to be positive. DeGennaro assumed that factors such as competition from other gambling opportunities or sources of entertainment were too small to matter or that they were uncorrelated with the other regressors and were included in the intercept.

DeGennaro estimated equation (15.1) using 436 daily observations, also using the 280 observations for which Pick 6 was available. The results show that most coefficient estimates are intuitively pleasing and much like those in previous research. The exception is *OtherFeature*, which is negative and not close to statistically significant. The economic implications are tiny, as the point estimate implies a decline of less than 1 percent of the average daily mean. The coefficient on the number of races contested on a program is reliably positive, with each extra race associated with an increased betting handle of about 16.6 percent. Sundays and Tuesdays have lower handles, all else being equal (by 13.8%), while Fridays and Saturdays have higher handles (by almost 20%). Breeders' Crown events dwarf these estimates. A Breeders' Crown event is associated with an increase of almost twice the mean total handle. Having this variable undoubtedly contributes to the adjusted R^2 of 0.725 in the model without *Pick6* and 0.637 in the model using the subset of observations for which *Pick6* is available. The time trend is positive in both models and reliably so in the model using all observations, implying that each subsequent race day tends to be associated with an increased handle of about 0.04 percent. The Pick 6 carryover is reliably associated with an increase in handle, and the estimate means that a dollar of carryover tends to increase total betting by $1.90.

DeGennaro (2009) concluded that the determinants of wagering volume have remained fairly constant since his earlier study, despite significant differences in the nature of the data. The two studies used data from racetracks in different countries, operating on different days of the week, and separated by over 20 years, during which regulations, betting technology, and institutional features have introduced new competition for gamblers' dollars. More extensive data allowed him to confirm that the number of races (constant in his previous study) and high-quality races tend to increase betting volume. The Pick 6 carryover is a subsidy from previous patrons to current patrons. Current patrons bet more because the payout is larger, making the expected return better. One noticeable change in the later study is that nonworking days have no reliable effect on mutuel handle. Perhaps modern patrons have more flexible working hours than their predecessors, or perhaps off-track betting makes workdays less important.

State Breeding Subsidies

Proponents of subsidization of the racetrack and breeding industry argue that tangible benefits accrue to the state by means of increased tax revenues through larger wagering handles and agriculture industry job creation (see DeGennaro 1989, 2009). Popular subsidies are breeder incentives which come in several forms, typically for performance and residency. For example, the sales taxes paid for breeding a stallion to a mare in Kentucky are used to pay incentives to breeders whose horses have won races. The amount of incentive varies by type of stakes, with quality receiving more, and with regard to whether the race is in- or out-of-state. In particular, rewarding quality with a strong industry marketing appeal, a Kentucky breeder of the Kentucky Derby and Oaks winners are awarded an incentive of $100,000 each. Typically conditions for eligibility that support the breeding industry are attached. Under this particular subsidy the mare must board in Kentucky from live cover until foaling.

Breeders' Awards programs are also commonplace for other racing breeds. For example, the New York Breeders' Awards program is 10 percent of the state breeding farm account, which comes from remittances as a percentage of total betting handle from the state's harness tracks and off-track betting parlors and a legislatively approved percentage of VLT revenues. The stated purpose of the program is to provide an incentive to promote agriculture through the breeding of standardbred horses in New York State. Awards are allocated on both performance and residency. Performance based awards reward for each level of NYSS (New York Sire Stakes) racing, with higher percentages of payouts going to the higher quality classifications of the race and to the top five finishers. The residency component of the bonus awards rewards breeders for keeping their mares in New York State. Eligibility requires the mare to reside in state for at least 180 consecutive days during breeding season.

Yet an important question remains: how much of a return are taxpayers getting for such subsidies? Using data from standardbred racing, DeGennaro (1989) investigated the influence of sires stakes, races open only to state-bred horses, on the volume of wagering in standardbred races. He found that assistance to the racing industry in the form of subsidized purses for horses bred or foaled within a given jurisdiction has no statistically reliable influence on the volume of wagering in daily data. If there are no benefits other than increased wagering handle, DeGennaro concluded, the subsidy amounts to a direct transfer from taxpayers to the breeding industry.

We have considered DeGennaro's (2009) research in the context of wagering, but that analysis also revisited the standardbred sire stakes subsidy. Despite the dramatic differences between the data in his 1989 and 2009 papers, DeGennaro, using daily data, found similar qualitative results for the effect of the sire stakes subsidy on total betting handle. However, using data from individual races (which were not available for his earlier study), he found that subsidized races *do* show a small but statistically significant increase in wagering, indicating that the government earns at least some return on its investment. Despite the statistical significance, he concluded that the increase is simply too small to detect in daily data, and far too small to justify the government subsidy purely on the economic benefits of increased wagering.

For example, the coefficient on the binary variable *Subsidy*, which equals one if a sire stakes is contested on the race card, never has more than about half the influence of a feature race other than a sire stake, and in the specifications with *Pick6*, the effect of the subsidy virtually vanishes. By contrast, a dollar of Pick 6 carryover is associated with almost a $2 increase in total betting.

The data in the 2009 paper are much better suited to study the subsidy because they include individual race data as well as more than three times the number of daily observations as DeGennaro (1989), and he uses the national unemployment rate and a time trend. These allow for changes in the economic welfare of the track's patrons (which Ali and Thalheimer 1997 found to be important). DeGennaro also includes binary variables to control for Breeders' Crown races. Similar to thoroughbred Breeders' Cup races, standardbred Breeders' Crowns are of immense importance to the sport. DeGennaro's (1989) data do not include any Breeders' Crown races.

DeGennaro's (2009) data include the ages of the horses. This variable was unavailable in 1989. This is particularly important because the state subsidizes races only for two- and three-year-old horses. Younger horses tend to be unproven and more erratic, reducing the advantage of handicapping skill. Gramm et al. (2007) suggested that informed bettors, having less information to process, would bet less. In addition, the owners and trainers of these young horses have a larger information advantage over the betting public. Less informed serious gamblers probably bet less in the face of this adverse-information problem. Because the state only subsidizes races for young horses, this adverse-information problem might explain the sometimes negative sign on *Subsidy* in the daily data. The subsidy might induce patrons to bet more, but if this is entangled with a tendency to bet less on younger horses, it might be impossible to detect in the data. Three-year-old horses are less prone to this than are two-year-olds because most three-year-old horses raced the previous year. DeGennaro (2009) disentangled the effect of the subsidy from the effect of the age group by adding variables to control for unsubsidized races restricted to two- and three-year-old horses.

The results, though, rule out this explanation for why the sire stakes does not increase wagering. First, all of the estimates on control variables for age are positive, though never statistically significant. Gamblers do not bet less, and may even bet *more*, on unsubsidized races that are restricted to younger horses. In contrast, the coefficients on variables which measure deviations from the unsubsidized races that are restricted to the same age groups are negative in four of six cases.

The large increase in exotic wagers does not explain the insignificance of the subsidy, either. DeGennaro repeated his tests using only the straight wagering totals and found no important changes.

DeGennaro's (2009) results, therefore, mirror his earlier results. After controlling for several other factors that explain wagering volume, the sire stakes subsidy simply is not positively correlated with total betting volume on the day these races are held. The evidence supports previous research showing that this particular government subsidy does not generate additional revenue—at least in the short term—to offset the cost of the subsidy. The loss to the state from the subsidy is not as large as it first appears, though, because the industry itself funds part of the sire stakes' purse through stallion assessments, contributions from breeding farms, and from entry fees paid by horse owners. In addition, a portion of revenues from slot machines located at the racetrack is directed to the sire stakes program at the racetrack that supplied the data.

In addition, DeGennaro's results using individual races are the first to show that the state may in fact recoup part of its investment because the sire stakes subsidy is associated with increased betting on subsidized races. However, the estimated magnitude of the increase is far too small to compensate for the subsidy, and taxpayers can still ask why states continue to subsidize this aspect of the racing industry. DeGennaro offers some possible explanations. One traces to public choice theory. Governments subsidize many things, not just horse racing. Perhaps the better question is why governments subsidize racing rather than some other activity. To justify the subsidy on economic grounds, one must find more (taxable) economic activity either in other areas or in other time periods that trace to the subsidy. For example, because the horses compete for higher purses, their earning potential is higher, and they sell for higher prices. This means that the state could recoup part of the subsidy through higher taxes on horses sold. Another possibility is that the state overtaxes the horse racing industry, and the subsidy is a way to reduce the effective tax rate. It would be simpler and more efficient economically to reduce the tax that is too high, but this might be politically less viable. "Subsidizing jobs" makes a better campaign sound bite than reducing "sin taxes."

Furthermore DeGennaro conjectures that conceivably the racing industry offers a politically attractive way to legalize casino games, including table games and VLTs. Taking the step from no gambling to casino games might be a bridge too far, but adding gaming to existing gambling opportunities might not be. Preserving horse racing as an avenue to institute gaming in the future might make sense. One might also argue that *future* tax revenues from gambling are higher because of the current subsidy. Races with better horses tend to generate more betting than those featuring lesser animals, and sire stakes horses are better than most others. However, we know of no research that supports the link between racehorse quality (or subsidies) and higher betting *through time*. This represents a fertile ground for additional research.

Promoters of the subsidy could argue as well that the racing industry provides a public good. For example, perhaps having a racetrack or breeding farm nearby attracts businesses and tourists, which locate near or visit an area with a racetrack instead of a competing location. The Kentucky Cabinet for Economic Development produced an advertisement that reflects this (see *US Airways* magazine, June 2007, p. 129). Breeding farms and racetracks offer open areas, green areas, and access to animals, which many people enjoy, and many tracks feature areas for children to play. The atmosphere is similar to a baseball stadium or even a park or zoo, and for better or worse, states do subsidize these activities. Viewed this way, racetracks provide non-pecuniary benefits to the area. Perhaps, then, the state's subsidy to the racing industry permits it to economize on other parts of its budget.

Efficiency of the Race Track Wagering Market

The efficiency of wagering markets is of great interest to researchers, bettors, and the industry. Many industry players believe that betting on horse racing is similar to investing in stocks. That is, there is a return for gathering and processing information, and thus, horse betting is not just pure speculation or gambling. Researchers have hypothesized that these markets have an opportunity to be more efficient than even the stock market because the wager value becomes known at a fixed termination point, there is quick feedback, low cost access through off-track venues, and learning is enhanced with repeated opportunities to bet the same wager or betting interest Hausch, Ziemba, Rubinstein (1981). Offsetting this, however, is the inability of bettors to take short positions. As in the financial markets, there is strong interest in how informed bettors, dubbed the "smart money," wager and how they place bets.

In the United States, the final odds at a racetrack that determine potential winnings on a race are not known until the betting windows close at post time. Thus payoffs are unknown to the bettor until after the race starts. Camerer (1998) found that approximately half the money in his sample was bet three minutes before post time. It is widely accepted that informed bettors bet near the end of the wagering period. Supporting this, Asch, Malkiel and Quandt (1982) found that bets made near the end of the wagering period are better predictors of the order of finish than are the final odds. Asch and

Quandt (1987) examined exotic bets and reported that the smart money systematically bets on exactas in order to avoid signaling their actions.

Manipulation of markets with large trades is of particular interest in many markets. Camerer (1998) ran an interesting field experiment using pari-mutuel betting on thoroughbred racing to examine how information aggregation occurred in these centralized pari-mutuel markets and to test whether these markets were easily manipulated by large bets. By computer he strategically placed large bets that visibly moved the odds on horses (relative to matched-pair control horses in the same race with similar pre-bet odds) and then canceled these bets before the race was run. The temporary bets were placed approximately 17–22 minutes before post time and cancelled 5 to 8 minutes prior to post time. Even though there was a slight movement of money toward the horse bet it was temporary such that the net effect was statistically insignificant. Camerer found no evidence that bettors systematically responded to his manipulated change in odds relative to his control sample horse at any time when the temporary bet was "live." He concluded that information aggregates well in these pari-mutuel markets and bettors know enough to ignore a large bet before post time that is not backed by consistently increasing wagers.

Other studies have looked for market inefficiencies for the possibility of exploitable wagering patterns. The most notable empirical regularity in racetrack betting markets documented in both bookmaker and pari-mutuel data is the favorite-long shot bias first documented by Griffith (1949). The favorite-long shot bias suggests that betting odds are biased estimates of the probability of a horse winning. In particular, longshots are overbet while favorites are underbet, giving consistently greater returns to favorites as compared to long shots. Thaler and Ziemba (1988), Sauer (1998), and Snowberg and Wolfers (2007) provide literature surveys.

Focusing on the post-simulcast era, Gramm and Owens (2006) examined whether the favorite-longshot bias persists when the final odds pool is comingled with bets from a variety of sources: on-track betting, casinos, off-track betting hubs, phones, and online betting venues. They found that the favorite-longshot bias still exists and is more severe in the place and show pools than in the win pool. Interestingly, they went on to analyze whether one could profitably arbitrage this inefficiency and found it to be unprofitable. Since roughly only 60 percent of the final pool totals are recorded when the betting windows close, Gramm and Owens found that late money reduced or eliminated the inefficiencies that appeared exploitable.

Examining the bias from a different significance perspective and noting that bets on longshots account for a small amount of the actual wagers in aggregate, O'Conner (2007) used the common method of value weighting in financial portfolios to investigate the economic significance to bettors in the aggregate of exploiting the longshot bias. He weighted each horse in the portfolio according to the dollar amount bet on it instead of assigning an equal wagered amount to all horses. O'Conner found that the longshot bias is only one-third as strong in a value-weighted portfolio relative to an equal-weighted portfolio and that it falls considerably as pools increase in size. It would be interesting to see if future research using different data sets across countries and time replicated this result.

New Landscape Issues

Facing a decline in the volume of pari-mutuel wagering in recent years, horse racing advocates have considered several changes to enhance the sport's visibility. One recurring discussion centers on the tendency for the most successful racehorses, particularly stallions, to retire from racing and begin lucrative careers in the breeding shed. The conventional wisdom is that people, particularly new and potential fans, need time to build an affinity for the sporting aspect of racing rather than just the appeal of betting. Keeping top horses on the racetrack longer would probably offer media outlets additional story lines, too, as last year's competitors begin training for the upcoming racing season.

Thoroughbred racing champion Zenyatta is the perfect example of how one horse's four-season racing career can captivate not only the nation's but the world's attention. Her unprecedented consecutive wins, signature paddock dancing, and electrifying come-from-behind finishes propelled her into the spotlight, but it was also the generous access and astute decisions of her owners and racing team that allowed her legacy and popularity to benefit the industry. After becoming the first filly to win the Breeders' Cup Classic she was brought back from a brief retirement as a six year old to challenge more industry records. Although it is hard to measure the impact that a superstar horse has on overall wagering handle, the *Daily Racing Form* (Nov. 7, 2011) noted that the total handle for the Breeders' Cup races was down 5.1 percent relative to 2010's record level increase of 15.5 percent when Zenyatta raced for the second time in the classic.

What is more important for the industry is how she won the hearts of thousands of non-regular race fans and brought back visibility to thoroughbred racing. Her iconic popularity is so strong that even the Los Angeles Dodgers baseball team has used her on billboards for promotion, and she was featured on *60 Minutes* among other national media outlets. Thousands of fans came to her retirement ceremonies at Hollywood Park and Keeneland Race Track. Her popularity has not waned since retirement, as she has more than 90,000 fans on Facebook who follow almost daily blogs by racing manager Dottie Shirreffs, and she continues to raise a great deal of money for several racing and cancer research charities. Team Zenyatta exemplifies what many people feel the sport of racing should be, and they were recognized with a special Eclipse Award for extraordinary service, individual achievement, and contributions to thoroughbred racing. As economists we realize that not all owners are so philanthropic and that the industry needs to create credible incentives to capture the externality of longer racing careers. Perhaps one credible way would be to fund valuable multiyear racing career awards.

Another aspect of concern is racing very young horses, whose developing tissues and bones are at greater risk for injury. This is a double blow to the sport because not only does it add risk to owners, but it also is a source of bad publicity—few people want to see animals injured, and such injuries reflect badly on the sport. Particularly for standardbred racing, the present structure of races, with the most lucrative contests restricted to two- and three-year-old horses, provides strong economic incentives to retire the top stallions to breeding farms, where they have a longer period of time to sire potentially valuable offspring. In economic terms racing horses beyond three years of age is a positive externality to the industry, whereas the private benefits are small to negative for the individual owner.

To address this issue in standardbred racing in the United States, Jeff Gural, chairman of American Racing and Entertainment (which owns and operates Tioga Downs and Vernon Downs), has implemented new rules for qualifying to race at his most recent acquisition, the Meadowlands. Specifically, he will ban the offspring of stallions that were four years old at the time of conception from competing in stakes races. Gural would offer exceptions for stallions that are sick, injured, or otherwise unable to race.

We understand and sympathize with the goals of such proposals. As economists, though, we know that when prices are subverted in favor of non-price approaches such as bans, unintended consequences result. Market participants will surely attempt to game the system to circumvent such a ban. We can think of several possible ways, and creative minds in the industry are sure to find several more. For example, an owner could pay a veterinarian to diagnose a career-ending illness or injury or, even worse, create one. Owners of promising young stallions need not breed their horses in states where such bans exist or even in the United States. Instead, they could ship their horses to other parts of the country where the ban is not in effect or even abroad to begin their stallion careers. This would have the perverse effect of reducing media coverage of the sport at the tracks where the ban is in place. We can imagine that sponsors of stakes races currently at these three tracks will encounter pressure to move to other racetracks not subject to the ban. If enough quality owners move to another track then sponsors will follow. The real winners of such a ban are the current owners of breeding stock because the ban delays competition from young stallions that current owners of bloodstock would otherwise face.

As we noted above, statistics from the *Daily Racing Form* for 2010 found that off-track simulcast betting now drives wagering handle. This has subtle implications for the future of racing. Given that even small tracks strategically positioned in the racing schedule can draw respectable betting volumes, a key driver of a racetrack's success is no longer just its location in geographical space but also its location in time. Technology makes geography much less important for wagering handle than in the past. This means that racing's efforts to bring live racing to new areas need to be conscious of strategically placing the race meet in the wagering market schedule. As a case in point, during recent legislative hearings in the state of Georgia (USA) horse racing proponents emphasized not only the benefits of increased employment to the state but also the strategic geographical advantage that allows them to place short meets within the racing schedule as horses move back and forth from Florida to other parts of the country during the different seasons. Not surprisingly, proponents are also seeking a state constitutional amendment to allow pari-mutuel betting in the state in order to take advantage of the simulcasting wagering market revenues while simultaneously arguing there is no need for other forms of gambling.

Currently 44 of the 109 tracks devoted to thoroughbred racing in the United States are subsidized by gaming. Many states are also turning to casino gambling to help revive the racing industry. For example, New York is one of 12 recently allowing casinos at racetracks, commonly called racinos. In particular the new casino at Aqueduct Racetrack, Resorts World Casino, brought in an additional $14.8 million to race purses and $15.96 million to the New York Racing Association in the first six months of operations, 6.5% and 7% respectively of revenues (New York Daily News, Sports, March 25, 2012). Higher purse levels attract better quality horses which in turn tend to lead to higher wagering handles. But even before Aqueduct's racino's fruits can be fully realized, the dynamics of competition for the gaming dollar continues to change. The Shinnecock Indian tribe has announced that it is contemplating putting a racino at Belmont Park race course, just seven miles away from Aqueduct. Many players in the gaming industry believe this concentrates too many venues within too small a geographical space.

Summary

Horse racing is an important global industry. Along with other forms of gambling this industry has provided financial support for beleaguered state governments. Technology has changed the way track patrons bet, with the vast and increasing majority of total bets being made away from live racing. The menu of wagering options continues to grow with a wide variety of exotic bets, futures wagering markets, program betting, and on line betting venues becoming more prevalent. We know a fair amount about the determinants of wagering volume, but our knowledge of the demand for wagering on these different types of bets within and across venues remains in its infancy. In particular, different types of informed traders will take advantage of these new venues to hedge their bets as the new betting exchanges offer different bet types, for example, limit price betting. As in financial markets, these market microstructure issues will induce changes in the wagering market, constituting fertile ground for future research.

Past studies in standardbred racing found that subsidies to purses, in the form of sire stakes, to the standardbred racing industry, do not increase wagering nearly enough to justify their use solely on the growth in handle. Policy makers must make the case that they help taxpayers in other areas. For example, they might argue that higher tax revenues will be generated by increased wagering in the future or by attracting gamblers to other forms of taxable gaming, such as VLTs, or that racetracks provide a public good. Perhaps, too, the subsidy provides a politically viable way of offsetting tax rates on gambling that are too high. Future research is needed to value these positive externalities and to measure whether these same conclusions apply equally to the thoroughbred and quarter horse racing industries, and in more recent times with simulcasting.

Additional research is needed to assess the efficiency of wagering in racing markets in the United States in light of recent types of wagering options coming into the marketplace. With exchange betting venues having entered the United States for the first time in 2011, and with the relatively recent onset of program betting, it will be interesting to reexamine the survivability of the favorite-longshot bias. In addition, increasing numbers of informed bettors with more venues and bet options are likely to alter the information content and manipulative ability of the on-track pari-mutuel market odds in interesting ways.

Finally, we described several new issues that will continue to affect the racing landscape and avenues to reignite interest in the racing industry, such as longer racing careers for top horses. One important dynamic that could have a profound effect on the racing and wagering industry would be the creation of a national racing association. Many different stakeholders in the industry have promoted such an idea as a way to control the image and integrity of the industry, with particular regard to drug usage standards, and to organize the racing schedule to maximize the total wagering handle. With states heavily vested in the subsidization, pricing/taxing of relative wagering venues and ownership of racetracks, this would involve complex political negotiations and implications. An interesting area of future research might examine how this transition might, come about, whether or not this structure would mimic other U.S. sports leagues and its governance pitfalls.

Notes

1. Data in this paragraph are from the Jockey Club Online Fact Book (www.jockeyclub.com/factbook.asp).
2. Track patrons call such bettors "bridge-jumpers," a colorful term referring to the strong regret that losers of such bets feel.
3. This section draws heavily on DeGennaro (2009).

References

Ali, Mukhtar M., and Richard Thalheimer. 1997. Transportation Costs and Product Demand: Wagering on parimutuel horse racing. *Applied Economics* 29(4):529–542.

Asch, Peter, Burton G. Malkiel, and Richard E. Quandt. 1982. Racetrack betting and informed behavior. *Journal of Financial Economics* 10(2):187–92.

Asch, Peter, and Richard E. Quandt. 1987. Efficiency and profitability in exotic bets. *Economica* 54(215):289–298.

Bossert, Jerry. 2012. Increased purses from Aqueduct casino seemed like winning bet until horses started breaking down. New York Daily News, Sports, March 25.

Camerer, Colin F. 1998. Can asset markets be manipulated? A field experiment with racetrack betting. *Journal of Political Economy* 106(3):457–482.

Chou, Nan-Ting, Ramon P. DeGennaro, and Raymond D. Sauer. 2000. The efficiency of the price system: Evidence from an alternative market. *Applied Economics Letters* 7:703–706.

DeGennaro, Ramon P. 1989. The determinants of wagering behavior. *Managerial and Decision Economics* 10:221–228.

———. 2009. New evidence on the link between government subsidies and wagering. *Journal of Gambling Business and Economics* 3(2):37–62.

DeGennaro, Ramon P., and Ann B. Gillette. 2013. Syndicates' choices under different payoff distributions. Kennesaw State University Working paper.

Deloitte Consulting LLP. 2005. The economic impact of the horse industry in the United States, June.

Gramm, Marshall, and Douglas H. Owens. 2006. Efficiency in pari-mutuel betting markets across wagering pools in the simulcast era. *Southern Economic Journal* 72(4):926–937.

Gramm, Marshall, C. Nicholas McKinney, Douglas H. Owens, and Matt E. Ryan. 2007. What do bettors want? Determinants of pari-mutuel betting preference. *American Journal of Economics and Sociology* 66(3):465–491.

Griffith, R. M. 1949. Odds adjustments by American horse race bettors. *American Journal of Psychology* 62(2):290–294.

Hausch, Donald B., William T. Ziemba, and Mark Rubinstein. 1981. Efficiency of the market for racetrack betting. *Management Science* 27(12):1435–1452.

O'Conner, Philip F. 2007. The economic significance of the longshot bias in horse race wagering. Working paper, June. Hamilton, New Zealand: University of Waikato Management School.

Paulick Report. 2011. *DRF* study: Slots tracks drive far less handle, Aug. 18; http://www.paulickreport.com/news/the-biz/drf-study-slots-tracks-drive-far-less-handle.

Sauer, Raymond D. 1998. The economics of wagering markets. *Journal of Economic Literature* 36(4):2021–2064.

Snowberg, Erik, and Justin Wolfers. 2010. Explaining the Favorite-Longshot Bias: Is it Risk-Love or Misperceptions? Natioanl Bureau of Economic Research working paper 15923.

Thaler, Richard H., and William T. Ziemba. 1998. Parimutuel betting markets: Racetracks and lotteries. *Journal of Economic Perspectives* 2(2):161–174.

Thalheimer, Richard and Mukhtar M. Ali. 1995. The Demand for Parimutuel Horse Race Wagering and Attendance. Management Science 41(1): 129–143.

Walker, Douglas M., and John D. Jackson. 2011. The effect of legalized gambling on state government revenue. *Contemporary Economic Policy* 29(1):101–114.

CHAPTER 16

WHAT EXPLAINS THE EXISTENCE OF AN EXCHANGE OVERROUND?

DAVID MARGINSON

1 Introduction

The advent of person-to-person Internet betting constitutes a radical change in the nature of gambling. Prior to such "exchanges" as Betfair (www.betfair.com), bettors faced what may be described as wagering asymmetry; it was possible to bet on something *to* happen but not to bet on something *not to* happen. Before the advent of the exchanges, only licensed bookmakers could lay something or someone to lose.[1] Since 2000, however, and because of the exchanges, unlicensed individuals can now either or both "back" and/or "lay." The exchanges enable or establish symmetry in wagering. This symmetry is particularly evident for horse race betting. Punters can now either back one or more horses in a race to win and/or lay one or more horses in the same race to lose. Horse race betting is the market focus for this chapter.

The introduction of Internet-based person-to-person betting also represents an important structural development for horse race and other betting markets. This structural development concerns the establishment of an order-driven type market to complement traditional bookmaker-led, quote-driven markets. Broadly, a quote-driven market is one in which dealers or market makers offer to buy or sell assets/securities to investors (Theissen 2000). All transactions are done with or through the market maker (normally at the bid and ask prices quoted by the dealer), who provides immediacy to the market.[2] Bookmaker-led markets are essentially quote-driven markets. The bookmaker is the market maker, providing immediacy for betting. In turn, bettors (investors) must bet at the prices offered by the bookmaker who, in reflecting quote-driven arrangements, takes the opposite side (as "layer") of every (betting) transaction.

An order-driven market operates *without* specialist market maker intermediation (Handa, Schwartz, and Tiwari 2003). Instead investors trade directly among themselves. This direct trading occurs in one of two ways. Any seller (buyer) can place offers (bids) as limit orders and wait until the order is executed; or, alternatively, seller or buyer can trade immediately by placing a market order against an existing limit order (offer or bid).[3] Buy and sell prices are publicly displayed, with the transaction price being the result of the equilibrium of supply and demand (Handa, Schwartz, and Tiwari 2003).

Internet person-to-person betting exchanges (henceforth exchanges) may be considered the betting equivalent of an order-driven market. First, the exchanges are similarly nonintermediated markets; there is no market maker. Put colloquially, the exchanges represent "betting without bookies" (O'Connor 2007). Second, trading occurs directly between bettors in essentially the manner described above for order-driven markets. The terminology is that bets are "matched."[4]

An important question that applies equally to both quote-driven and order-driven markets is that of market efficiency; i.e., how efficient is the market? In seeking to address this question, one line of research, often referred to as market microstructure, has focused on exploring the idea that asset prices need not equal full information expectations because of a variety of market frictions (Madhavan 2000). Market frictions (inefficiencies) include the cost of transacting or trading in a particular market (Stigler 1967). The most significant and observable transaction cost is the bid-ask spread (Kumar 2004). The bid-ask spread represents the difference between asset buying and selling prices. Its existence is generally explained in terms of the effects of inventory holding costs, order processing costs, liquidity, and adverse selection costs (Levin and Wright 2004). While differing in magnitude, bid-ask spreads are characteristic of both quote-driven and order-driven markets (Theissen 2000).

For horse race betting markets, the bookmaker's overround or "vig" may be viewed as akin to the bid-ask spread in financial markets (Shin 1991). The explanation is as follows. Consider the market for bets in a horse race comprising n runners. This market "corresponds to a market for contingent claims with n states, in which the ith state corresponds to the outcome if the ith horse wins the race" (Shin 1993, 1142). In this market, the value of the securities (say £1 if a particular state occurs; 0 otherwise) is determined by betting odds (implied win probabilities). The sum of implied win probabilities on all race entrants gives the price of a portfolio which pays £1 for sure at the end of the race (Shin 1993). For operational efficiency, sum of implied win probabilities = 1 ($x = 1$). Thus if x diverges from 1 ($x - 1$), an overround (OR) or market spread exists (Coleman 2007; Law and Peel 2002). Overrounds (ORs) based on publicly reported starting prices (sum of starting prices – 1) provide a clear, unambiguous, and readily accessible measure of the size of the (final) market spread (Shin 1993, 1142).[5]

Since December 12, 2007, Betfair, the leading Internet-based person-to-person betting exchange (O'Connor 2011), has reported its own horse race starting prices to rival industry- or bookmaker-determined starting prices. Betfair starting prices are seen as more "efficient" than industry starting prices, and evidence does suggest that longshot prices in particular are considerably higher than the equivalent bookmaker starting price (SP) (O'Connor 2007; Marginson 2010). At the same time, it follows that the reporting of Betfair starting prices (SPs) helps to establish an equivalent "sum of Betfair (starting) prices." To the extent that this sum of SPs diverges from 1, an equivalent Betfair overround is created. A cursory review of the data confirms the existence of Betfair ORs, both < and >1. The existence of "underrounds" (sum of SPs <1) is interesting, not least as it implies that, in theory at least, every participant in a horse race could be backed to yield a profit, however small. More importantly from an operational efficiency perspective, divergence of sum of Betfair SPs from 1 gives rise to the following research question: why does this bid-ask spread exist? Put slightly differently, we may ask: what factors might help to explain the existence of an over/underround in the Betfair exchange market?

The purpose of this chapter is to examine the reported overrounds on Betfair in an attempt to explain their existence. The study draws on both the finance and horse racing literatures as a basis for identifying possible determinants for investigation. The factors examined include, from the finance literature on order-driven bid-ask spreads, balance of investor opinion/activity, trading volume, and adverse selection costs (Handa, Schwartz, and Tiwari 2003). Drawing on the horse racing literature, the following are examined: number of runners, grade of race, age limit of the race (three proxies for adverse selection), last race effect, type of race (handicap versus nonhandicap), type of race event (Flat vis-à-vis National Hunt), and, given the increasing association between exchange and racetrack betting (O'Connor 2011), bookmaker overrounds. Hypotheses relating to each are developed and subject to empirical analysis.

The empirical analysis is based on data collected from 2,184 horse races tha took place in the United Kingdom between autumn 2008 and spring 2010. The empirical analysis suggests that Betfair overrounds vary in relation to: balance of activity (whether weighted toward back or lay), trading volume (monies wagered), grade of race, and bookmaker overrounds. Significant effects (at $p < 0.05$) are also observed for type of race (handicap versus nonhandicap), "last race," and number of runners. Nonsignificant results are reported for type of race event and age limit of the race. Findings support the finance literature concerning, for instance, the effects of balance of activity on bid-ask spreads in order-driven markets (Handa, Schwartz, and Tiwari 2003). In contrast, findings for grade of race and type of race challenge arguments in the horse racing literature about the effects these two factors are expected to have on betting overrounds (Vaughan Williams and Paton 1997). Possible reasons for these finding are considered. The analysis suggests the need for further research into the operational efficiency of exchange betting markets. More broadly, the present research suggests that microstructure analysis of order-driven betting markets such as Betfair offers a potentially fruitful line of inquiry for those interested in understanding market efficiency.

The next section describes exchange betting and explains how an overround contributes to market inefficiency. Section 3 develops the study's hypotheses. Section 4 outlines the research methodology. Section 5 presents the empirical results. Section 6 concludes.

2 Exchange Betting

"Betting exchanges exist to match people who want to bet on a future outcome at a given price with those who are willing to offer that price. The person who bets on the event happening at a given price is the "backer" [or "bettor"]; the person who offers the price is known as the 'layer"' (Smith, Paton, and Vaughan Williams 2006, 674).

For example, someone may wish to lay a particular horse at, say, odds of 3 to 1 to a maximum stake of, say, £100 (maximum liability is £300). Someone else, the bettor, may accept these odds and place a bet with the layer, say £100 at 3 to 1. A wager thus occurs between the two parties. Equally, exchange bettors can indicate (1) the odds at which they would be willing to *bet* on a particular horse (say 7–2) and (2) the magnitude of the wager at that odds level (say £100). Anyone wishing to *lay* the same horse at those odds can accept this bet and thereby form a transaction with the bettor. A wager thus occurs once more—between layer and bettor.

From an order-driven market perspective, the above two betting examples describe both limit orders and market orders. By specifying both the order/volume (e.g. £100 wager, £300 liability) and the price (e.g.7–2), those bettors who place wagers (lay or bet) are placing limit orders; they must wait until the wager gets executed. The wager is executed by those who accept the lay or bet. Accepting is akin to placing a market order; these bettors are trading immediately with an existing limit order (lay or bet).

The multiplicity of limit orders on Betfair are displayed under the headings "back" and lay," with each category comprising three columns of prices. The term *back* represents a situation in which a bettor has placed a limit order about something *not* happening. In our example, for instance, the wager (limit order) of the person offering 3–1 against a particular horse winning (stake £100) will be displayed under "back." In this context, back represents the response (bet) of the person (bettor) accepting the limit order (thereby making a market order). The term *lay* represents the actions of the participant accepting the limit order placed by the bettor seeking to back the horse to win at 7–2 (stake £100).

Both "back" and "lay" help to demonstrate the symmetry of betting provided by the exchanges. At the same time, the development allowing non-licensed individuals to bet on something *not* to happen has proved controversial. The main and continuing concern is that, by "allow[ing] punters to bet on horses that lose" (O'Connor 2007), betting exchanges create "the perfect recipe for malpractice, such as race fixing" (statement by chairman of the Australian Jockeys Association, Oct. 14, 2004). There is evidence to support these assertions (Marginson 2010). Such concerns, notwithstanding, exchange betting has proved extremely popular. Betfair, for instance, reported, for the year 2010, yet another increase in revenue to £340.9 million. While this increase, at 13 percent, is less than 2009's 24 percent increase, and while gross profit before tax, at £17.8 million, represents a fall of 63 percent from 2009's gross profit of £47.5 million, Betfair remains by far the biggest Internet-based betting exchange globally (O'Connor 2011).

The continuing success of Betfair appears to have, in part at least, informed the company's desire to get involved with SPs. In October 2007 it was reported that representatives of Betfair had made a submission to the Starting Price Regulatory Commission, suggesting that Betfair prices should form part of the overall calculation of the SP. This offer was rejected by the commission. In response, Jack Houghton, a spokesperson for Betfair, is quoted as saying: "Betfair offered to provide data from its audit trail to assist in returning an SP that was more equitable, transparent and fair than the incumbent system . . . Betfair's concern is that the new approach, regrettably, does not prioritise the interests of punters" (O'Connor 2011).[6]

The launch of Betfair's own SP service in December 2007 provided bettors with a clear way of comparing industry and exchange SPs. As with bookmaker-based SPs, the production of Betfair SPs is governed by certain rules and procedures. These are presented on Betfair's website (www.betfair.com). The algorithms involved can be detailed and complex, not least because, and controversially for some (see O'Connor 2011), they allow the incorporation of some unmatched limit orders into the calculation of Betfair SPs. The reader is referred to Betfair for a full account of how the SPs are calculated. Perhaps the two key points to mention here are that (1) to a large extent, Betfair SPs reflect the average of the best back and lay odds immediately prior to "the off" and (2) based on publication of Betfair SPs, it is possible to determine Betfair ORs, calculated as sum of SPs – 1.

From an efficiency perspective, transactional or operational *in*efficiency exists if sum of prices (SPs) diverges from 1. Normally, for bookmaker-led markets at least, sum of prices exceeds 1 and may do so by 30 percent or more (Cain, Law, and Peel 2003). The effect of ORs >1 is to compress prices (implied win probabilities) below their "true" or objective level, as represented by the frequency of winning outcomes.[7] For example, suppose each horse in a four-runner race is priced at 2/1. This implies that each should win 33.33 percent of races or, say, 4 races out of 12. But of course, there are *four* runners, and so, assuming equal ability (as indicted by the equal prices), the "true" outcome would be three wins each (out of 12 races), not four. Objectively, this frequency is represented by odds of 3/1 (25%) and a zero OR. Thus for efficiency, market prices should approximate to their "true" or objective win probabilities. Outcomes priced at 1/1 should occur 50 percent of the time. Outcomes priced at 100/1 should happen, on average, 1 in 101 cases. In our example, however, each horse wins 25 percent of races, a lower ratio than the 33.33 percent implied in the odds of 2/1. Not all prices in a given betting market may be affected in this manner. Nonetheless, so long as an OR exists, inefficiency exists; at least one price does not match its "true" value.[8] (Sum of prices <1 would create a situation where one or more implied odds are potentially greater than their objective or true odds, as represented by race outcomes.)

For the sample of 2,184 races used in the present study,[9] Betfair sum of SPs (sum of SPs – 1) ranges from 86 to 112 percent. Thus ORs based on SPs reported by Betfair range from −14 to +12 percent. This variation is nontrivial; it highlights the extent of potential operational inefficiencies for an order-driven betting market such as Betfair. This leads us to ask: what might explain the existence of such SP-based ORs on Betfair? Drawing on both finance and horse racing literatures, the next section develops testable hypotheses that seek to answer this question.

3 Explaining Betfair Overrounds

3.1 Explanations from Finance

The finance literature posits that three factors in particular influence the magnitude of bid-ask spreads in order-driven markets (see, e.g., Chakravarty and Holden 1995; Chung, Van Ness, and Van Ness 1999; Handa and Schwartz 1996; Handa, Schwartz, and Tiwari 1998, 2003; Goettler, Parlour, and Rajan, 2009). The factors are (1) "balance of opinion" based on the proportion of buyers (buying activity) vis-à-vis sellers (selling activity) in the market, (2) trading volume, and (3) adverse selection; the risk of trading with informed traders (Handa, Schwartz, and Tiwari 2003). The third factor is also a feature of the horse race betting market literature (Crafts 1985; Vaughan Williams 2005) and will be discussed under "Explanations from Finance and Horse Racing." The following discusses "balance of opinion" and trading volume.

3.1.1 *Balance of Opinion in an Order-Driven Market*

Puneet Handa, Robert Schwartz, and Ashish Tiwari (1998) argued that bid-ask spreads in order-driven markets are related to the activities of buyers and sellers, specifically differences of opinion relating to asset values. Fundamentally, markets exist because investors hold heterogeneous beliefs and opinions (Madhavan 2000). For various reasons, traders reach different opinions based on the same public information. In the context of order-driven financial markets, a "balance of opinion" (equal weighting of buyers/sellers) is associated with "equilibrium" or widest bid-ask spreads. In contrast, an imbalance in favor of either buying or selling activity is argued to drive bid-ask spreads away from equilibrium and toward minimum values (Handa, Schwartz, and Tiwari 1998).

For instance, suppose two groups of investors, both with access to the same public information, place different values, V_{high}, V_{low}, on a particular security/asset. V_{high}

investors seek to purchase shares; V_{low} investors seek to sell shares. In this context difference of opinion has two dimensions: (1) divergence of opinion, $V_{high} - V_{low}$, and (2) the proportion, p, of investors with valuation V_{high}. Handa, Schwartz, and Tiwari (1998) argued that, for a given $V_{high} - V_{low}$, the difference of opinion reduces or disappears at extreme values of p (0 and 1) and is maximized at $p = 0.5$.

Handa, Schwartz, and Tiwari (1998) further argued that differences of opinion translate into or help to establish bid and ask prices, and thereby the magnitude of the bid-ask spread. Their reasoning is predicated on the notion of a "gravitational pull" effect that a previously posted buy (or sell) order exerts on an incoming sell (or buy) order. The pull is the attractiveness of trading with certainty by market order (investors placing limit orders face the risk that their order may not be executed, in Betfair terms, not matched). According to Handa, Schwartz, and Tiwari (1997, 49), a spread thereby results because "any limit order sitting on the book must be far enough away from a counterpart quote to lie outside the gravitational pull of the counterpart quote."

Consistent with the gravitational pull effect, market bid, B, and market ask, A, both depend on the nonexecution risks faced by buyers relative to sellers. Further, nonexecution risks for each participant depend on the relative proportion of buyers, p (valuation, V_{high}). For p close to unity, nonexecution risk is high for buyers and low for sellers (and vice versa for p close to zero). For this reason, the equilibrium spread, A – B, also depends on p. Handa, Schwartz, and Tiwari (1997, 2003) have shown that spreads in an order-driven market are widest at $p = 0.5$ and lowest when p is at either of its extreme values (0 or 1).[10]

It is possible to apply the above line of reasoning to betting exchanges. The arguments are threefold. First, as with financial markets, exchange betting markets exist because bettors, based on the same public information, hold different opinions about the winning chances of a given race entrant. Those bettors (buyers) holding a relatively high opinion of a horse's chances will back the horse to win. Bettors (seller) holding a relatively low opinion will lay the horse to lose.

Second, similar to order-driven financial markets, it also seems reasonable to suggest that a gravitational pull effect may influence bettor behavior on the exchanges. The point is that exchange bettors face similar uncertainties that their limit orders (back or lay) may not be "matched." As such, there is an attractiveness to trading with certainty by matching existing limit orders rather than by placing additional limit orders. And third, the gravitational pull effect is likely to increase as the proportion of backers to layers diverges from equilibrium (same proportion of each) toward either extreme (more backers to layers and vice versa). As the extreme positions are approached, betting activity is "pulled" toward either the back or the lay prices, depending on the nature of the imbalance between bettors and layers.

However, while the conceptual arguments may equally apply on exchange order-driven betting markets, it is argued here that the nature of betting results in a different effect on the bid-ask spread or OR compared to that proposed by Handa, Schwartz, and Tiwari (1997), who suggest that order-driven bid-ask spreads are maximized for $p = 0.5$, that is, where there is a balance of investor opinion. In contrast, a balance of opinion/activity on the exchanges ($p = 0.5$) will result in ORs that are minimized (closest to zero) rather than maximized (extent of divergence from 1). ORs will be greatest when p is at either of its extreme values (0 or 1).

Consider a single race entrant; if neither back nor lay activity predominates, there will be no "pull" on price away from equilibrium. In contrast a preponderance of betting vis-à-vis laying activity (or vice versa) will either shorten or lengthen the odds of that race entrant. The cumulative effect (regarding all race entrants) will be to pull each price, and thereby sum of SPs, either higher or lower, thereby maximizing the divergence of sum of SPs from 1 (either above or below 1). This leads to the following hypothesis:

H1: Betfair ORs vary with the balance of betting vis-à-vis laying activity.

3.1.2 *Trading Volume/Liquidity*

Trading volume or liquidity is generally regarded as a determinant of bid-asks spreads (Madhavan 2000). Although arguments can differ,[11] the enduring view appears to be that, for both quote-driven and order-driven markets, the width of the bid-ask spread will vary inversely trading volume/liquidity (Amihud and Mendelson 1980). This view is supported by empirical evidence (Benston and Hagerman 1974; Chung and Charoenwong 1998; Stoll 1978) and is predicated on the following two hypotheses: (1) that market makers/investors normally face more competition with high volume securities than with low volume securities and/or (2) that it is generally less risky to make a market in a high-volume security, with the result that spreads for such securities would be lower, even if there were competing market makers/investors (Smidt 1971, 64). Within finance/economics, it is seen as a truism that both increasing competition and decreasing risk will reduce market inefficiencies, in the present case, by reducing bid-ask spreads.

The above is generally discussed in relation to a single security; the greater the volume of trading in a given security, the lower the bid-ask spread. It follows from this that, for given *market* of securities, if trading volumes increase for each or all securities, the lower the *market* spread. This is simply applying the cumulative or aggregative effect of the inverse relationship between trading volume and spread widths. Recall that for horse race betting markets, the market spread for the exchange, Betfair, is represented as sum of SPs – 1. Considering volume as a potential determinant of sum of prices, the question is: how might sum of SPs – 1 be affected by trading volume on the exchanges?

Applying the inverse relationship discussed above, we should expect increasing volume of trade (monies wagered) for a given race entrant to reduce the spread between the best back and lay prices. In aggregate, therefore, increasing trade (monies wagered) on all race entrants would reduce the *market* spread. At the same time, as the practical effect of increasing trade would be to increase the back price while reducing the lay price, the midpoint between the two sets of prices is not necessarily altered and neither therefore is the sum of SPs.[12] To the extent this applies, variation in volume is unlikely to alter sum of prices (bid and ask prices may vary, but the mid-point/average of these, on which SPs are calculated, is not affected). This means that, in effect, and in contrast to financial markets, variation in volume should not affect Betfair ORs. It is therefore hypothesized that

H2: Betfair ORs are invariant to changes in betting volume (monies wagered).

3.2 An Explanation Shared by Finance and Horse Race Betting Literatures

3.2.1 *Adverse Selection/Information Asymmetry*

Studies of financial markets and horse race betting conjecture that adverse selection affects the magnitude of bid-ask spreads. In both cases it is suggested that market makers (bookmakers) seek to manage adverse selection costs through the bid-ask spread (overround) (Shin 1991, 1992, 1993; Smith, Paton, and Vaughan Williams 2006). The basic idea is that the cost of dealing with insiders is passed on to outsides through higher bid-ask spreads (overrounds) (Law and Peel 2002). In financial markets (both quote-driven and order-driven markets) the cost of dealing with insiders is considered to increase as liquidity decreases. Market makers respond by increasing bid-ask spreads.

For horse race betting markets, liquidity tends to decrease as odds increase (that is, more money is wagered on favorites than on long shots). Increasing odds increases the payout to the bettor for a given wager. Given these two points, a commonplace argument is that bookmakers disproportionately compress the odds on long shots to protect against dealing with insiders (Law and Peel 2002; Schnytzer and Shilony 1995, 2003; Shin 1993; Smith, Paton, and Vaughan Williams 2006). Higher ORs are needed to allow bookmakers to pass the costs arising from insider activity on to "outsiders," that is, recreational bettors who tend to overbet long shots and underbet favorites (Cain, Law, and Peel 2003; Smith, Paton, and Vaughan Williams 2006). Such actions create a favorite-longshot bias (Vaughan Williams 1999; Vaughan Williams and Paton 1997).[13]

Adverse selection risk is used to explain the observed positive relationship between OR and number of runners in a given race (see, e.g., Smith, Paton, and Vaughan Williams 2006); $x - 1$ gets larger as the number of race entrants increases). As Michael Cain, David Law, and David Peel (2003, 270) argued: "a larger field of competitors leads to higher odds against any individual winning the event and thus higher winnings for insiders. In these circumstances bookmakers need enhanced margins to protect themselves." The effect is that sum of prices, x "increases with the number of runners as the bookie tries to recoup greater losses to the insider by raising the prices faced by outsiders" (Shin 1993, 1152).

To the extent that bettors on the exchanges may seek to manage adverse selection risks, it is reasonable to argue that the effect may be increasing bid-ask spreads. In particular, back limit orders (offers to lay a given race entrant) may be subject to increasing pricing constraint as the number of race entrants increases. As with bookmakers, the point would be to try and limit the payout to insiders as odds increase with increasing number of race entrants. Moreover, to the extent that any compression of prices occurs as *n* increases, the "gravitational pull" effect (discussed above) could be expected to reinforce any compression of odds. Thus, from the perspective of adverse selection, the following hypothesis is presented:

H3: Betfair ORs increase with number of race entrants.

The horse race betting literature posits three further manifestations of adverse selection; bookmaker ORs are expected to vary according to (1) grade of race (Vaughan Williams and Paton 1997), (2) age of horses competing in a given race (Peirson and Smith 2010), and (3) type of race (handicap versus nonhandicap) (Vaughan Williams and Paton 1997). The underpinning argument to (1) is that insider trading is deemed more or most likely to occur in lower grade races, for which prize money is less and media interest is often minimal or absent (Smith, Paton, and Vaughan Williams 2006). Aware of this, bookmakers increase their ORs accordingly in order to pass on adverse selection costs onto "outsiders." Regarding (2), the argument here concerns the relationship between age and publicly available information about a horse's ability; younger horses (e.g., two-year olds for Flat racing) typically have less racetrack experience, and thereby less public form, than older horses, which, by their greater age, have generally more publicly exposed form.[14] Increasing public information is considered to reduce the potential for insider trading; the horse's ability is more exposed (Vaughan Williams and Paton 1997). As such, an inverse relationship is predicted between age and bookmaker OR; lower aged races attract higher ORs as bookmakers seek to manage adverse selection costs (Peirson and Smith 2010).

Finally, for reasons similar to those presented for (2) above, bookmaker ORs are expected to be lower for handicap races compared to nonhandicap races. Insider trading is expected to be more prevalent in nonhandicap events (see Bruce and Johnson 2003) because, in such events, some or all of the participants may have little or no prior public form. There may be a degree of owner/trainer influence on the race conditions (e.g., via the assignment of weight in claiming races), the level of media and official scrutiny of the running of the race is lower, and the magnitude of gains from betting coups relative to prize money is higher. Each of these factors would, according to Hyun Song Shin (1991, 1992, 1993), be likely to result in relatively high and low ORs in nonhandicap and handicap races, respectively.

For (1), (2) and (3) the implication is that bookmakers actively manipulate prices (in particular, by compressing the prices of long shots) in response to information asymmetry and adverse selection risks. Bettors placing limit order bets on Betfair are, as suggested, acting as de facto market makers; they are providing liquidity and immediacy to the market. Given this, the arguments pertaining to how bettors on the exchanges may seek to manage information asymmetry in relation to number of race entrants are applied here to (1), (2) and (3) to further test adverse selection as an explanation for Betfair SP-based ORs. The following hypotheses are presented:

H4: Betfair ORs are inversely related to grade of race.

H5: Betfair ORs are inversely related to age limit of race.

H6: Betfair ORs are higher for nonhandicap races than for handicap races.

3.3 Explanations from Horse Race Betting

3.3.1 *Last Race Effect*

A commonly observed empirical phenomenon in racehorse betting markets concerns (on-course) bettors' particular appetite to bet on the last race of the day in order to recoup losses incurred earlier in the race meeting. This behavior, known colloquially to bookmakers as "Charlie chasing," is formalized in both Harry Markowitz's (1952) model of utility and more recently David Peel and David Law's (2009, 253) "general non-expected utility model" as behavior that is "risk-seeking over losses." It relates to the strong probability that, having engaged in regular wagering throughout the course of a race meeting, bettors will be facing a loss on their trading prior to the last race. Initial risk-averseness regarding losses is then replaced by risk-seeking behavior as bettors attempt to retrieve a favorable trading outcome (Peel and Law 2009).

Implicit in this type of behavior is the notion that bettors may not view the menu of betting opportunities in a particular race as an isolated decision problem but rather as part of a series of decisions within a larger set of events, for example, the races constituting a given race meeting. As such, the behavioral influences on a decision that form part of a larger set of related decisions may be expected to differ from those that operate on isolated decisions (see, e.g., Keren and Wagenaar 1985; Rachlin 1990). In this context a number of contributions point to the relationship between declining capital and increasing risk preference (see, e.g., McGlothlin 1956; Ali 1977; Gilovich 1983; Asch, Malkiel, and Quandt 1984; Metzger 1985; Gilovich and Douglas 1986; Golec and Tamarkin 1995), which results in overbetting in the last race, particularly on long shots (Johnson and Bruce 1993).

To the extent that exchange bettors also view betting as a series of decisions within the wider context of the race meeting or race day, it is possible to argue that a "last race effect" also may also arise. Bettors may seek to recoup losses and retrieve a favorable trading outcome by wagering on the last race. If so, and as laying can only ever represent an even-money "win" outcome (if the horse loses, the layer retains the stake money), attempts to recoup losses may push more bettors toward betting to win, particularly betting long shots to win. This is in order to achieve greater returns for a given wager. The result will be a greater balance of activity toward betting rather than laying, with concomitant effects on odds; these will generally shorten, thereby increasing the sum of prices and OR. It is therefore hypothesized that

H7: Betfair ORs for the last race are significantly greater than the OR for prior races at a given race meeting.

3.3.2 *Flat versus National Hunt Races*

The study's final hypothesis focuses on the possibility that part of the explanation for variation in Betfair ORs may lie in differences between Flat and National Hunt (jumps) racing. The argument is based on the idea that back vis-à-vis lay activity may vary between Flat and National Hunt (NH) racing. Betting activity may vary because of the greater number of unpredictable factors for NH racing compared to Flat racing. The greater number of unpredictable factors may encourage more *laying* compared to betting activity, with concomitant effects on prices (lengthened), sum of prices, and Betfair OR (lower for NH races compared to Flat races).

The increased unpredictability for NH racing includes, for instance, the presence of hurdles and fences. These introduce the possibility that horses may fall or be brought down. The longer distances covered and the generally more testing conditions (NH racing is predominantly a winter sport) mean that it is quite common for horses to "pull up" (i.e., drop out) without completing the race. The NH season is frequently badly disrupted by weather-based cancellations, which negatively affects the analysis and interpretation of form. Finally, there may be an effect relating to the competition for betting revenue which NH racing has been forced to confront over the past two decades in the form of all-weather Flat racing. This form of racing, relatively invulnerable to the vagaries of weather and offering unparalleled consistency in terms of track conditions, has ended NH racing's monopoly of horse race betting opportunities during the winter months. For the reasons outlined above, the study hypotheses that

H8: Betfair ORs are significantly higher for Flat racing than for NH racing.

3.4 Control Variable

Bookmakers are known to be active on Betfair (O'Connor 2011). To the extent that this activity and the experience of bookmakers (e.g., regarding the laying of multiple runners) influence prices on Betfair, the study controls for the potential interplay between bookmaker and Betfair ORs.

4 Data Collection and Measurement

To test our hypotheses we analyzed data on 2,184 horse races that took place in the United Kingdom between autumn 2008 and spring 2010. Sample races were drawn from both Flat and NH racing (in roughly equal measure) and include an approximately equal number of handicap vis-à-vis nonhandicap races. Data collection was spread across 330 race meetings and all 60 racetracks located throughout the United Kingdom. All days of the week are involved, including Sunday. For each race in the sample set both grade of race and age limit were recorded. The sample set includes 330 "last races." Sample statistics are presented in table 16.1.

Table 16.1 Sample Statistics

Variable	Number
Type of race	
Flat races and National Hunt (NH) races	1,100
	1,084
Type of race event	
Nonhandicap races	1,056
Handicap races	1,128
Racetracks	60
Last races	330
Prior races	1,854
Race day	
1. Monday	275
2. Tuesday	266
3. Wednesday	243
4. Thursday	250
5. Friday	372
6. Saturday	494
7. Sunday	284

Note: $n = 2,184$ races.

The study's focus and dependent variable is Betfair SP-based ORs. Data on these were collected via the Timeform website (www.timeform.com), and cross-checked using data collected directly from Betfair's own website (www.betfair.com). The data were also randomly audited and sum of prices calculated to confirm the accuracy of the reported ORs. No errors were found in the published data, though the reporting to only two decimal places potentially restricts the sensitivity of the empirical analysis. Bookmaker SP-based ORs were similarly collected from Timeform and random audited to confirm accuracy. Again, the reporting of ORs to only two decimal places is potentially restricting.

For the 2,184-race sample, Betfair ORs range from −0.14 (sum of prices: 0.86) to +0.12 (sum of prices: 1.12), with a mean of 1.0047, standard deviation of 0.0189, and mode of 1.00 ($n = 544$ for mode of 1.00). Bookmaker ORs range from −0.04 (sum of prices: 96%) to +0.42 (sum of prices 1.42), with mean of 1.1834 and standard deviation of 0.0628.

Betfair provides regularly updated information on monies matched for each horse race. Total volume traded for each sample race was recording immediately prior to the start of a given race. Monies wagered "in play" (in running) were excluded from the analysis. These monies relate primarily to a horse's "real time" performance and do not represent "normal" betting activity. Trading volumes for the 2,184-race sample ranged from £89,160 to £3,592,042, with a mean of £559,706. Given the skewness of the data, logN of trading volumes was used in the analysis.

For the purposes of this study, a measure of "balance of opinion"/"balance of betting activity" was calculated as follows. First, for each of the 2,184 races in the sample, Betfair prices for all race participants were recorded at two points during the betting auction: at 10 minutes prior to the scheduled start time of the race and immediately at "the off."[15] The prices were recorded directly from Betfair's website and included both back and lay odds for each of the two measurement points.[16]

Second, from back and lay odds, average prices were calculated for each of the two measurement points. Third, average price at 10 minutes was divided into average "off" price in order to determine both the direction and magnitude of any odds change for each race entrant over the 10-minute period. Finally, for each race in the 2,184-race sample, race entrants' price differentials were summated to determine the overall balance of trading activity. Overall balances in favor of betting were coded 1, lay balances were coded −1, and equilibrium positions were coded 0.

For analysis handicap races were coded 1 and nonhandicap races 0. Flat races were coded 0, NH (jumps) races 1. Grade of race was coded as per official grading, from 1 (top grade) to 6 (lowest). To assess possible age limit effects two-year-old Flat races and three/four-year-old NH races were coded 0 and all other races 1. For each race meeting prior races (races before the last race) were coded 0, the last race 1. "Number of runners" is based on the number of *starters* for each race in the 2,184-race sample (i.e., withdrawn horses were not included). Numbers range from a minimum of 2 to a maximum of 22 runners (mean = 9.989; standard deviation 3.48).

5 Analysis and Results

Table 16.2 presents descriptive statistics, including means, standard deviations, and Pearson correlation coefficients for the variables examined in this study.

The hypotheses were tested using ordinary least squares (OLS) multiple regression analysis. Tests for multicollinearity among the predictor variables revealed variance inflation factors and tolerances substantially within acceptable limits (< 1.2 and > 0.85, respectively).[17] These test results confirm no significant problems in terms of using OLS multiple regression analysis to test the study's predictions (Belsley, Kuh, and Welsch 1980).

The dependent variable for analysis is Betfair OR. For robustness, "Betfair OR" is operationalized in two related but distinct ways: as sum of prices (SPs) – 1 and as sum of SPs – 1/runners. The former represents the OR for a given race. The latter establishes the OR per runner for a given race. This dual operationalization of Betfair OR results in two regression models, with each taking the following standard form:

$$Y = \beta_0 + \beta_1 + \cdots \beta_n + \acute{\varepsilon},$$

where Y = dependent variable; sum of SPs – 1 (model 1) and sum of SPs – 1/runners (model 2)

Table 16.2 Descriptive Statistics, including Pearson Correlation Coefficients

	Mean	SD	1	2	3	4	5	6	7	8	9	10
1BetOR	0.005	0.019	-									
2IObal	0.085	0.777	0.25	-								
3LgMon	13.052	0.505	0.12	0.01	-							
4Run[a]	9.989	3.540	0.22	0.05	−0.14	-						
5Grade	4.165	1.298	−0.15	−0.08	−0.33	0.05	-					
6Age	3.736	1.428	−0.01	0.00	0.07	−0.14	−0.06	-				
7H'cap	0.511	0.500	0.07	0.01	−0.20	0.14	−0.07	0.28	-			
8Last	0.152	0.359	0.06	0.01	0.05	0.12	0.19	0.01	0.11	-		
9FlatNH	0.494	0.500	−0.02	0.05	−0.15	0.02	−0.21	0.08	−0.05	−0.01	-	
10OR	18.341	6.189	0.21	0.03	−0.19	0.79	0.10	−0.07	0.25	0.27	−0.11	-

Note: $n = 2,184$
BetOR = Betfair sum of prices −1.
IObal = in-out-balance of price movements (in = prices shortening; out = prices drifting).
LgMon = LogN of monies (trading volume).
Run = number of runners in a given race.
Grade = grade of race.
Age = age limit for a given race.
H'cap = type of race (handicap versus nonhandicap).
Last = last race of a given race meeting.
FlatNH = type of race event (Flat versus National Hunt).
OR = bookmaker sum of prices −1.
SD = standard deviation.
Correlations above 0.05 significant at $p < 0.05$ or better.
[a] Minimum 2 runners; maximum 22 runners.

$$\beta_{1,\ldots}\beta_n = \text{independent variables.}$$

The results of the regression analyses are shown in tables 16.3 and 16.4. The results across both tables support hypotheses 1, 3, and 7. That is, balance of activity, number of runners, and last race effect all influence Betfair ORs in the predicted way. For instance, a balance of activity in favor of backing rather than laying increases the reported OR. Betfair ORs are also higher for increasing number of runners and for the last race compared to prior races at a given race meeting.

The results across the two tables also show significant effects for trading volume (logmonies), grade of race, and type of race (handicap versus nonhandicap). However, findings here are contrary to expectations. For instance, Betfair ORs increase (rather than decrease) with both trading volume and grade of race. ORs are also higher for handicap races compared to nonhandicap races. Nonsignificant results are reported for age limit of the race and type of race event (Flat versus NH). The study's findings are discussed in the concluding section.

Table 16.3 Test of Hypotheses, Betfair Sum of SPs – 1 as Dependent Variable

Variable	Standardized Beta Coefficient	*t*-value
Inoutbalance (H1)	0.223	7.130***
Logmonies (H2)	0.143	4.081***
Number of runners (H3)	0.128	2.336*
Grade of race (H4)	−0.208	−5.811***
Age limit of race (H5)	0.037	0.758
Handicap versus nonhandicap (H6)	0.069	1.977*
Last race (H7)	0.072	2.008*
Flat versus NH (H8)	−0.070	−1.438
Bookmaker sum of SPs – 1 (control)	0.159	2.805**
Adjusted $R^2 = 21.0\%$	F value = 24.556***	

Note: $n = 2,184$.
*p <0.05; **p <0.01; ***p <0.001.

Table 16.4 Test of Hypotheses, Betfair Sum of SPs – 1/runner as Dependent Variable[a]

Variable	Standardized Beta Coefficient	*t*-value
Inoutbalance (H1)	0.214	6.292***
LogMonies (H2)	0.116	3.028**
Grade of race (H4)	−0.173	−4.423***
Age limit of race (H5)	−0.028	0.549
Handicap versus nonhandicap (H6)	0.081	2.234*
Last race (H7)	0.084	2.450*
Flat versus NH (H8)	−0.037	−0.709
Control[b]		
Adjusted $R^2 = 11.6\%$	F value = 14.542***	

Note: $n = 2,184$.
[a] H3 (number of runners) n/a for this analysis.
[b] Bookmaker sum of SPs – 1/runner omitted due to unusually high value (t-value > 100.00).
**p* <0.05; ***p* <0.01; ****p* <0.001.

5.1 Sensitivity Analysis

Analytical robustness of the empirical examination was assessed via two further tests of the data. First, to check whether isolating the Betfair OR from sum of prices affects the analysis, sum of SPs and sum of SPs/runner were substituted for sum of SPs – 1 and sum of SPs – 1/runner as the dependent variables. The regression analyses were then repeated as per above. Results are qualitatively very similar to those reported in tables 16.3 and 16.4 except that, for sum of SPs/runner, type of race event shows a significant effect while grade of race does not. These changes compared to the previous analysis suggest that both type of race event and grade of race may be sensitive to variation in the number of race entrants.

Second, therefore, given the above, partial correlation analysis was conducted to test for possible spurious relationships involving the eight predictor variables and Betfair ORs. Partial correlation analysis allows examination of the effect of one variable, the control, on the relationship between two other variables (Meng, Rosenthal, and Rubin 1992). For the present study, the aim was to examine the effect of number of race entrants (B) on the association between each predictor variable (A) and Betfair OR (C) by computing a partial correlation coefficient to remove the association that B has with both A and C (Blaikie 2003).

Results for the partial correlation analysis suggest that, apart from last race effect (the significance of which is reduced), each of the predictor variables found to be significant in the main analysis remains broadly as significant (no change in the value of p) when controlling for variation in the number of race entrants. Put slightly differently, partial correlation analysis suggests that each predictor variable exerts an effect on Betfair OR that is independent of the number of race entrants. No spurious relationships are revealed.

6 Discussion and Conclusions

The study reported in this chapter sought to explain both the existence and variation in exchange (Betfair) overrounds. Since December 2007, Betfair, the leading person-to-person Internet betting exchange worldwide (O'Connor 2007, 2011), has reported its own starting prices and accompanying overrounds. Based on our sample, we find that Betfair overrounds varied by a factor of 26 percent, from −14 to +12 percent. The empirical results imply that "balance of activity," trading volume, number of race entrants, grade of race, type of race (handicap versus nonhandicap), and bookmaker overrounds are key determinants of the existence and magnitude of Betfair overrounds. In contrast, insignificant results were reported for age limit and type of race event (Flat versus National Hunt).

The finding that "balance of activity" may be a key explanatory variable supports arguments in the finance literature concerning the determinants of bid-ask spreads in order-driven markets (Handa, Schwartz, and Tiwari 1998, 2003). The basis of these arguments is that an "imbalance of opinion/activity," combined with the "gravitational pull" effect, will serve to minimize bid-ask spreads, whereas a "balance of opinion" will maximize the magnitude of spreads. This study has argued that, in the context of exchange betting, an imbalance of activity (for a given race proportionally greater backing to win/laying to lose) will combine with "gravitational pull" effect to move Betfair ORs either higher or lower. Results support these theoretical arguments, suggesting that bettor behavior on the exchanges may, to some extent at least, mirror investor behavior in order-driven financial markets more generally.

The results for trading volume are intriguing, as are the results for type of race, grade of race, and number of race entrants. Contrary to expectations, the study reports a positive and significant relationship between volume and Betfair OR. It is possible that increased trading volume may reflect increased betting rather than laying activity. The significant correlation coefficients between trading volume and grade of race and between grade of race and balance of activity (see table 16.2) seem to support this conjecture; they suggest that a greater proportion of the increasing trading volume may be directed at betting to win rather than laying to lose, with concomitant effect on Betfair ORs. The issue merits further attention.

Consistent with hypothesis 3, the study finds a significant positive relationship between number of race entrants and Betfair OR. That said, results here are minimally significant ($p < 0.05$) and contrast with the strong relationship between bookmaker ORs and number of race entrants (correlation coefficient 0.79; see table 16.2). Results for grade of race and type of race are contrary to expectations. For instance, instead of finding increasing ORs for lower grade races compared to higher grade races, results show higher Betfair ORs for higher grade races. Taken alongside results for number of race entrants and type of race, the present findings raise questions about the nature of betting activity on the exchanges, particularly bettors' response to adverse selection risks. In essence, findings suggest reasons other than information asymmetry and adverse selection that may explain the existence of exchange bid-ask spreads. Again, these reasons might include an increasing tendency to bet to win rather than lay to lose in higher grade races. The correlation coefficients shown in table 16.2 may help to explain the results for grade of race; the higher the grade, the greater the volume of monies wagered aimed at betting to win, which leads to higher ORs on Betfair. The correlations shown in table 16.2 do not support this explanation for type of race; trading volume is higher for nonhandicaps vis-à-vis handicaps, yet Befair ORs are higher for the latter compared to the former. The question of exchange betting behavior in relation to information asymmetry and adverse selection risks is an issue which merits closer scrutiny. The present results provide a glimpse into potential idiosyncrasies of exchange-based bettor behavior.

Notwithstanding the above, results for last race effect suggest some familiar bettor behavior on the exchanges. Results showing a (weakly) significant effect support extant arguments that bettors may often "Charlie chase" on the last race, placing bets aimed at recovering the day's losses. Given this observed racetrack behavior, there seems little reason to think that such behavior would not also feature on the exchanges, with increasing betting to win activity increasing Betfair ORs on the last race compared to previous races. Finally, the insignificant results for type of race event and age limit of race may suggest that bettors neither differentiate between flat and NH racing nor adjust their betting according to the degree of public form available for younger vis-à-vis older race entrants.

Taken as a whole, it appears from the present analysis that betting activity on the exchanges may show some similarities with investing behavior in order-driven markets (balance of activity and gravitational pull effect) while also demonstrating behavior unique to horse race gambling (e.g., last race effect). Extant arguments in both the finance and horse racing literatures appear to help explain the existence of and variation in exchange (Betfair) ORs. That said, there may be behaviors, such as behavior toward grade of race, which may be potentially unique to exchange betting. Examining betting activity on the exchanges appears to offer a potentially fruitful line of inquiry for researchers interested in empirical tests of the efficient market hypothesis. Additional microstructure analysis may be especially fertile ground for future research.

Notes

1. It has always been theoretically possible for bettors to lay—by backing all other possibilities to happen. For instance, a bettor could, before the exchanges, lay any given horse(s) in any given race by backing all other race entrants to win. That said, nontrivial practical and cognitive difficulties have always precluded this form of laying. For two-runner races it is possible to lay one entrant by backing the other. For three-runner races and above, however, the practical and cognitive difficulties of laying one horse by backing all others increase exponentially.
2. Immediacy is the ability investors have to buy or sell assets at any time the market is open (Handa, Schwartz, and Tiwari 1998).
3. The trader placing (buy or sell) limit orders is providing both prices and immediacy to the market, thereby acting, in essence, as de facto market maker.
4. Exchange betting is discussed further in the following section.
5. Shin (1993, 1142) argued that the measurability of bookmaker overrounds in betting markets is "in marked contrast to more sophisticated markets, such as the stock market, in which the spread varies across assets and also across volumes traded." The magnitude of the overround can vary during the period of betting (the betting auction). These overrounds are observable by racegoers but generally not reported and therefore difficult to access. Overrounds based on starting prices are, however, readily accessible and measurable through sources such as Timeform and Raceform.
6. The reference to "new approach" relates to changes to the way industry SPs are determined. Previous to 2007, SPs were determined by the official SP returner(s), who had the discretion to investigate the entire "betting ring" to find the best odds being traded. The new rules, set by the Starting Price Regulatory Commission, require that (1) a set number of "pitches" are surveyed (normally 10 and normally to involve the same list of bookmakers on an ongoing basis) and (2) the "majority price" should be used if prices among the list of 10 are not uniform.
7. For bookmaker-led markets, divergence from "true value" is typically unidirectional; sum of prices is usually greater than 1, which means (1) a positive OR and (2) that prices at which recorded transactions occur tend to be below but not above their objective value.
8. While a horse race example has been used to illustrate the OR and betting market operational efficiency, the points made apply to any event, sporting or otherwise, for which there is more than one possible outcome and for which betting odds are available.
9. Data collection procedures are described in section 4.
10. The reader is referred to Handa, Schwartz, and Tiwari (1998, 2003) for a thorough exposition of the arguments and analysis relating to gravitational pull effects on bid-ask spreads in an order-driven market.
11. Arguments differ in that, on the one hand, if trading volumes are generally low, market makers will find it difficult to adjust their inventory levels and will increase their spreads

to compensate. On the other hand, high trading volumes may suggest investors trading on superior information. Market makers will therefore increase their bid-ask spreads to compensate for perceived adverse selection risk.

12. For instance, increasing trade (monies wagered) on a given race entrant will likely increase the competition among bettors seeking to ensure that their limit order is "matched" and/or that they are able to execute a market order to match an existing limit order. The effect will be to bring back and lay odds closer together, potentially reaching the limit of closeness as determined by the minimum "tick" price available for a given odds range.
13. The favorite-longshot bias refers to the empirical observation that compared to long shots, implied probabilities for favorites more closely match their objective win probabilities (Schnytzer and Shilony 1995).
14. In the limit (beginning of Flat race turf season), *all* entrants in a two-year-old maiden race will be competing in their first race under license. In this context, there is *no* publicly available information by which to adjudge a horse's chances.
15. The time period for recording prices, beginning at 10 minutes before the scheduled start time, is based on the observation that the vast majority of trading activity on Betfair tends to occur in the last 10 minutes before the start of a race. During this period betting volumes frequently increase at least six-fold, from around £100,000 to in excess of £600,000.
16. Piloting of data collection procedures showed that a maximum of 44 data points (equivalent to 22 runners) could be accurately recorded within Betfair's automatic "refresh" period (approximately 25-second intervals). Data collection was therefore limited to races comprising no more than 22 runners. In turn, creation of a 22-runner race limit resulted in the omission of several races from the data collection process. In total, because field size exceeded 22 runners, 25 races were omitted from, in total, 23 of the 330 race meetings. The 2,184-race sample is the net number of races, following the omission of the 25 races.
17. Basically the closer both values are to 1, the lower the collinearity among the predictor variables (Blaikie 2003).

References

Ali, Mukhtar M. 1977. Probability and utility estimates for racetrack bettors. *Journal of Political Economy* 85(4):803–815.

Amihud, Yakov, and Haim Mendelson. 1980. Dealership market: Market making with inventory. *Journal of Financial Economics* 8(1):31–53.

Asch, Peter, Burton G. Malkiel, and Richard E. Quandt. 1984. Market efficiency in racetrack betting. *Journal of Business* 57(2):165–175.

Belsley, David A., Edwin Kuh, and Roy E. Welsch. 1980. *Regression diagnostics: Identifying influential data and sources of collinearity.* New York: John Wiley & Sons.

Benston, George J., and Robert L. Hagerman. 1974. Determinants of bid-ask spreads in the over-the-counter market. *Journal of Financial Economics* 1(4):353–364.

Blaikie, Norman. 2003. *Analyzing quantitative data.* London: Sage.

Bruce, Alistair C., and Johnnie E. V. Johnson. 2003. Market ecology and decision behaviour in state-contingent claims markets. *Journal of Economic Behavior & Organization* 56(2): 199–217.

Cain, Michael, David Law, and David A. Peel. 2003. The favourite-longshot bias, bookmaker margins and insider trading in a variety of betting markets. *Bulletin of Economic Research* 55(3):263–273.

Chakravarty, Sugato, and Craig W. Holden. 1995. An integrated model of market and limit orders. *Journal of Financial Intermediation* 4(3):213–241.

Chung, Kee H., and Charlie Charoenwong. 1998. Insider trading and the bid-ask spread. *Financial Review* 33(3):1–20.

Chung, Kee H., Bonnie F. Van Ness, and Robert A. Van Ness. 1999. Limit orders and the bid-ask spread. *Journal of Financial Economics* 53(2):255–287.

Coleman, Les. 2007. Measurement of insider trading in wagering markets. *Applied Financial Economics* 17(5):351–356.

Crafts, N. F. R. 1985. Some evidence of insider knowledge in horse race betting in Britain. *Economica* 52(207): 295–304.

Gilovich, Thomas. 1983. Biased evaluation and persistence in gambling. *Journal of Personality and Social Psychology* 44(6):1110–1126.

Gilovich, Thomas, and Christine Douglas. 1986. Biased evaluations of randomly determined gambling outcomes. *Journal of Experimental Social Psychology* 22(3):228–241.

Goettler, Ronald L., Christine A. Parlour, and Uday Rajan. 2009. Informed traders and limit order markets. *Journal of Financial Economics* 93(1):67–87.

Golec, Joseph, and Maurry Tamarkin. 1995. Do bettors prefer long shots because they are risk-lovers, or are they just overconfident? *Journal of Risk and Uncertainty* 11(1):51–64.

Handa, Puneet, and Robert A. Schwartz. 1996. Limit order trading. *Journal of Finance* 51(5):1835–1861.

Handa, Puneet, Robert A. Schwartz, and Ashish Tiwari. 1997. *Determinants of the bid-ask spread in an order driven market.* Working Paper. University of Iowa and City University of New York.

——. 1998. The ecology of an order-driven market. *Journal of Portfolio Management* 24(2): 47–55.

——. 2003. Quote setting and price formation in an order driven market. *Journal of Financial Markets* 6(4):461–489.

Johnson, Johnnie E. V., and Alistair C. Bruce. 1993. Gluck's second law: An empirical investigation of horserace betting in early and late races. *Psychological Reports* 72(2):1251–1258.

Keren, Gideon B., and Willem A. Wagenaar. 1985. On the psychology of playing blackjack: Normative and descriptive considerations with implications for decision theory. *Journal of Experimental Psychology: General* 114(2):133–158.

Kumar, P. C. 2004. Bid-ask spreads in the U.S. equity markets. *Quarterly Journal of Business and Economics* 42(3/4):85–111.

Law, David, and David P. Peel. 2002. Insider trading, herding behaviour and market plungers in the British horse-race betting market. *Economica* 69(274):327–338.

Levin, Eric J., and Robert E. Wright. 2004. Estimating the profit markup component of the bid-ask spread: Evidence from the London Stock Exchange. *Quarterly Review of Economics and Finance* 44(1):1–19.

Madhavan, Ananth. 2000. Market microstructure: A survey. *Journal of Financial Markets* 3(3):205–258.

Marginson, David. 2010. Market inefficiency in person-to-person betting: Examining "known loser" insider trading on the exchanges. *Southern Economic Journal* 76(4):932–952.

Markowitz, Harry. 1952. The utility of wealth. *Journal of Political Economy* 60(2):151–158.

McGlothlin, William H. 1956. Stability of choices among uncertain alternatives. *American Journal of Psychology* 69(4):604–615.

Meng, Xiao-li, Robert Rosenthal, and Donald B. Rubin. 1992. Comparing correlated correlation coefficients. *Psychological Bulletin* 111(1):172–175.

Metzger, Mary Ann. 1985. Biases in betting: An application of laboratory findings. *Psychological Reports* 56(3):883–888.

O'Connor, Niall. 2007. Betting exchanges—Disruptive innovation at work. Bettingmarket.com.

———. 2011. Betfair—An alternative exchange. Bettingmarket.com.

Peel, David A., and David Law. 2009. A more general non-expected utility model as an explanation of gambling outcomes for individuals and markets. *Economica* 76(302): 251–263.

Peirson, John, and Michael A. Smith. 2010. Expert analysis and insider information in horse race betting: Regulating informed market behavior. *Southern Economic Journal* 76(4): 976–992.

Rachlin, Howard. 1990. Why do people gamble and keep gambling despite heavy losses? *Psychological Science* 1(5): 294–297.

Schnytzer, Adi, and Yuval Shilony. 1995. Inside information in a betting market. *Economic Journal* 105(431):963–971.

Schnytzer, A. and Shilony, Y. (2003) 'Is the presence of insider trading necessary to give rise to a favourite-longshot bias?' In the Economics of Gambling, ed. L. Vaughan-Williams, London: Routledge: 14–17.

Shin, Hyun Song. 1991. Optimal betting odds against insider traders. *Economic Journal* 101(408):1179–1185.

———. 1992. Prices of state contingent claims with insider traders, and the favourite-longshot bias. *Economic Journal* 102(411):426–435.

———. 1993. Measuring the incidence of insider trading in a market for state-contingent claims. *Economic Journal* 103(420):1141–1153.

Smidt, Seymour. 1971. Which road to an efficient stock market: Free competition or regulated monopoly? *Financial Analysts Journal*, Sept.–Oct., 18–69.

Smith, Michael A., David Paton, and Leighton Vaughan Williams. 2006. Market efficiency in person-to-person betting. *Economica* 73(292):673–689.

Stigler, George J. 1967. Imperfections in the capital market. *Journal of Political Economy* 75(3):287–292.

Stoll, Hans R. 1978. The supply of dealer services in securities markets. *Journal of Finance* 33(4):1133–1151.

Theissen, Erik. 2000. Market structure, informational efficiency and liquidity: An experimental comparison of auction and dealer markets. *Journal of Financial Markets* 3(4):333–363.

Vaughan Williams, Leighton. 1999. Information efficiency in betting markets: A survey. *Bulletin of Economic Research* 51(1):1–30.

———, ed. 2005. *Information efficiency in financial and betting markets.* Cambridge: Cambridge University Press.

Vaughan Williams, Leighton, and David Paton. 1997. Why is there a favourite-longshot bias in British racetrack betting markets? *Economic Journal* 107(440):150–158.

CHAPTER 17

INSIDER TRADING IN BETTING MARKETS

ADI SCHNYTZER

1 Introduction

A betting market is a market in which the agents trade bets, where a bet is a contingent commodity (or portfolio of contingent and, perhaps, regular commodities). One side to the trade believes that the contingency (or contingencies) will occur while the other believes it (they) will not. Accordingly, both are willing to enter into a trade that, by definition, is subject to either risk (there are known or estimable probabilities associated with contingencies) or uncertainty (probabilities are unknown and inestimable). A betting market may be either legal or illegal; this chapter deals only with legal betting markets.[1] Examples of operations in betting markets include betting on two flies crawling up a wall, buying or selling an option,[2] trading in real estate, playing roulette in a casino, and betting on a horse in the Melbourne Cup.

These examples are sufficient to determine when insiders are likely to be found in the market and thus determine the boundaries of this chapter. When Australians bet on which of two flies crawling up a wall will first reach the ceiling, they are generally[3] betting in a situation where probabilities are neither known nor estimable, and thus there is no relevant information to determine winning probabilities. Accordingly, there can be no human insiders and betting on such an event may be termed *pure gambling* and will not be discussed further.

Buying or selling options are bets in which payoffs are determined either by the price of the underlying asset(s) on the expiry date of the option if the option is held to expiration or by the price of the option when sold if this occurs before expiry. In either case, there is sufficient information regarding the underlying asset(s) to attempt prediction of the likely value of the option at some point in the future. Whether such predictions, if made by outsiders, are economically useful is a matter of debate and need not be discussed further.[4] However, it is undeniable that company insiders who are privy to important information not yet publically disseminated are able to predict the short-term direction in the price of the relevant share.[5] Accordingly, this type of betting market certainly falls within the domain of this chapter. However, since insider trading in financial markets is illegal in the world's major markets, and since very little has been published in this area,[6] this type of betting, too, will be discussed no further. It should be noted that these factors are related; illegal insiders will wish to enter the market in such a way as to be undetectable and given the complexity of derivative markets,[7] this renders data collection by researchers (and the SEC) extremely difficult.

Trading in real estate is similar to trading in derivatives markets in that price prediction is extremely difficult. This difficulty is exacerbated by the fact that most properties are literally unique.[8] Also, since the current owner generally has inside information with regard to the property, the prospective buyer is subject to adverse selection. However, the exploitation of inside information in this market is virtually immeasurable and thus, short perhaps of courtroom anecdotes, has not been a subject of considerable research.[9]

Playing roulette in a casino is, in some sense, a mirror image of betting on two flies climbing up a wall. If the latter is pure gambling because the bettors are subject to uncertainty and no insider can get an edge,[10] in roulette all probabilities are known and so, again, insiders have no edge. Accordingly, insiders do not feature in such betting markets. To summarize, inside information cannot exist when either all or no probabilities are known to all participants in a betting market.

This leaves the last example, betting on a horse in the Melbourne Cup, and here the conditions are ripe for the presence of inside traders and, possibly, for their detection. First, the only people connected to a runner in the race who are not permitted to bet on the race are the jockeys. On the other hand, in harness racing, drivers are permitted to bet and in greyhound racing all the dogs' connections may bet. Hence in principle there will be asymmetry of information, with the animals' connections having a more accurate idea of the true winning probabilities in the race than the betting public at large, and the insiders will be able to bet freely should they wish to do so. The focus in this chapter will thus be on the three racing betting markets, gallops, harness, and greyhounds.[11] A key issue to be addressed is whether the presence of betting insiders at the track implies that their presence is easily measured. It will be shown that this is a function of the microstructure of the particular betting market. In some markets the impact of insider trading is readily measured while in others it has hitherto proven virtually impossible. The different microstructures of betting markets and the implications for insider trading and its measurement are presented in the next section.

2 The Microstructure of Betting Markets

In terms of the functioning of insiders,[12] the microstructure of betting markets has three dimensions, the types of bet offered, the legal betting mechanism(s), and the location of available mechanisms—at the track where the race is held (henceforth on-course) and/or elsewhere (henceforth off-course). While there are many types of bets available,[13] for the purposes of this chapter, only the most common type, win betting, will be considered.[14] The most common betting mechanism worldwide is the pari-mutuel, also known as the totalizator and henceforth referred to simply as the tote.

The relevant locations for betting are on-course, off-course, and online (Internet betting). In any country or state that permits betting, it is always permitted on-course, and it make sense to assume that insiders will generally bet at the track unless considerably better odds are available off-course, since they may thereby be sure of the state of their horse or dog shortly before the race.[15] The legality or otherwise of off-course betting varies considerably from country to country and will be considered below wherever relevant to insider trading.

The tote is a betting mechanism that takes bets on all horses and then, at the end of betting, removes its profit from the total pool. This tends to be in the range of 12 to 15 percent. The remaining money is then divided among the bets on the winning horse. Thus if $1,000 is bet altogether on all horses in a given race and the tote take is 15 percent, then $850 is left in the pool. If $100 of the original $1,000 had been bet on the winning horse, then each dollar bet would yield a payout of $8.50, that is, a profit of $7.50 per dollar.[16] The only complication that may arise in the tote market is the case where most of the money is bet on one horse. Suppose, then, that $900 of the $1000 had been bet on the winning horse. This implies, with a payout of $850 among the winning bettors, that each dollar bet on a winner yielded a loss of 10 cents in the dollar![17] To avoid this patently unfair situation, totes guarantee a minimal payout of either $1.05 or $1.10 for each dollar bet on a winning horse.

The interesting feature of the tote is that it is not a seller of (contingent) commodities in the classic sense. The owners make a more or less fixed[18] rate of return given by the take and maximize nothing except possibly revenue.[19] In particular, the tote owners do not set winning odds, these being determined entirely by the relative demand for different horses of the betting population. Hence the owners of the tote are untroubled by insider trading. There is a tote on every racetrack in the world where gambling is legal.[20]

A second betting mechanism related to the tote is known as SP (starting price) betting in Commonwealth countries and is similar to the mechanism whereby odds paid on winning bets by bookmakers who operate legally in the State of Nevada (USA) are determined. Basically this is a mechanism whereby the payouts to winning bets are determined by the final odds available at the end of betting with bookmakers on-course, in the case of SP betting, and somewhat less than the on-course tote payout in the case of Nevada betting.[21] In such cases, the bookmakers who own the services would appear to be acting as either a remote tote or a service that provides odds out of their control, but this is not so. Since these providers may themselves bet with either the tote (in the United States) or with on-course bookmakers (in the United Kingdom), it can be shown that they may be able to manipulate And the final odds; thus insiders are unlikely to be interested in this mechanism.[22]

The third betting mechanism, and the one that has generated the most research with respect to insider trading, is the (fixed-odds) bookmaker. Where bookmakers are legal, they are generally legal on-course. They are also legal off-course in the United Kingdom and Ireland, while being illegal off-course in Australia. In New Zealand there are no on-course bookmakers, but the New Zealand tote offers a bookmaking service on selected races. The crucial difference between bookmakers and the betting mechanisms described above is that bookmakers offer fixed odds on race outcomes. In other words, bettors know as soon as they have bet how much they stand to win if their chosen animal wins.

This provides an obvious advantage for the possessor of inside information because it enables the holder to make more accurate calculations regarding the expected return when large sums of money are bet. With an SP-type mechanism, this is impossible, whereas with the tote it is plausible only if the insider waits until seconds before the end of betting before placing the bet[23] and no other insider has the same idea! It is thus broadly accepted in the literature that insiders bet with on-course bookmakers wherever possible. And if there are no legal bookmakers on-course or readily accessible from the track by phone or online? This question has only rarely been addressed in the literature.[24] However, before presenting some new empirical results on this issue, it will be useful to consider the literature on insider trading in the presence of bookmakers.

3 Insider Trading in the Presence of Bookmakers

The pioneering paper in this field is by Jack Dowie (1976), who presented an interesting, albeit flawed, test of whether there is profitable insider trading in the U.K. horse betting market. Dowie (1976, 147) defined SP "as the odds at which a sizeable bet could have been placed at the 'off'," in other words as very close to the end of betting, from which he argued that

> SP can be taken to incorporate any superior or inside information that exists in relation to the event. If inside information plays a significant role in horse race betting markets, then the correlation between the probabilities embodied in the SP returns and the realized probabilities should be significantly higher than the correlation between the latter and any other set of probabilities assigned prior to the "off" (and certainly any set assigned prior to betting on the event). If, then, the correlation between the probabilities embodied in the betting forecasts in a morning newspaper and the realized probabilities is as high as the SP correlation, we can conclude that the existence of superior "inside" information is in doubt (149).

Having shown that the correlation between the winning probabilities incorporated in forecast prices and ex post winning probabilities are no less than those between the latter and the winning probabilities implied by SP, Dowie concluded that the market is strongly efficient.[25] The problem with the argument is that these correlations being tested refer to all horses in Dowie's sample and not simply to those horses on which insiders bet. Thus SP might be highly accurate for insider-traded horses and quite inaccurate for all other horses, whereas forecast odds might be reasonably accurate overall.

This chapter will provide a simple counter-test to Dowie's after the behavior of insiders in such a market are described. Since bookmakers are not required to take bets of unlimited size, an insider wishing to bet large sums on a horse will spread the bet by placing acceptable quantities of money on the horse simultaneously with a number of different bookmakers so as to obtain the best odds possible. This highly visible activity is called a plunge and is accompanied by an immediate reduction in the fixed odds of the plunged horse offered by the bookmakers. Since insiders bet at greater odds than SP, a better test of the presence of insiders in the U.K. betting market would be to check whether horses whose odds shorten between forecast price and SP would yield profits were they to have been backed at forecast prices.[26] This is the essence of N. F. R. Crafts's (1985) attack on Dowie, which he supplemented with interesting anecdotal evidence of heavily plunged winners. Crafts also showed that horses that drift from forecast prices to SP tend to belong to "a class of outstandingly poor value bets" (303).

Crafts used these results to suggest that horse betting markets may require more government regulation than was the case. After all, the sport's own governing body, the Jockey Club, was apparently of the same view: "The Jockey Club's own Committee of Inquiry argued that 'the public is entitled to be satisfied that every precaution is taken to ensure that racing is fairly conducted and that malpractices are reduced to a minimum"' (Crafts 1985, 303). Adi Schnytzer and Yuval Shilony (2007) set out to test the implicit hypothesis provided by Crafts that insiders might have an incentive to mislead the betting public and bookmakers by having their horses deliberately underperform. This would provide the insiders with better odds at a later start.

Schnytzer and Shilony presented a simple model that shows the conditions under which it is optimal for insiders to rig prices by deliberate underperformance in some races. Using tote[27] data for Australian greyhound, harness, and thoroughbred racing markets, they then showed how an empirical analysis of the relationship between win and place probabilities in conjunction with observed patterns of betting behavior can be used to establish the presence of price rigging. It was shown that there is no significant systematic price rigging in these markets. Quite simply, animals that are "not on the job" do not, on average, underperform other animals. This does not reject the hypothesis that there is some corruption in these markets, but it does reject the hypothesis that there is *systematic* corruption.

Proceeding chronologically from Craft's contribution in this area, three papers by Hyun Song Shin[28] represent pioneering work on the theory of bookmakers' behavior in the presence of insiders. Shin (1991) presented an extensive form game in which nature first chooses the winning probabilities of the horses in a two-horse race.[29] This distribution is observed by a monopoly bookmaker, who sets odds according to this probability distribution. The single insider then observes the actual result of the race and decides how much to bet subject to a budget constraint. Outsiders are effectively noise traders and provide the bookmaker with the means to pay the winning insider while still allowing the bookmaker a profit. Shin proved that in the unique equilibrium in this game, the optimal prices set by the bookmaker will display a favorite-longshot bias[30] and that the value of insider trading is a decreasing function of the size of the insider's (assumed known to the bookmaker) budget, since as this budget grows, the implied optimal prices for the bookmaker rise. The maximum value of insider trading is shown to be one-third, a modest proportion given that the insider always wins, but is explained by monopoly power of the bookmaker.

Shin (1992) presented a generalization of the previous model to the case of a horse race with more than two starters and a single bookmaker who, however, has lost his monopoly power because he must bid against another candidate to set odds on the race. The game analyzed here has four stages. In the first stage two potential bookmakers bid the sum of prices[31] in the race and the lowest bidder becomes the bookmaker in the race. In the second stage the bookmaker sets prices for each horse in the race, subject to the constraints that no price is negative or greater than one and that the sum of prices is no greater than that bid in the first stage. In the third stage, nature decides on the winner of the race and whether the single bettor in the race will be an outsider or an insider. The difference between insider and outsider is that the former knows the identity of the winning horse while the latter believes with certainty that a particular horse (not necessarily, but possibly, the actual winner) will win the race. In the final stage the chosen punter bets and the payout determined. The equilibrium in this game when the bookmaker is subject to a zero profit condition is that, as in Shin (1991), prices will reflect a favorite-longshot bias.

In order to produce a method of measuring the extent of inside traders operating at the track, Shin (1993) modified the previous model slightly and derived an iterative procedure for determining the extent of insider trading, obtaining a value of somewhat over 2 percent. He again modeled the behavior of competitive bookmakers facing two types of bettors, those with and those without inside information, and maximizing expected profit. The main result is the existence of a favorite-longshot bias if and only if there is inside money in the market. The model hinges on the absence of any bias in the participants' forecasts of race results. Shin assumed that the proportion of outsiders backing any given horse is equal to its true winning probability and that insiders know the winning horse and always back it. Thus, ex post, the proportion of money bet on all horses by all bettors is exactly in accordance with the true winning probabilities. Observing the optimal prices charged by bookies in a race, the model can be used to solve the underlying winning probabilities and the incidence of insider trading. This solution cannot be obtained analytically, but numerically, using an iterative process as described by Michael Cain, David Law, and David Peel (1996).

An interesting result obtained by Shin is that there is a linear relationship between the sum of prices and the number of horses in a race, and he argued that this is due to the presence of inside traders. Leighton Vaughan Williams and David Paton (1997) showed that the relationship may be due also to bettors counting only a fixed fraction of their losses. They also showed that the favorite-longshot bias may derive from the demand as well as the supply side. In other words, while bookmakers may bias their prices in the presence of insiders, there are other factors explaining the bias.[32]

In a simple test of Shin's model against Craft's idea of the plunge as an indicator of insider trading, Cain, Law, and Peel (2001b) showed that there is a significant relationship between Shin's measure of the extent of insider trading and the extent of plunging in the race. In an alternative test of Shin's model, Schnytzer and Shilony (2003) tested whether the Shin (1993) model-derived probabilities correspond to actual winning frequencies. They showed that bettors display a favorite-longshot and that accordingly bookmakers would evaluate their expected profits, taking this into account and setting prices accordingly. However, Shin did not take this correction into account in his model, and this undermines the essential result of Shin's model, which is that the *sole* cause of the bias is insider trading.

Cain, Law, and Peel (2001a) used Shin model estimates of insider trading in the betting on individual races to show that the market anomaly observed by Paul Gabriel and James Marsden (1990), that U.K. tote payments on winning bets consistently exceeded those paid by bookmakers, is in fact more subtle than originally reported. Use of more appropriate statistical methods suggests that bookmakers pay more generously than the tote on winning bets on favorites but less generously on winning longshot bets. They showed that this new anomaly is associated with the incidence of insider trading in the betting on each race and argued that it cannot be arbitraged away owing to the bookmakers' dominant market position.

In a different application of the Shin model, Law and Peel (2002) tried to distinguish between insider trading and herding (as evidenced by plunged horses) by arguing that in races where the Shin measure of the extent of insider trading increases as between opening prices and starting prices, then insider trading may be imputed, whereas if there is a reduction in the Shin measure and the horse has been plunged, the plunges would appear to indicate herding. Law and Peel (2002, 327) found "that significant positive betting returns are achieved when shortening odds are accompanied by a rise in the Shin measure; when they are accompanied by a fall, returns are negative, suggesting herd behaviour."

One difficulty with the Law and Peel approach is the implicit assumption that the reduction in a horse's odds is due either to insider trading or to herding. While the authors do provide a brief discussion of various possible triggers for herding, there is no theoretical model and no empirical allowance for the possibility that a horse might be plunged by insiders early in the betting and subsequently drift in the betting, allowing profits to be made by betting late. One reason for this omission might be that Law and Peel had access only to opening and starting odds and not to odds in the middle of betting. This lacuna was filled Schnytzer and Avichai Snir (2008a), who showed that profits may be made by betting on horses plunged early in the betting, which then drift, even at starting prices in both the British and Australian on-course bookmakers' markets. Based on the results obtained, it would seem that, on average, horses which are plunged late generally trigger herding and that only some of those plunged early are followed by herding behavior.

The mention of early and late plunges raises the question: when do insiders bet in a bookmakers' market? This issue has been analyzed by Schnytzer and Shilony (2002), who modeled a game of timing where the agents are insiders wishing to bet on different horses in a race. The basic arguments are as follows. At the opening of the betting, odds are determined by a cartel and incorporate profit margins so that odds are lower than the equivalent perceived winning probabilities for all horses. As soon as betting begins, each bookmaker is on his own, and the cartel collapses.[33] Thus, *ceteris paribus*, odds tend to rise over time. However, there are risk-neutral informed traders whose estimates of winning probabilities for the horses with which they are associated are more accurate than those of the bookmakers. Should the odds be greater than the insiders' valuation of the corresponding horse's winning probability, the relevant insiders will place large amounts of money on the horse via a plunge.[34] This leads to an immediate reduction in the odds of that horse and an increase in the odds of all other horses in the race. Suppose, now, that there are two such groups of insiders, each wishing to plunge its own horse. Since a plunge increases the odds for other horses, each group has an incentive to wait for the other to plunge first. On the other hand, since the information concerning any given horse is known to more than one person,[35] the longer the insiders wait, the greater the risk that the information will leak to a third party. The recipient of the leak will then plunge the horse and the group of insiders—except perhaps the one responsible for the leak—may be left with odds at which betting is no longer worthwhile. This conflicting incentive gives rise to the game of timing. Schnytzer and Shilony showed that the game has no equilibrium in pure strategies but, for the two-player case,[36] derived the unique equilibrium in mixed strategies. Based on this model, three rather intuitive empirically testable hypotheses were derived:

1. The lower the level of opening odds, the later any plunge activity will take place.
2. An increase in the number of horses that have insiders associated with them leads to an earlier optimal plunge time.
3. An increase in the number of horses also leads to an earlier optimal plunge time.

Using data from the 1997–1998 horse racing season in Australia, the authors showed that these hypotheses were strongly supported by the data.

Three additional studies complete this survey of the literature on insider trading in the presence of on-course bookmakers. The first is Schnytzer and Shilony (1995), which found that insider trading takes place in this type of market. A second key study is an extension of this line of research by Schnytzer, Shilony, and Richard Thorne (2003), and the third is based on an alternative measure of the extent of insider trading provided by Schnytzer, Martien Lamers, and Vasiliki Makropoulou (2008 and 2009). The Schnytzer and Shilony (1995) paper built on Craft's 1985 insight that plunges provide evidence of insider trading and used the unique market environment pertaining to horse race betting in Melbourne in the 1980s to make the point without access to bookmakers' odds data and without employing unintuitive theoretical modeling. The argument of the paper may be summarized as follows:[37]

The Melbourne horse betting market produces[38] data on the betting behavior of two mutually isolated populations who bet with the tote, one with and the other without access to inside information. In this market, off-course bettors may place bets with the government-run tote at offices located throughout Victoria. During the 1980s the Victoria government had a legal monopoly in off-course betting.[39] In addition to the public information already available to bettors, all off-course tote offices provided the projected odds for the different horses in a race from around 30 minutes before starting time. There are also tote windows at the racecourse. Some 15 minutes before race time, the off-course pool is amalgamated with the on-course pool so that the tote payout on a race is based on the combined pools. Since the size of the off-course pool is generally more than three times that of the on-course pool, unless the projected odds in the two pools are very different, changes in projected odds owing to differences in the betting behavior of the two populations were not easily picked up by off-course bettors following the amalgamation.

On-course bettors, of course, also may bet with bookmakers. Well over one hundred private bookmakers competed[40] among themselves and with the tote. The important feature of this market, which facilitated the monitoring of inside information, was that, with the exception of tote information—whose presentation to bettors was identical on- and off-course—information flows between these two segments of the market were restricted. There were no public telephones at the track, and the communication of race-related information via official phones was forbidden. Bettors were free to leave the track and pay to return, but since the on- and off-course tote payouts are identical, they had no (legal) incentive to leave the track. Given that with bookmakers the payout contingent on a win is known when the bet is placed and that inside information is likely to be more accurate as race time draws near, it is to be expected that most "insiders" would bet with bookmakers via plunges. A plunge leads to an immediate reduction in the odds offered by bookmakers about the horse in question. Since bookmakers operate in very close proximity to one another, bettors have no difficulty in discerning a plunge. A plunge provides bettors with an indication that a horse's connections believe that the true probability the horse will win is higher than that reflected in the pre-plunge available odds. Bettors' budget considerations and herding aside, the plunge continues to a point at which the odds have shortened to reflect the new subjective probability that the horse will win. The projected tote odds would now have been longer than those offered by bookmakers. "Outsiders," who observed the plunge, would bet on the plunged horse with the on-course tote to take advantage of this "expected arbitrage opportunity." Hence the informed population, on-course tote bettors, used "secondhand" inside information. However, since plunge information does not leave the race course, it is not incorporated into the behavior of off-course bettors.

Under these circumstances, testing the hypothesis that inside traders are betting on-course is fairly simple. First it needs to be shown that the information contained in the proportions of money bet both on-course and off-course significantly contributes to an explanation of the horses' ex post winning probabilities. Schnytzer and Shilony (1995) did this by regressing a win dummy on the proportions bet on-course and off-course using conditional logit regression[41] and found that both explanatory variables received positive and significant coefficients. They remained positive and significant when each was run alone in a separate regression. The interpretation of these variables is that the proportion of money bet off-course represents the outsiders' assessment of public information, while the relative amounts bet on-course represent public information as modified by inside information.

Second, and even more convincing, is the following betting simulation: Suppose a bettor had access to these proportions just before the race and bet on any horse for which the proportion of money bet on-course exceeded the proportion bet off-course. This would have yielded a profit, net of the 15 percent tote take, of 29.8 percent over 168 races during the 1984 season. This result makes it clear that, prior to the cellular phone era, insiders, bookmakers, the tote, and some on-course outsiders were all profiting at the expense of off-course bettors in this betting market.

The connection between on-course bookmakers' odds and the transmission of plunge information to the tote is considered in greater detail in Schnytzer, Shilony, and Thorne (2003). It showed that when the Victoria tote offers bets on races being held both in Victoria and other states, the transmission of odds information from the on-course bookmakers to the tote is more efficient for local races, since on-course punters betting on interstate races are not aware of the precise timing and extent of all plunges on those races, as local bookmakers fielding interstate races receive only sporadic information from the relevant interstate betting market.

Prior to a discussion of insider trading in tote-only betting markets, it remains to present preliminary results on ongoing research that provides an alternative approach to Shin (1993) to measuring the extent of insider trading at the track. This is the work of Schnytzer, Lamers, and Makropoulou (2008 and 2009).[42] The model is an extension of that developed by Makropoulou and Markellos (2011), which was applied to the European soccer betting market. The basic intuition underlying the model is that fixed odds[43] offered by bookmakers at the track are examples of call options and that, while bookmakers hope to offer only net of premium out-of-the-money options, when they err by underestimating a particular horse's true winning probability they are liable to offer a net in-the-money option, which the insider (who is assumed to know a horse's true winning probability) will be glad to snap up. Building on Schnytzer and Shilony (1995 and 2002), the model tracks the value of the options in the face of plunges and, using Monte Carlo simulations, estimates the extent of insider trading as between 20 and 22 percent, a value considerably greater than Shin's (1993) and others applying his model of 2.5 or so percent. The major reason that Schnytzer, Lamers, and Makropoulou (2008 and 2009) obtained so high a value is that they did not focus on SP alone and they relaxed Shin's two particularly restrictive assumptions, namely, that the proportion of outsiders backing any horse is equal to its true winning probability and that inside traders always win.

4 Insider Trading in a Tote-Only Market

It is perhaps surprising that very little has been written about the operation of inside traders in tote-only markets, but the reason is simple: the returns from tote betting are a strictly decreasing function of the take and the quantity bet on the winning animal while they are a monotonically increasing function of the amount bet on all losing animals. Thus the more the insider bets on an animal, *ceteris paribus*, the lower the return (whether the animal wins or loses). Further, if the insider plunges an animal relatively early in the betting and the public spots the plunge,[44] the return may be pushed down further by herding. Hence simple logic suggests that insiders will wait until almost the end of the betting when returns may be calculated with greater accuracy and the impact of potential herding minimized. An alternative strategy might be to bet very small amounts continuously so that no real plunge becomes evident, but this is only likely to work if the projected payout is sufficiently high to provide a high contingent profit for a relatively small outlay.

One paper that has tested the hypothesis that "smart" money bets late in a tote-only market is that produced by Peter Asch, Burton Malkiel, and Richard Quandt (1982). They argued as follows:

> We have suggested above that because of the potential signaling effect, bettors who feel they have inside information would prefer to bet late in the period so as to minimize the time that the signal was available to the general public. As table 3 [p. 193] shows, the marginal odds of the late bettors appear to be at least as good as and perhaps better than the final odds in predicting the order of finish. Horses that win have marginal odds that average 79 to 82% of their morning line odds (depending on which definition of marginal odds [p. 190][45] is employed). In other words, winning horses are especially favored by the late bettors.

This paper concludes with some preliminary evidence on the presence of insiders in the horse betting market in Hong Kong. The data consist of 4,245 Hong Kong races in which 54,335 horses took part between the third of September 2000 and the eighteenth of October 2006. For each horse[46] prospective payouts at three different time slices were observed: at overnight, 5 minutes before race start, and at the close of betting. The actual win payout and the horse's finishing place in the race also are known. The data are provided for both tracks in Hong Kong: Happy Valley, and Sha Tin.

Using the same data set, Schnytzer and Barbara Luppi (2008) have shown that this market has two features that are of potential relevance for the tracking of inside traders. First there was no favorite-longshot bias in the betting at any of the available stages of betting. This is perhaps surprising since those people betting on the day prior to the race are most likely "normal" outsiders, and it is these bettors who tend to bet with a bias in almost all other animal betting markets hitherto studied. The relevance of this finding is that such insiders as might be betting in Hong Kong are unlikely to be arbitraging away biases and thus are more likely to be focusing on specific mispriced horses. Indeed, this is likely to be the case also for syndicate bettors, and the model hinted at by Bill Benter (1994) certainly seems to have been designed with just such a purpose in mind.

Second, using a linear probability model to regress a win dummy on the price equivalents of the three sets of prospective payouts and horse-jockey interactions, it was shown that all three sets of prices contributed significantly to an explanation of horses' winning probabilities at a one percent level of statistical significance. In other words, odds are changing during the betting but always with the addition of valuable information. Now, suppose that the (overnight) opening prospective payouts represent a rough, albeit unbiased, estimate of public information; in that case the subsequent odds changes seem, at least in part, to reflect either insider trading or syndicate expert betting or both. For the purposes of this chapter, it is not possible to attempt a formal distinction between the two,[47] but the presence of winners in this market is readily demonstrated. Informally, the one thing that points in the direction of insider trading is that winning plunges (as shown in table 17.1) are all in the range of middle to long shots, and to the extent that expert syndicate betting is based on modeling public information, it is to be expected that some favorite categories would yield profits. Put differently, public information in the absence of insiders is unlikely to consistently misprice winning horses at odds greater than 10 to 1.

The results shown in table 17.1 are based on the following simple natural experiment: Suppose that a better is able to bet when all others bettors have finished betting but the race has not yet begun. This bettor places \$1 on every horse that has been plunged in different payout categories and also \$1 on each favorite[48] and each long shot[49] in the race. When there are two or more equal favorites or long shots in the race, \$1 is put on each. Table 17.1 shows the returns to such betting for favourites, long shots and every final payout category in which a profit resulted. For all other payout groups losses were incurred and not shown.

Table 17.1 presents results for both tracks together and separately and shows the number of races in the sample, the number of bets placed, the payout group where 11–20, for example, means that the winning horse paid somewhere between \$11 to \$20 inclusive for a dollar bet on the Hong Kong tote, the type of plunge, and the actual percentage return. There are three types of plunge: early plunged is defined as a reduction in the prospective tote payout between the overnight and 5 minutes to start payouts, a late plunge takes place when the prospective payout falls during the last 5 minutes of betting, and throughout indicates that there has been both an early and late plunge on the horse.

The first thing to notice about the results is that backing favorites in Hong Kong is not a bright idea. While doing so loses considerably less at Sha Tin than betting on rank outsiders, it fares considerably worse when betting at Happy Valley. This discrepancy between the tracks may be explained by the relatively high number of bets on rank outsiders at Sha Tin compared with Happy Valley and by the fact that Happy Valley is in general a far more difficult track for bettors than is Sha Tin.[50]

Table 17.1 Returns to Tote Betting on Plunged Horses in Hong Kong, 2000–2006

Track	Number of Races	Number of Bets	Payout Group	Plunge Type	Return (%)
Both	4,245	4,312	Favorite	-	−0.1934369
Both	4,245	5,934	Long shot	-	−0.4496124
Both	4,245	1,870	21–30	Late	0.1508021
Both	4,245	970	31–40	Late	0.0412371
Both	4,245	1,524	11–20	Throughout	0.003937
Both	4,245	399	21–30	Throughout	0.0100251
Happy Valley	1,428	1.457	Favorite	-	−0.2046671
Happy Valley	1,428	1.593	Long shot	-	−0.086629
Happy Valley	1,428	1.395	11–20	Late	0.0415771
Happy Valley	1,428	423	11–20	Throughout	0.0661939
Happy Valley	1,428	25	31–40	Throughout	0.36
Sha Tin	2817	2,855	Favorite	-	−0.1877058
Sha Tin	2,817	4,341	Long shot	-	−0.582815
Sha Tin	2,817	407	51-max	Early	0.031941
Sha Tin	2,817	1,342	21–30	Late	0.2406855
Sha Tin	2,817	738	31–40	Late	0.1287263
Sha Tin	2,817	315	21–30	Throughout	0.1365079

The results in table 17.1 suggest that insiders are betting on horses in the range 10 to 1 and up. It is also clear that most winning plunged horses are plunged either late only or throughout, the only exception being the 3 percent profit on early plunged longshots at Sha Tin. Finally, the returns to betting plunged horses in the winning categories appear to be both greater and more even at Sha Tin than at Happy Valley, the 36 percent profit in the 31–40 range at the latter track being for only 25 bets.

In conclusion, it is clear that more research needs to done on the behavior and impact of insider traders in tote-only markets. While both Asch, Malkiel, and Quandt (1982) and this chapter provide evidence that insider trades occur late in the betting, the evidence from Hong Kong is hardly unambiguous, with money being made also on horses plunged early and throughout the betting. However, a puzzle remains to be solved: how is it that all profitable plunged odds groups in Hong Kong offered odds at 10 to 1 or greater?

Notes

1. Excluded therefore are any betting markets that are illegal in the jurisdiction(s) in which they operate. Thus the Australian horse racing markets operated by various pari-mutuels in that country will be analyzed even though participation in these markets is illegal for Americans. SP bookmakers who operate illegally in some Australian pubs are, however, not considered.
2. It makes no difference if it is an American or a European option.

3. It is assumed that either these flies are untrainable or that even if flies can be trained to race the flies in the market under consideration have not been trained and are unknown to the bettors in the market.
4. In particular, if the options are on shares and share returns move in accordance with Brownian motion (as is argued in some texts, e.g., Wilmott 1998), then attempted predictions of future prices would be largely useless. On the other hand, supporters of technical analysis believe that, under appropriate circumstances, share returns are predictable (see Bollinger 2002).
5. Accurate prediction of actual share prices over different time horizons is rendered impossible by the fact that exogenous shocks of sufficient magnitude influence the prices of all shares to a greater or lesser degree.
6. There is a considerable literature dealing with the impact of legal insider trades in the U.S. markets, where the insider has reported the trade with the SEC, but these are of no interest here since such legal trades cannot, by definition, make use of information unknown to the public. For an example of a paper that deals with a priori illegal insiders in this market see Coleman and Schnytzer (2008).
7. For example, an insider who believes that the price of the relevant shares will fall may sell a range of calls at different strike prices and expiry dates and may also buy various puts. This is the equivalent of plunging a horse by betting small amounts with many bookmakers.
8. Even identical new apartments on the same floor of the same building differ in regards to position and thus, for example, views.
9. Research into moral hazard in real estate brokerage is unrelated to this topic. For an example of this literature see Munneke and Yavas (2001).
10. Short of such surreal possibilities as drugging flies.
11. Organized camel racing is popular in the Middle East, but gambling is illegal. Ostrich racing is popular in Africa and in parts of the United States, and pari-mutuel betting is legal at the annual Virginia City, Nevada, Camel Races festival (http://j-walkblog.com/index.php?/weblog/comments/20295), but a discussion of these markets is beyond the scope of this chapter.
12. For an insightful analysis of the distinction between insiders and experts (an issue discussed very briefly in section 4), see Pierson (2011).
13. For example, betting on a horse to win a race, betting on it to occupy either first, second, or third place in the race, betting on which two horses will finish first and second in the race either in the correct order or not are but some of the many betting options. See http://www.dannysheridan.com/horse-racing/horses-101.php and http://www.freebettingonline.co.uk/Horse-Racing-Betting/ for information regarding betting in the U.S. and U.K. markets, respectively.
14. This involves no loss of generality since whether insiders bet and the possibility of tracking their behaviors are determined by mechanism and location and not by bet type.
15. A trivial example of the importance of being on-course will suffice: when the animal is brought to the track it may be involved in a minor traffic accident or even simply suffer a shock from strange traffic noises which would reduce its winning chances in the race. An insider who is at the track with the animal—and better yet travels with it to the track—is best placed to monitor the animal's fitness for racing. One the other hand, in the era of mobile phones, it may be possible for the insider to remain off-course and still be fully informed.

16. Note that most totes round their payouts down to the nearest 5 or 10 cents so that if the correct payout as described in the text should be, say, \$4.29, then either \$4.25 or \$4.20 would be the actual payout.
17. After rounding, the payout in the example would be 90 cents instead of 94 cents.
18. Subject to fluctuations in profit from rounding.
19. If the tote is privately owned, its take may be fixed by the government, though state-owned totes might set the take to maximize returns. For more on this topic see Gruen (1976) and Suits (1979).
20. Dubai features the horse race with the world's largest prize, but there are no legal gambling anywhere facilities in Dubai. For details and betting facilities online see http://www.dubai-horse-racing.com/dubai-world-cup-odds.html, where British bookmakers sell bets.
21. A service that pays full on-course tote odds is available via the Internet in some states, though winnings may be subject to taxation. See http://www.tvg.com/Default.aspx for more details.
22. See Schnytzer and Snir (2008b). The authors provide empirical evidence based on U.K. data that SP betting is a self-enforcing cartel. Outside the United Kingdom there are Internet bookmakers who offer SP plus betting; that is, they offer odds greater than SP. As one Australian online bookmaker put it: "If you are currently placing your bets at SP you really should open an account with Centrebet. They offer SP + odds on all UK horse racing and greyhounds." http://www.racing-index.com/bookmakers/centrebet.html.
23. Unless the tote in question does not provide up-to-date data on the amounts bet on the race, in which case the tote is similar to the SP-type mechanism.
24. See Asch, Malkiel, and Quant (1982), which is discussed in section 4.
25. That is, there is no meaningful insider trading in this market.
26. Of course no one bets at forecast prices, but they may be taken as a reasonable proxy for bookmakers' opening prices.
27. Given the presence of both tote and bookmakers on-course in Australia, it has been shown that plunges with the bookmakers are reflected in tote odds. See Schnytzer and Shilony (1995), which is discussed below.
28. Shin (1991, 1992 and 1993).
29. For an alternative model of bookmakers faced with insiders and the resulting favorite-longshot bias, see Schnytzer and Shilony (2005).
30. This well-known bias is discussed elsewhere in this volume.
31. The price is the reciprocal of one plus the odds. Thus for odds of 3 to 1, the price is 0.25. The price represents the outlay necessary to ensure a total payout of 1 should the horse win the race.
32. See elsewhere in this volume.
33. This is a reasonable picture of bookmaker odds-setting in the Australian market where off-course bookmaking was illegal in 1995.
34. See Schnytzer and Shilony (1995), which is discussed below.
35. Owner(s), trainer, jockey or driver, stable hands, perhaps family members and friends, etc.
36. It should be noted that the number of plunges per race for most Australian datasets is around 2 to 3.
37. The following paragraphs paraphrase closely pages 963–964 of Schnytzer and Shilony (1995).

38. The use of the present tense here should be understood as applying to the era prior to the introduction of cellular phones in Australia. In the description that follows, the past tense is used wherever the microstructure of this market has changed since the 1995.
39. The Victoria tote has subsequently been privatized.
40. The changing structure of this market has led to a significant reduction in the number of bookmakers who field at Melbourne tracks. A discussion of this interesting subject is beyond the scope of this chapter.
41. McFadden (1973).
42. For the most recent theoretical model in this area see the chapter by Schnytzer, Makropoulou, and Lamers elsewhere in this volume.
43. For the purposes of this chapter, "odds" means that odds of, say, 5 to 1 represent a net profit of $5 for every $1 bet on the winning horse.
44. Tote projected payout updates are today very frequent at all tracks.
45. Asch, Malkiel, and Quandt defined marginal odds as the odds implied by those bettors betting during a particular part of the betting period. They defined the marginal odds for those betting during the 8 and 5 minutes of betting, respectively, and showed that the latter bet relatively more on winners than the former, who themselves bet relatively more on winners than all bettors.
46. Note that the number of starters in a race runs from 7 through 14, with most races having 12 or 14 starters.
47. If indeed this is ever possible, since both would appear to be net winners and because an empirical distinction between a bettor who has a better estimate than outsiders of a horse's winning probability because he knows the horse and one who has an edge because he is a superior data processor would seem to be difficult to draw, even if the conceptual differences are self-evident. In the case of the bookmakers' market in Australia (such as that studied in Schnytzer and Shilony 1995), bettors who make up the plunging agents are invariably observed running from the stables area, suggesting that they are insiders. Plunging on the tote is observable only via observable prospective payout changes.
48. That is, the shortest priced horse(s) in the race at the close of betting.
49. The rank outsider(s) in the field at the close of betting.
50. This issue is discussed in Schnytzer and Luppi (2008).

References

Asch, Peter, Burton G. Malkiel, and Richard E. Quandt. 1982. Racetrack betting and informed behavior. *Journal of Financial Economics* 10(2):187–194.

Benter, William. 1994. Computer based horse race handicapping and wagering systems: A report. In *Efficiency of racetrack betting markets*, edited by Donald B. Hausch, Victor S.Y. Lo, and William T. Ziemba. San Diego, Calif.: Academic Press, 183–198.

Bollinger, John A. 2002. *Bollinger on Bollinger Bands*. New York: McGraw Hill.

Cain, Michael, David Law, and David A. Peel. 1996. Insider trading in the greyhound betting market. Salford Papers in Gambling Studies No. 96–01. Lancashire: Center for the Study of Gambling and Commercial Gaming, University of Salford.

———. 2001a. The incidence of insider trading in betting markets and the Gabriel and Marsden anomaly. *Manchester School* 69(2):197–207.

——. 2001b. The relationship between two indicators of insider trading in British racetrack betting. *Economica* 68(269)97–104.

Coleman, Les, and Adi Schnytzer. 2008. Shorting the bear: A test of anecdotal evidence of insider trading in early stages of the sub-prime market crisis. *Journal of Prediction Markets* 2(2):61–70.

Crafts, N. F. R. (1985). Some evidence of insider knowledge in horse race betting in Britain. *Economica* 52(207):295–304.

Dowie, Jack. 1976. On the efficiency and equity of betting markets. *Economica* 43(170):139–150.

Gabriel, Paul E., and James R. Marsden 1990. An examination of market efficiency in British racetrack betting. *Journal of Political Economy* 98(4) :874–885.

Gruen, Arthur. 1976. An inquiry into the economics of race-track gambling. *Journal of Political Economy* 84(1):169–178.

Law, David, and David A. Peel. 2002. Insider trading, herding behaviour and market plungers in the British horse-race betting market. *Economica* 69(274):327–338.

McFadden, Daniel. 1973. Conditional logit analysis of qualitative choice behaviour. In *Frontiers in Econometrics*, edited by Paul Zarembka. New York: Academic Press, 105–142.

Makropoulou, Vasiliki, and Raphael N. Markellos. 2011. Optimal price setting in fixed-odds betting markets under information uncertainty. *Scottish Journal of Political Economy* 58(4):519–536.

Munneke, Henry J., and Abdullah Yavas. 2001. Incentives and performance in real estate brokerage. *Journal of Real Estate Finance & Economics* 22(1):5–21.

Schnytzer, Adi, and Barbara Luppi. 2008. Painful regret and elation at the track. *Journal of Gambling Business and Economics* 2(3):85–99.

Schnytzer, Adi, and Yuval Shilony. 1995. Inside-information in a betting market. *Economic Journal* 105(431):963–971.

——. 2002. On the timing of inside trades in a betting market. *European Journal of Finance* 8(2):176–186.

——. 2003. Is the presence of insider trading necessary to give rise to a favorite-longshot bias? In *The economics of gambling*, edited by Leighton Vaughan Williams. London: Routledge, 14–17.

——. 2005. Insider trading and bias in a market for state-contingent claims. In *Information efficiency in financial and betting markets*, edited by Leighton Vaughan Williams. Cambridge: Cambridge University Press, 287–312.

——. 2007. The optimality and statistical detection of price rigging in betting markets. *Journal of Gambling Business and Economics* 1(1):13–29.

Schnytzer, Adi, and Avichai Snir. (2008a). Herding in imperfect markets with inside traders. *Journal of Gambling Business and Economics* 2(2):1–16.

——. (2008b). SP betting as a self-enforcing implicit cartel. *Journal of Gambling Business and Economics* 2(1):45–65.

Schnytzer, Adi, Martien Lamers, and Vasiliki Makropoulou. 2008. *Measuring the extent of inside trading in horse betting markets*. Working Paper No. 2009-10. Bar-Ilan, Israel: Department of Economics, Bar-Ilan University; http://econ.biu.ac.il/files/economics/working-papers/2009-10.pdf.

——. 2009. The impact of insider trading on forecasting in a bookmakers' horse betting market. *International Journal of Forecasting* 26(3):537–542.

Schnytzer, Adi, Yuval Shilony, and Richard Thorne. 2003. On the marginal impact of information and arbitrage. In *Economics of gambling*, edited by Leighton Vaughan Williams. London: Routledge, 80–94.
Shin, Hyun Song. 1991. Optimal betting odds against insider traders. *Economic Journal* 101(408):1179–1185.
———. 1992. Prices of state contingent claims with insider traders, and the favourite-longshot bias. *Economic Journal* 102(411):426–35.
———. 1993. Measuring the incidence of insider trading in a market for state-contingent claims. *Economic Journal* 103(420):1141–1153.
Suits, Daniel B. 1979. The elasticity of demand for gambling. *Quarterly Journal of Economics* 93(1):155–162.
Vaughan Williams, Leighton, and David Paton. 1997. Why is there a favourite-longshot bias in British racetrack betting markets? *Economic Journal* 107(440):150–158.
Wilmott, Paul. 1998. *Derivatives: The theory and practice of financial engineering*. Hoboken, N.J.: Wiley.

CHAPTER 18

PRICING DECISIONS AND INSIDER TRADING IN FIXED-ODDS HORSE BETTING MARKETS

ADI SCHNYTZER, VASILIKI MAKROPOULOU, AND MARTIEN LAMERS

1 Introduction

This chapter examines the decision a bookmaker makes when setting prices under uncertainty in a fixed-odds betting market.[1] We argue that this decision can be modeled in a call option framework to measure the degree of insider trading in racetrack betting markets. Vasiliki Makropoulou and Raphael Markellos (2011) first developed an option-pricing framework for the pricing of bets in fixed-odds markets and in particular for the European soccer betting market. In this market the odds are offered by bookmakers via fixed-odds coupons several days before the game, and they remain largely unchanged throughout the betting period. Their model deals with expert traders who either exploit public information in a manner superior to that of bookmakers or obtain access to new public information sooner than bookmakers do. Our approach differs in that we focus on racetrack betting, where odds change frequently during the half-hour betting period. In our context, public information is irrelevant since it can be incorporated into new odds as soon as it hits the market. On the contrary, we deal with insiders who possess private information. Of course, the implications of trading with insiders in the racetrack betting market where the bookmaker frequently changes the odds can be quite similar to those of trading with experts in a market where the odds remain unchanged. However, the Makropoulou–Markellos framework could not be readily applied to the racetrack betting market, due to structural differences between the two markets. In order to fill in this gap, Adi Schnytzer, Martien Lamers and Makropoulou (2010) developed a model for the pricing of bets in a market with insiders relying on the Makropoulou–Markellos framework and applied it to the Australian racetrack betting market. In this chapter, we extend the research of Schnytzer, Lamers, and Makropoulou (2010) in several aspects. First, we relax their assumption of continuous arrival of information by employing a more realistic specification in which information arrives in discrete amounts, and therefore the true probability of a horse winning exhibits quantum jumps and dives. Second, the model is extended to allow for more periods in which betting takes place. Also, whereas Schnytzer, Lamers, and Makropoulou (2010) assumed a betting period in which the bookmaker sets prices once (at opening prices), we extend the model to accommodate more time periods in which a bookmaker quotes prices. More specifically, we follow the data at our disposal and allow for betting by insiders both at opening prices and middle prices instead of only at one of those. Finally, to derive the probability of insider trading, the zero-profit condition of the bookmaker does not have to hold for every single horse. This condition is necessary only for the race, allowing the bookmaker to make losses only on the horses he or she expects insiders to bet on and to make profits on the horses backed by outsiders, who bet according to subjective winning probabilities in accordance with public information as explained below.

The remainder of this chapter is organized as follows. In the second section, we discuss the general framework. In section 3 we build the theoretical model, which is then discussed in section 4. Finally, our findings are presented in section 5.

2 General Framework and Model Assumptions

Our objective is to build a model of bookmaker optimal pricing, assuming that there are two populations of bettors in the market, namely, outsiders and insiders. We begin by describing the general framework and primary assumptions with respect to the information possessed by the market agents and their betting criteria along with the trading process and the pricing response by bookmakers.

Assume there are N horses in a race. The problem of the bookmaker is that of determining the opening odds. We denote by $\theta_j(0)$ the odds quoted by the bookmaker at time 0 against horse j winning, where $j = 1, \ldots, N$. If a bet is successful then, ignoring taxes, the bettor receives back $1 + \theta_j(0)$ on a one-unit bet. An opening price $OP = \phi_j(0)$ implies odds $\theta_j(0) = \frac{1-\phi_j(0)}{\phi_j(0)}$.

Suppose also that the horses' true winning probabilities at any point in time t are given by $P_j(t), j = 1, \ldots, N$, where $\sum_{j=1}^{N} P_j(t) = 1$. These true winning probabilities are assumed to evolve according to the flow of information, both public and private, throughout the betting period until the race starts and are therefore stochastic. Moreover, we assume that the flow of information is tied to the flow of bets. In this sense new information is said to have hit the market only if bets arrive in the marketplace in a way that alters the horses' winning probabilities, as those were until then perceived by the bookmaker. The stochastic process for the true winning probabilities could be either continuous or discontinuous, that is, a jump process or a mixture of the two. Strictly speaking, the process that affects the true probability should be seen as discontinuous, since the flow of information from small events that may affect the outcome of the race is not continuous. Moreover, we assume that the flow of public information during the betting period is negligible, at least compared to the amount of private information that may hit the market. This assumption makes sense especially if one considers the nature of racetrack betting and the short betting period (about 30 minutes). Moreover, it implies that whenever the bets arrive in a way different from the bookmaker's expectation, it is due to trading on inside information, unknown to the bookmaker until the actual trade has taken place. The above suggests that the expected value of $P_j(t)$ at any point in time, $E\left[P_j(t)\right]$, should be equal to the initial value $P_j(0)$.

Regarding the information possessed by the two presumed populations of bettors, namely outsiders and insiders, and their betting behavior, we make the following assumptions. First, nobody, not even an insider, knows in advance which horse will win the race, in contrast to Hyun Song Shin (1991, 1992, 1993), who assumed that insiders know which horse will win the race. Second, an insider knows only the true winning probability of one horse, k, $\hat{P}_k$, before this knowledge becomes public. However, the insider does not know how $1 - \hat{P}_k$ is distributed among the other horses. Given the quoted opening price for horse k, $\phi_k(0)$, this true winning probability might involve a profit opportunity for the insider. A risk-neutral insider will wager on horse k only if he or she expects a positive return. The expected return of the insider on a one-unit bet is the expected value of either $\left(-1 + \hat{P}_k \big/ \phi_j(0)\right)$ or zero, whichever is greater, since the insider bets only if $-1 + \hat{P}_k \big/ \phi_j(0) > 0 \Leftrightarrow \hat{P}_k > \phi_j(0)$. This is similar to saying that bookmakers actually offer insiders (call) options on the horses. Obviously it is in the bookmakers' interest to offer net out-of-the money options. However, when they err by underestimating a particular horse's true winning probability, they are liable to offer a net in-the-money option on this particular horse, which the insider (who knows the horse's true winning probability) will be glad to snap up.

Outsiders have access only to public information regarding past performance and current conditions. Therefore, we would expect outsiders to support the horses in proportion to the winning probabilities implied by "public information," $P_j(0)$, which are equal to the expected values of the true winning probabilities at the closing of betting, $E\left[P_j(T)\right]$. However, in reality the winning probabilities perceived by the outsiders should also account for their attitudes toward risk as well as for the existence of any behavioral biases among them. Consequently, outsiders are assumed to support the horses in proportion to their subjective winning probabilities, denoted by $\pi_j(t)$. A favorite-longshot bias may arise if bettors are risk-loving (e.g., Quandt 1986) or due to behavioral biases, such as those considered by Daniel Kahneman and Amos Tversky (1979). There may of course also be herding, which would lead to plunge horses being overbet. The bookmakers are also assumed to know the horses' winning probabilities implied by "public information," that is, $E\left[P_j(T)\right]$. Compared to outsiders, bookmakers are particularly skillful in gathering and processing public information and are therefore assumed to also know the marginal density function of each horse.[2] In addition, we assume that the bookmaker can accurately predict the expectations of outsiders, that is, the outsiders' subjective probabilities are known with certainty to the bookmaker.

Trading proceeds in a number of stages. At time zero the bookmaker declares the opening prices (OPs), $\phi_j(0)$, based on the bookmaker's perception of the true winning probabilities at this time, $P_j(0)$. At this first stage, a proportion of the outsiders bet in the market at the OP set by the bookmaker. Suppose now that a private signal revealed to a group of insiders indicates that the true winning probability of k is actually higher than the quoted OP, that is, $\hat{P}_k > \phi_k(0)$. The insiders will then bet on this horse, say at time t^*. Note that such signals indicating mispricing might be revealed for more than one horse. The bookmaker observes the insider betting pattern and therefore the new value of the true winning probability and adjusts the prices accordingly. At the other stages, the rest of the outsiders bet at the new updated prices. Note that insiders are faced with a timing dilemma. To understand this, suppose that there are two such groups of insiders, each wishing to plunge their own horse. Since a plunge reduces the prices of other horses, each group has an incentive to wait for the other to plunge first. Insiders must utilize any special information they have during the betting, since it loses all value once the race starts. Furthermore, since insider trading is both legal—only jockeys are forbidden to bet—and takes place at fixed prices, insiders have no incentive to hide their trading behavior from outsiders. Moreover, since the insider information concerning any given horse is likely known to more than one person, the longer insiders wait, the greater the risk that the information will leak to a third party. The recipient of the leak will then plunge the horse, and the group of insiders—except perhaps the one responsible for the leak—may be left with odds at which betting is no longer worthwhile (see also Schnytzer and Shilony 2002).

In the option pricing framework developed in this study to model the effect of information asymmetries on prices, we did not account for competitive interactions among insiders, since this would increase significantly the complexity of the problem in hand while offering limited additional insight. For simplicity, we assumed that insiders will place their bet once they receive the private signal.[3]

Price updating effectively continues until the last stage, during which starting prices (SPs) are determined as the equilibrium prices observed in the market at the end of betting. Since in contrast to the British market there is no legal SP betting in the Australian market, these prices may be assumed to embody all the available useful information regarding the race's outcome. Although price updating might actually take place several times throughout the betting period, our empirical analysis considered only three stages, the first, an intermediate, and the last stage, at which opening prices (OPs), middle prices (MPs) and starting prices (SPs), respectively, are set.

The chapter develops a model of bookmaker pricing that can be used to derive not only the OP but also any intermediate prices. At each point in time the prices are modeled as the equilibrium of a perfectly competitive bookmaker market. Specifically, the bookmaker is assumed to be risk-neutral, (i.e., an expected profit maximizer) and there is free entry in the market. Thus the long-run competitive equilibrium will be established when all bookmakers earn zero expected profits in the market corresponding to each race. Moreover, assuming perfect competition allows for the simplifying assumption of inelastic outsider demand. Note that if the bookmaker were a monopolist and demand were totally inelastic, maximizing profits would lead to unbounded prices (a formal proof is presented in the appendix to this chapter).

Insiders are assumed to have a collective wealth, W_i. When bookmakers price horses according to the methodology developed in this chapter, they assume that insiders bet to the full extent of W_i should the opportunity arise and that W_i is evenly distributed among insider horses. Therefore, in a race of N horses, up to $(1/N)\, W_i$ can be placed by insiders on each horse.

We do not make any assumptions concerning the likelihood of inside traders vis-à-vis either favorites or long shots. Finally, transaction costs are assumed to be negligible.

3 The Theoretical Model

3.1 Development of the Mathematical Model

The problem of the bookmaker is that of determining the opening odds $\left(1+\theta_j(0)\right)$ for each one of the N horses in a race such that the expected profit is equal to zero. Assume for the moment that only outsiders exist in the market. Then, ignoring the time-value of money, the expected profit of the bookmaker at time zero (stage 1) is equal to the total amount of money, W_n, bet by outsiders at stage 1 on the N horses minus the amount of money that the bookmaker is expected to pay out to the winners. Assume also that $w_{n,j}$ is the amount bet on horse j, where $j = 1, 2, \ldots, N$ and $E_0\left[P_j(T)\right]$ is the expected value of the true winning probability of horse j at the end of the betting period (time T). Note that as explained in the previous section, the proportionate amount of money bet by outsiders on each horse, $w_{n,j}/W_n$, is known to the bookmaker. Regarding the true winning probabilities, these might change throughout the betting period, since they evolve according to the flow of information, public and private, as this information is revealed through the flow of bets. However, in the absence of insiders and under the assumption that the flow of public information during the betting period is negligible (see section above), then $E_0\left[P_j(T)\right] = P_j(0)$. The expected profit of the bookmaker is

$$E_0(\Pi) = W_n - \sum_{j=1}^{N} P_j(0)\, w_{n,j}\left(1+\theta_j(0)\right). \tag{18.1}$$

Setting $\phi_j(0) = \frac{1}{1+\theta_j(0)}$, where the notation $\phi_j(0)$ is used to denote the opening prices (OPs), we obtain

$$E_0(\Pi) = W_n - \sum_{j=1}^{N} \frac{P_j(0)}{\phi_j(0)} w_{n,j}. \tag{18.2}$$

Given that $\sum_{j=1}^{N} P_j(0) = 1$, for the bookmaker to have a zero expected profit it is sufficient that for each j the OP satisfy the following equation:

$$\phi_j(0) = \frac{w_{n,j}}{W(W_n)}. \tag{18.3}$$

Therefore, if only outsiders exist in the market and, as assumed earlier, the bookmaker can accurately predict their expectations, for the latter to have zero expected profit on each horse it is sufficient that opening prices are set equal to the expectation of the bookmaker about the proportion of money bet on each horse, that is, $\phi_j(0) = \pi_j$, where $\pi_j = w_{n,j}\big/W_n$ is the winning probability of horse j as perceived by outsiders. Considering that π_j actually reflects outsiders' beliefs as shaped by public information, risk attitudes, and behavioral biases, under the assumption that the flow of public information is small, there is no reason for the opening prices to change during the betting period.

Suppose now that insiders also exist in the market. Obviously the final distribution of bets will depend on the expectations of both outsiders and insiders. The bookmaker can predict with accuracy the expectations of outsiders but not those of insiders, since the latter are shaped according to the private information they receive; moreover, this information is revealed to the bookmaker only after an inside trade has taken place.

Assume again that the bookmaker gives at time zero (opening) prices $\phi_j(0)$ for each one of the horses and that the betting period is again T periods of time. It is assumed that only part of the outsiders will bet at OP, $\omega_n^{OP} = \left(W_n^{OP}\big/W_n\right)$, while the other part will bet at later stages after observing insider behavior. A risk-neutral insider will wager on horse k only if the insider expects a positive return. The expected return of the insider on a one-unit bet is the expected value of either $\left(-1 + \hat{P}_k\big/\phi_j(0)\right)$ or zero, whichever is greater, since the insider bets only if $-1 + \hat{P}_k\big/\phi_j(0) > 0 \Leftrightarrow \hat{P}_k > \phi_j(0)$.

Under the above assumptions, the bookmaker is always expected to lose from trading with insiders. In particular, the bookmaker's expected loss to an insider at time zero on a one-unit bet (placed at time t^*) is

$$E_0(\Pi_i) = -E_0\left[\max\left(-1 + \frac{\hat{P}_k}{\phi_j(0)}, 0\right)\right]. \tag{18.4}$$

It holds that $\hat{P}_k = P_k(t^*) \neq P_k(0)$, where $P_k(t^*)$ is the true winning probability of horse k at the time the insiders place their bet (which is now revealed to the bookmaker).

The expected profit of the bookmaker is therefore

$$E_0(\Pi) = W_n^{OP} - \sum_{j=1}^{N} \frac{E_0\left[P_j(T)\right]}{\phi_j(0)} w_{n,j}^{OP} - \sum_{j=1}^{N} w_{i,j}^{OP} E_0\left\{\max\left(-1+\frac{P_j(t^*)}{\phi_j(0)},0\right)\right\}, \tag{18.5}$$

where $w_{i,j}^{OP}$ is the amount of money bet by insiders at OP on horse j. Note that now that insiders also exist in the market, the bookmaker cannot know what the true wining probability will be at the end of the betting period. However, as explained in the previous section, the bookmaker is assumed to know the expected value of the true winning probability, $E\left[P_j(t)\right]$, at any time t.

The expression above is complicated by the fact that t^* cannot be known a priori to the bookmaker, and hence it should be treated as stochastic. In order to simplify this we assumed that private information that may alter the true winning probability of a given horse may arrive only once for each horse. Then we can safely state that $\hat{P}_k = P_k(t^*) = P_k(T)$, where $P_k(T)$ is the value of the true winning probability at the closing of betting since, as pointed out earlier, private information regarding a certain horse may arrive in the marketplace only once. Of course one might argue that the true winning probability of horse k may be lowered if at a later time new (positive) information regarding a second horse r hits the market, implying $\hat{P}_r > P_r(0)$. Obviously, this would always be true in a race of two horses only. However, in a race of many horses, one could accept the supposition that this new information would reduce the true winning probabilities of all other horses (for which no inside information has hit the market) except for horse k.

Given that $w_{i,j}^{OP} = (1/N)W_i$, for the bookmaker to have zero expected profit, the following condition must be met:

$$1 - \sum_{j=1}^{N} \frac{w_{n,j}^{OP}}{W_n^{OP}} \frac{E_0\left[P_j(T)\right]}{\phi_j(0)} = \frac{1}{N}\frac{W_i}{W_n^{OP}} \sum_{j=1}^{N} E_0\left\{\max\left(-1+\frac{P_j(T)}{\phi_j(0)}\right),0\right\}, \tag{18.6}$$

or

$$1 = \sum_{j=1}^{N} E_0\left[P_j(T)\right]\left\{\frac{w_{n,j}^{OP}}{W_n^{OP}}\frac{1}{\phi_j(0)} + \frac{1}{N}\frac{W_i}{W_n^{OP}}\frac{E_0\left\{\max\left(-1+\frac{P_j(T)}{\phi_j(0)}\right)\right\},0}{E_0\left[P_j(T)\right]}\right\}. \tag{18.7}$$

Given that $\sum\limits_{j=1}^{N} E_0\left[P_j(T)\right] = 1$, for the above equation to hold, it is sufficient that the opening price of each horse j satisfies the following equation:

$$\frac{w_{n,j}^{OP}}{W_n^{OP}}\frac{1}{\phi_j(0)} + \frac{1}{N}\frac{W_i}{W_n^{OP}}\frac{E_0\left\{\max\left(-1+\frac{P_j(T)}{\phi_j(0)}\right),0\right\}}{E_0\left[P_j(T)\right]} = 1, \tag{18.8}$$

or, multiplying with the term $\left(\frac{W_n^{OP}}{W_n}\right)$ and rearranging we obtain

$$\left(\frac{W_n^{OP}}{W_n}\right) - \left(\frac{W_n^{OP}}{W_n}\right)\frac{w_{n,j}^{OP}}{W_n^{OP}}\frac{1}{\phi_j(0)} = \frac{1}{N}\left(\frac{W_i}{W_n}\right)\frac{E_0\left\{\max\left(-1+\frac{P_j(T)}{\phi_j(0)}\right),0\right\}}{E_0\left[P_j(T)\right]}. \tag{18.9}$$

The left-hand side of the above equation is the expected bookmaker gain from trading with outsiders while the right-hand side is the expected bookmaker loss to insiders. The optimal price is the one that equalizes the gain from outsiders to the loss to insiders. It can be found by solving the above equation through trial and error, given the proportion of outsiders who bet at OP, $\omega_j^{OP} = W_n^{OP}/W_n$, outsider's subjective probabilities, $\pi_j^{OP} = w_{n,j}^{OP}/W_n^{OP}$, the bookmaker's expectation about the true winning probability at the closing of betting, $E_0\left[P_j(T)\right]$, the number of runners in a race, N, and, of course, the degree of insider trading (as perceived by the bookmaker) defined as the ratio of total insider money to total outsider money, $\left(W_i/W_n\right)$.

Note that the left-hand side of this equation should be greater or equal to zero since the right-hand side is always nonnegative. Therefore, if insiders exist in the market, in order for the bookmaker to have zero expected profit, prices should be set greater than outsiders' subjective probabilities, that is,

$$\phi_j(0) \geq \frac{w_{n,j}^{OP}}{W_n^{OP}} = \pi_j^{OP}. \tag{18.10}$$

To summarize, our model suggests that since any private information is conveyed to the bookmaker only after an informed trade takes place, the latter should include a premium in the OP to compensate for this risk. Moreover, this premium is related to the cost of trading with insiders, which in turn is a function of the degree of insider trading (W_i/W_n) and the potential value of private information that may be exploited by insiders (as captured by the term $E\left\{\max\left(1-+\frac{P_j(T)}{\phi_j(0)},0\right)\right\}$). Under these considerations, the sum of OP would always be greater than one.

Suppose now that at a later point in time, time τ (stage 2), after the bookmaker has observed insider trading, the bookmaker will set new prices (called MP). For those horses on which insider trading has taken place, say m horses, prices will be set equal to the horses' new true winning probabilities (in the absence of any bookmaker margin). The reason is that insiders pose no further risk to the bookmaker since they can only bet at either OP or MP on a given horse but not both.[4] For the rest of the horses $(N-m)$, the bookmaker will set prices as above. Therefore, at the second stage the total amount of money available by insiders is $W_i - \frac{m}{N}W_i \leq W_i$. Thus, the bookmaker will quote MP as if $\frac{1}{N-m}\left(W_i - \frac{m}{N}W_i\right) = \left(1/N\right)W_i$ would be wagered by insiders on each one of the remaining $N-m$ horses should the opportunity arise. Therefore we have

$$\left(\frac{W_n^{MP}}{W_n}\right) - \left(\frac{W_n^{MP}}{W_n}\right)\frac{w_{n,j}^{MP}}{W_n^{MP}}\frac{1}{\phi_j(\tau)} = \frac{1}{N}\left(\frac{W_i}{W_n}\right)\frac{E_\tau\left\{\max\left(-1+\frac{P_j(T)}{\phi_j(\tau)}\right),0\right\}}{E_\tau\left[P_j(T)\right]}, \tag{18.11}$$

where the term $\pi_j^{MP} = w_{n,j}^{MP} \big/ W_n^{MP}$ captures outsiders' new subjective probabilities as these have been shaped after observing the insider trading pattern at stage 1 and $\omega_j^{MP} = W_{n,j}^{MP} \big/ W_n$ is the proportion of outsiders that bets at this second stage.

Price updating effectively continues until the last stage at which starting prices (SPs) are determined as the equilibrium prices observed in the market at the end of betting. Under the assumption of zero bookmaker profit, the sum of SP would be equal to one. Then, following our model, in the presence of insiders the sum of OP should always be greater in any race than the sum of SP. In reality, the sum of OP is always greater in any race than the sum of SP, even in the apparent absence of insider trading.[5] The reason is that opening prices tend to have a so-called cartel level of profit built in since they are recommended to individual bookmakers by the bookmakers' association. Once betting begins competition among bookmakers also begins, and thus the sum of prices will tend to decrease. This practically means that the estimates of insider trading obtained when applying our model may overestimate its true extent if the premium included in OP is largely due to this cartel profit rather than to the risk that bookmakers face in the presence of insiders. On the other hand, it may be that the expected profit margins built into OP are designed just to compensate the bookmakers for inside trades.

3.2 The Option Analogy

The commitment made by bookmakers to sell at fixed prices, the quoted odds, can be analyzed as a call option. Specifically, the bookmaker gives an insider a call option on horse j, that is, the right to bet at a fixed price. Obviously the underlying asset whose value changes stochastically is horse j's true winning probability. Apparently, only insiders are entitled to the option. The reason is that while an insider has perfect information (both public and private) and therefore knows a horse's true winning probability, outsiders form their expectations, at least partially, according to the public component of information, and based on that they assign subjective probabilities. The insider will exercise the option to bet at the opening prices only if the insider expects a positive return, that is, if the true probability at the time the bet is placed, t^*, is greater than the opening price.

One could assume that insiders would be better off exercising their option at the last minute, that is, at the closing of betting. The reason is that since a plunge reduces the prices of other horses each group of insiders has an incentive to wait for the other groups to plunge first. However, since the insider information concerning any given horse is likely known to more than one person, the longer insiders wait, the greater the risk that the information will leak to a third party. The recipient of the leak will then plunge the horse and the group of insiders—except perhaps the one responsible for the leak—may be left with odds at which betting is no longer worthwhile. This timing dilemma is similar to the problem of the optimal exercise time faced by the holder of an American option on a dividend-paying stock. In the betting market the dividend equivalent is the potential value leakage as a result of competition among insiders. However, for simplicity we ignored competitive interactions among insiders. We assumed instead that insiders place their bet once they observe mispricing.

Assuming that the bookmaker is risk-neutral, today's option price (time zero) can be determined by discounting the expected value of the terminal option price by the riskless rate of interest. Therefore, neglecting the time-value of money, the value of the call option is

$$C_j(0) = C_j^{OP} = E_0 \left\{ \max \left(-1 + \frac{P_j(T)}{\phi_j(0)}, 0 \right) \right\}. \tag{18.12}$$

Similarly

$$C_j(\tau) = C_j^{MP} = E_\tau \left\{ \max \left(-1 + \frac{P_j(T)}{\phi_j(\tau)}, 0 \right) \right\}. \tag{18.13}$$

The value of the option can be derived by assuming a stochastic process for the true winning probability and performing Monte Carlo simulations (see section 4).

3.3 The Favorite-Longshot Bias

In this section we show that the optimal prices set by the bookmaker using equation (18.9) will exhibit the favorite-longshot bias.

Expected returns will exhibit the favorite-longshot bias if and only if $\partial E(R_j) / \partial \left(E_0[P_j(T)] \right) > 0$, where $E(R_j) = -1 + E_0\left[P_j(T)\right] / \phi_j$. This is equivalent to

$$\frac{\partial \left(E_0 \left[P_j(T) / \phi_j \right] \right)}{\partial \left(E_0 \left[P_j(T) \right] \right)} > 0. \tag{18.14}$$

Denoting $f_j(0) = \left(E_0 \left[P_j(T) \right] / \phi_j \right)$, equation (18.9) can be written as

$$\left(\frac{W_n^{OP}}{W_n} \right) E_0 \left[P_j(T) \right] - \left(\frac{W_n^{OP}}{W_n} \right) \left(\frac{w_{n,j}^{OP}}{W_n^{OP}} \right) f_j(0) = \frac{1}{N} \left(\frac{W_i}{W_n} \right) E_0 \left\{ \max \left(-1 + f_j(T) \right), 0 \right\}, \tag{18.15}$$

where $f_j(T) = \left(E_T[P_j(T)] / \phi_j \right) = P_j(T) / \phi_j$

Differentiating the above with respect to $E_0 \left[P_j(T) \right]$ and setting $\frac{\partial E_0 \{ \max(-1 + f_j(T)), 0 \}}{\partial E_0 \left[P_j(T) \right]} = \frac{\partial E_0 \{ \max(-1 + f_j(T)), 0 \}}{\partial f_j(0)} \frac{\partial f_j(0)}{\partial E_0 \left[P_j(T) \right]}$, we obtain

$$\frac{\partial f_j(0)}{\partial E_0 \left[P_j(T) \right]} = \frac{\left(\frac{W_n^{OP}}{W_n} \right) \left(1 - f_j(0) \frac{\partial \left(w_{n,j}^{OP} / W_n^{OP} \right)}{\partial \left(E_0 \left[P_j(T) \right] \right)} \right)}{\left(\frac{W_n^{OP}}{W_n} \right) \left(\frac{w_{n,j}^{OP}}{W_n^{OP}} \right) + \frac{1}{N} \left(\frac{W_i}{W_n} \right) \frac{\partial E_0 \{ \max(-1 + f_j(T)), 0 \}}{\partial f_i(0)}}. \tag{18.16}$$

We focus on the denominator first. The first term is always positive. The second term is positive too since the partial derivative $\frac{\partial E_0\{\max(-1+f_j(T)),0\}}{\partial f_j(0)}$ is always positive. Note that a higher level of $f_j(0) = \left(E_0\left[P_j(T)\right]/\phi_j\right)$ is equivalent to a lower quoted price for the same level of expected true probability. Therefore the potential profit of insiders, as captured by the term $E_0\left\{\max\left(-1+f_j(T)\right),0\right\}$, should increase since a lower quoted price makes underpricing more likely. In the terminology of options this is equivalent to saying that a lower strike price increases the value of a call option.

We turn our attention now to the nominator. For the nominator to be positive it is necessary that the term $1 - f_j(0)\frac{\partial\left(w_{n,j}^{OP}/W_n^{OP}\right)}{\partial(E_0[P_j(T)])}$ is positive. This obviously depends on the partial derivative of the outsiders' subjective probability with respect to the expected true winning probability. If the subjective winning probabilities of outsiders, $\pi_j(t)$, are equal to the winning probabilities implied by "public information," then this partial derivative will be zero and the nominator will be positive. Suppose instead that outsiders tend to overestimate the winning chances of long shots relative to those of favorites, as is often argued in the literature, that is,

$$\frac{\partial\left(E_0\left[P_j(T)\right]/\left(w_{n,j}^{OP}/W_n^{OP}\right)\right)}{\partial\left(E_0\left[P_j(T)\right]\right)} > 0 \Leftrightarrow 1 - \frac{E_0\left[P_j(T)\right]}{\left(w_{n,j}^{OP}/W_n^{OP}\right)}\frac{\partial\left(w_{n,j}^{OP}/W_n^{OP}\right)}{\partial\left(E_0\left[P_j(T)\right]\right)} > 0.$$

We know that

$$\phi_j > \frac{w_{n,j}^{OP}}{W_n^{OP}} \Rightarrow \frac{1}{\phi_j} < \frac{1}{\left(w_{n,j}^{OP}/w_n^{OP}\right)} \Rightarrow \frac{E_0\left[P_j(T)\right]}{\phi_j} < \frac{E_0\left[P_j(T)\right]}{\left(w_{n,j}^{OP}/W_n^{OP}\right)}$$

$$\Rightarrow \frac{E_0\left[P_j(T)\right]}{\phi_j}\frac{\partial\left(w_{n,j}^{OP}/W_n^{OP}\right)}{\partial\left(E_0\left[P_j(T)\right]\right)} < \frac{E_0\left[P_j(T)\right]}{\left(w_{n,j}^{OP}/W_n^{OP}\right)}\frac{\partial\left(w_{n,j}^{OP}/W_n^{OP}\right)}{\partial\left(E_0\left[P_j(T)\right]\right)}$$

$$\Rightarrow 1 - \frac{E_0\left[P_j(T)\right]}{\phi_j}\frac{\partial\left(w_{n,j}^{OP}/W_n^{OP}\right)}{\partial\left(E_0\left[P_j(T)\right]\right)} > 1 - \frac{E_0\left[P_j(T)\right]}{\left(w_{n,j}^{OP}/W_n^{OP}\right)}\frac{\partial\left(w_{n,j}^{OP}/W_n^{OP}\right)}{\partial\left(E_0\left[P_j(T)\right]\right)} > 0$$

Therefore we have proved that when the bookmaker sets optimal prices following our model, expected returns will exhibit the favorite-longshot provided that either outsiders have no biases in their expectations and therefore their subjective probabilities reflect the publicly available information or that they tend to overestimate the winning chances of long shots relative to those of favorites.

4 Empirical Model

4.1 Option-Pricing Specifications of the Model

The challenge faced here is that the assumed specification must be a realistic description of probability dynamics. In particular we want to model the true winning probability such that the following requirements are met: First, said probability is concentrated on [0,1). A probability of a certain horse winning equal to one implies that the probabilities of all other horses are zero. In practice this is never the case. For this reason we set as an upper boundary for the true winning probabilities the value $p_{\max} < 1$. In particular $p_{\max}$ could be the highest single probability in our sample, which is found to be 0.7197. Second, the sum of probabilities is equal to one at all times. Third, it may exhibit positive and/or negative jumps throughout the betting period following the release of new private information. Finally, in the long run it reverts to a mean equal to the reciprocal of the number of runners in a race. This assumes that over a long period of time all horses have equal chances of winning. Note that the behavior of this process in the absence of mean reversion is problematic since in this case the boundaries become absorbing.

Taking the above under consideration, the following stochastic process is assumed:

$$dP_j(t) = h\left(\mu - P_j(t)\right)dt + P_j(t)\left(P_{\max} - P_j(t)\right)Jdq, \tag{18.17}$$

where h is the speed of mean reversion, μ is the long-run mean (equal to $1/N$), J is the jump size, which is assumed to be normally distributed with mean zero and standard deviation σ_J, and dq describes a time-homogeneous Poisson jump process such that $dq = 1$ with probability λdt and $dq = 0$ with probability $(1 - \lambda dt)$. Parameter λ is known as intensity or arrival rate and is the expected number of "events" or "arrivals" that occur per unit time. The term $P_j(t)\left(P_{\max} - P_j(t)\right)$, which multiplies the jump component Jdq, is employed in order to ensure that the probability will remain inside the boundaries of zero and $p_{\max}$. Furthermore, given that the jump size has a mean of zero, it can be easily shown that the expected value of $P_j(t)$ at any $t > 0$ is given by

$$E\left[P_j(t)\right] = P_j(0)e^{-ht} + \mu\left(1 - e^{-ht}\right). \tag{18.18}$$

Note that when the speed of mean reversion is very small, as assumed herein, the expected value of $P_j(t)$, $E\left[P_j(t)\right]$, tends to the initial value $P_j(0)$. This is important since the theoretical model described previously in this chapter relied heavily on this assumption.

There is one final concern with respect to the specification for the true winning probability, which, as mentioned earlier, refers to the fact that the sum of probabilities must be equal to one at all times. Suppose that the probability of horse k follows the above stochastic process while for all other horses j, $j = 1, 2, \ldots, N$, $j \neq k$, it holds that

$$dP_j(t) = h\left(\mu - P_j(t)\right) + \varepsilon_j(t). \tag{18.19}$$

Then, taking the sum of all probabilities, setting it equal to one and observing that $\sum_{j=1}^{N} h\left(\mu - P_j(t)\right) dt = 0$, it follows directly that

$$\sum_{\substack{j=1 \\ j \neq k}}^{N} \varepsilon_j(t) + J_k P_k(t)(P_{\max} - P_k(t))dq = 0. \tag{18.20}$$

Therefore, although specification (18.17) does not warrant that $\sum_{j=1}^{N} P_j(t) = 1$, we can find a condition under which this holds. Thus the above specification is indeed a realistic description of probability dynamics. We now need to estimate the parameters that appear in the stochastic process followed by the true winning probability. For the purpose of this estimation we will ignore the mean-reverting component, assuming instead that the speed of mean reversion is very close to zero. Thus we only have to estimate the parameters of the jump process and, in particular, the standard deviation σ_J of the jump size and the intensity λ of the Poisson process. The intensity parameter tells us how often the true winning probability experiences a sudden jump while the parameter of jump volatility measures the size of these jumps. We calculate these parameters by computing the second and fourth (raw) moments. These are specified as follows:

$$\mu_2 = E\left(Y^2\right) = E\left(J^2\right) E\left(dq^2\right) = \sigma_j^2 \lambda \Delta t \tag{18.21}$$

$$\mu_4 = E\left(Y^4\right) = E\left(J^4\right) E\left(dq^4\right) = 3\sigma_J^4 \lambda \Delta t, \tag{18.22}$$

where $Y = \dfrac{\Delta P}{P(1-P)}$.

Those two moments completely identify the jump components. Moreover, they can be derived from the bookmakers' odds as following: as the dataset includes prices at three points in time (OP, MP, and SP), prices are available roughly every 15 minutes. The 15-minute moments may thus be calculated for each race

$$2M : m_2 = \frac{1}{s-1} (u_1 - u_2)^2 \tag{18.23}$$

$$4M : \ m_4 = \frac{1}{s-1} (u_1 - u_2)^4 \tag{18.24}$$

where $s = 2$ and $u_{1,} u_2$can be calculated as follows:

$$u_1 = \frac{\phi_{MP} - \phi_{OP}}{\phi_{OP} \cdot (1 - \phi_{OP})} \tag{18.25}$$

$$u_2 = \frac{\phi_{SP} - \phi_{MP}}{\phi_{MP} \cdot (1 - \phi_{MP})}. \tag{18.26}$$

Obviously the one-minute moments can be calculated from the fifteen-minute moments by dividing by fifteen. Using equations (18.21) and (18.22) for $\Delta t = 1$ minute and the estimated values for the one-minute moments, we determined the jump components λ and σ_J for all horses in each race:

$$\lambda = \frac{3\mu_2^2}{\mu_4}$$

$$\sigma_j = \sqrt{\frac{\mu_4}{3\mu_2}}.$$

Finally, we calculate the average values of σ_j and λ for our sample, which are then used in the options calculations. Note that these are "one-minute" values. For example, $\lambda = 0.1$ implies that we have a jump every 10 minutes. The results are presented below.

4.2 A Measure of Insider Trading

We focus now on the task of estimating the degree of insider trading, that is, the parameter $\left(W_i / W_n\right)$. To this end, we assume that in practice bookmakers set their prices according to the methodology described above. Thus using the actual prices we can infer the degree of insider trading by using expression (18.6) to directly solve $\left(W_i / W_n\right)$. However, the theoretical model was built under the assumptions of zero expected profit and zero transaction costs. This may yield estimates of insider trading that are biased upward. Starting from expression (18.6)

$$1 - \sum_{j=1}^{N} \frac{w_{n,j}^{OP}}{W_n^{OP}} \frac{E_0\left[P_j(T)\right]}{\phi_j(0)} = \frac{1}{N} \frac{W_i}{W_n^{OP}} \sum_{j=1}^{N} C_j^{OP}. \tag{18.27}$$

By multiplying with $\frac{W_n^{OP}}{W_n}$, the part of outsider trading that occurs at OP,

$$\frac{W_i}{W_n} \frac{1}{N} \sum_{j=1}^{N} C_j^{OP} = \frac{W_n^{OP}}{W_n} \left(1 - \sum_{j=1}^{N} \frac{W_{n,j}^{OP}}{W_n^{OP}} \frac{E_0\left[P_j(T)\right]}{\phi_j(0)}\right) \tag{18.28}$$

or that

$$q \sum_{j=1}^{N} C_j^{OP} = N \frac{W_n^{OP}}{W_n} D^{OP}, \tag{18.29}$$

where $D^{OP} = \left(1 - \sum_{j=1}^{N} \pi_j^{OP} \frac{E_0[P_j(T)]}{\phi_j(0)}\right)$ for each race. The superscript OP indicates that these values refer to the first stage at which the opening prices are set. This is the basic equation for our empirical analysis. Obviously this expression refers only to OP. However, as we said, we assumed that trading takes place in two stages. At stage 1 ($t_1 = 0$) a proportion of the outsiders bet in the market at the OP set by the bookmaker. At any subsequent point in time, $t_1 + \Delta t \leq T$, all insiders may bet should the opportunity arise. The bookmaker observes the insider trading pattern and at time t_2, $t_1 < t_2 \leq T$, updates prices. At stage 2 the rest of the outsiders bet at the new set of updated prices, denoted by MP. Again, at any subsequent point in time, $t_2 + \Delta t \leq T$, all insiders may bet should the opportunity arise. The bookmaker observes the insider trading pattern and at time T sets new updated prices denoted by SP (starting prices).

Similarly, for the second stage at which MP is set we have

$$q \sum_{j=1}^{N} C_j^{MP} = N \frac{W_n^{MP}}{W_n} D^{MP}, \tag{18.30}$$

$$\text{where } D^{MP} = \left(1 - \sum_{j=1}^{N} \pi_j^{MP} \frac{E_\tau\left[P_j(T)\right]}{\phi_j(\tau)}\right).$$

We can use equations (18.29) and (18.30) to calculate the proportion of outsiders that bet at OP and MP

$$\omega^{OP} = \frac{W_n^{OP}}{W_n} = \frac{D^{MP} \sum_{j=1}^{N} C_j^{OP}}{D^{MP} \sum_{j=1}^{N} C_j^{OP} + D^{OP} \sum_{j=1}^{N} C_j^{MP}}. \tag{18.31}$$

Then we can use equations (18.29) and (18.31) to calculate q. In order to do so we still have to explain how to calculate the option values at both OP and MP, C_j^{OP} and C_j^{MP}, as well as the quantities D^{OP} and D^{MP} for each race.

We begin with betting at OP. The option values C_j^{OP} can be estimated via Monte Carlo simulation. The required inputs to perform the simulations are $P_j(0)$, OP_j, T, and the specifications of the stochastic process followed by the true winning probability. OP_j is the observed opening price quoted by the bookmaker. T is assumed to be equal to 30 minutes. With respect to the specifications of the stochastic process followed by the true winning probability, we need to know the speed of mean reversion, h, the long-run mean, μ, which is set equal to $1/N$, the standard deviation of the jump size, σ_J, given that $J \sim N\left(0, \sigma_J^2\right)$ and the intensity λ of the Poisson process. A way to derive those parameters has been shown in the previous section of this chapter. The speed of mean reversion is assumed to be very small since we are dealing with a betting period of no more than 30 minutes and is therefore set at 0.001. The true winning probability, $P_j(t)$, is derived via a conditional logit regression on a dummy win, ensuring that the sum of probabilities in each race equals 1. The subjective probabilities π_j^{OP} are calculated by simply normalizing OP as suggested by Jack Dowie (1976), though this yields estimates with a favorite-longshot bias. The true winning probability is simulated in 1,000 steps using the stochastic process (18.17). When the simulated true winning probability after 1,000 steps is larger than the true winning probability in time $t = 0$, the option value is this difference; otherwise the option value is zero. Each horse is subject to 1,000 repetitions. The option value for the horse is the averaged value over all repetitions. A similar procedure is followed to calculate C_j^{MP}, using as inputs $P_j(\tau)$, MP_j, and $T - \tau$, where τ is assumed to be equal to 15 minutes. We used the same specifications for the stochastic process as above.

We still need to calculate D^{OP} and D^{MP} for each race. The expected true winning probabilities at the end of the betting period, $E_0\left[P_j(T)\right]$, are assumed to be equal to the true winning probabilities at time 0, that is, $E_0\left[P_j(T)\right] = P_j(0)$. This is derived from equation (18.18) if we assume that the speed of mean reversion is very small. This way we assumed that mean reversion has almost no effect on the true winning probabilities in the very short betting period of 30 minutes while any deviations from the initial value are due to the effect of jumps that come as surprises.

Next, we used equation (18.31) to calculate $\omega^{OP} = \frac{W_n^{OP}}{W_n}$. The extent of insider trading for each race is then $q = \frac{N\omega^{OP} D^{OP}}{\sum_{j=1}^{N} C_j^{OP}}$

The probability of insider trading is simply

$$a = \frac{q}{1+q}.$$

As we said, our model was built from the viewpoint of the bookmaker, and the approach we have followed so far effectively supposes that bookmakers know the probability of insider trading in advance. Or, more reasonably, such a measure is of the bookmakers' expectations regarding insider trading. However, we have access to ex post plunging information, which the bookmaker cannot know until after insider activity has taken place. We will use this (ex post) plunging information in order to get closer to the true probability of insider trading for a given horse that got plunged.

In order to calculate the probability of insider trading per horse, we used both an unweighted and a weighted average of q. The weight was derived as the absolute size of the plunge, called PW: max(MP − OP, 0) + max(SP−MP, 0). Using the unweighted and weighted average, the probability of insider trading for each horse in a given race in the sample was calculated. Note that when we weighted absolute plunges sizes, we were weighting on those horses that insiders were observed to have bet on more heavily in accordance with plunge size. Using these weights, the weighted average probability of insider trading for each of the races in the sample was calculated. The simple average of these values is the probability of insider trading in the dataset.

5 Results

We used the above model to derive a measure of the extent of insider trading. Our measure was applied to a dataset of the 1998 Australian Horse Racing season, covering 4,017 races with 45,296 runners.[6] The dataset includes for each horse prices at three moments (OP, MP, and SP). The time period during which betting takes place is set at 30 minutes, meaning that prices are available roughly every 15 minutes. Given that there were some cases in which the sum of OP, MP, or SP in a race was less than one, these races are dropped from the sample. This left us with 3,995 races out of the initial sample of 4,017 races.

The true winning probabilities at OP and MP, necessary for the measure, were derived by running a conditional logit regression, the results of which are displayed in table 18.1.

Table 18.2 displays descriptive statistics for OP, MP, SP; the subjective winning probabilities at OP and MP; the true winning probabilities following from table 18.1; and the sum of OP, MP, and SP per race.

The table shows clearly that the average sum of prices decreases between OP and SP. At the opening of betting this margin was 43 percent, but by the start of the race the margin had decreased to 24 percent. The decrease in the margin indicates competition among bookmakers, forcing them to decrease prices and leading to lower profits. Since the OPs are above the competitive level, this could deter insiders from trading at these prices, leading to a lower degree of insider trading.

The option values were generated via Monte Carlo simulation as explained in the previous section. The average 1-minute values for λ and σ in the dataset are 0.37 and 0.11, respectively. On average, there seems to be a jump every 3 minutes or 10 times per a 30-minute betting period, indicating quite some inflow of private information into the prices. Table 18.3 shows the values of the non-zero options generated at OP and MP.

Table 18.1 Conditional Logit Regression

	Win	Win
OP	6.713***	
	(0.133)	
MP		7.155***
		(0.141)
N	45266	45266
Log Likelihood	−8259.18	−8238.81

Note: Standard errors in parentheses.
*** significant at 1%.

Table 18.2 Descriptive Statistics

Variable	Mean	Min	Max	Standard Deviation
OP	0.1255	0.0019	0.8889	0.1017
MP	0.1107	0.0013	0.8670	0.0946
SP	0.1092	0.0010	0.8462	0.0972
$\pi_j(OP)$	0.0883	0.0014	0.7197	0.0731
$\pi_j(MP)$	0.0883	0.0011	0.7165	0.0768
$P_j(OP)$	0.0883	0.0026	0.9657	0.0970
$P_j(MP)$	0.0883	0.0020	0.9723	0.0984
$\sum_{j=1}^{N} OP$	1.4339	1.0225	2.0631	0.1117
$\sum_{j=1}^{N} MP$	1.2660	1.0003	1.8508	0.1008
$\sum_{j=1}^{N} SP$	1.2487	1.0122	1.7646	0.0921

There were 5,184 horses for which a zero option was generated at OP and 6,806 horses for which the option value was zero at MP. Moreover, we can see from table 18.3 that the options generated at MP have a higher value, indicating more profitable trading opportunities for insiders. This should not come as a surprise, as we already saw in table 18.2 that the MPs are lower, leading to a lower strike price for the insiders and a higher profit.

One last thing is required to calculate the degree of insider trading, namely ω^{OP}, the group of outsiders who bet at opening prices. Using the data from tables 18.2 and 18.3, the average ω^{OP} in the dataset was calculated to be 0.41. On average 41 percent of outsider trading occurs at OP and 59 percent at MP, though there are races in which almost no outsider betting is found to occur at OP. The ratio of insider betting to outsider betting, which is expected by the zero-profit bookmaker, was found to have a mean of roughly 25 or an average probability of insider trading of around 95 percent. The density plot is shown in 18.1. This seems very high, but there are a few considerations to take into account. First, the insider in our model was assumed to know only the true winning probability of one horse *k* and not the winning probabilities

Table 18.3 Option Statistics

Variable	N	Mean	Max.	Standard Deviation
C_j^{OP}	40,082	0.00454	0.20610	0.01385
C_j^{MP}	38,460	0.00723	0.25498	0.02087

of any other horses. When compared to the insider by Shin (1991, 1992, 1993), our insider does not know which horse will win but simply has a better understanding of the true winning probabilities compared to the probabilities as quoted by the bookmaker. Second, the definition of insider trading is the amount of money being bet by insiders compared to outsiders. The fact that 95 percent of the total money is bet by insiders does not mean that they place more bets, but it could mean that they wager more money per bet. The bulk of the bets placed could still be made by outsiders, but the amount that insiders bet compared to outsiders is just much higher; that is, for every Australian dollar bet by outsiders, 25 is bet by insiders. Third, the assumption underlying the measure is that the bookmaker set his expected profit at zero, as would be the case under perfect competition. This is, of course, a very strict assumption to make and may not suit the reality all that well. A solution would be to assume that the bookmaker sets prices that guarantee a certain level of profit. This level could be assumed to be equal to the profit the bookmaker would make in a market with no insiders. However, the prices that the bookmaker would set in this market are unobservable prices by definition. By looking at equation (18.32), keeping everything else constant and assuming the bookmaker sets prices to have a positive expected profit, it becomes obvious that we are estimating the degree of insider trading with an upward bias.

Finally, the measure that was generated is the bookmaker's expectation regarding insider trading. To have a zero expected profit, the bookmaker expects the probability of insider trading to be 95 percent. However, as mentioned in section 4, we used ex post plunging information to get closer to the true probability of insider trading in the dataset. The weight that we used is based on the absolute size of the plunge: $PW = \max(MP-OP, 0) + \max(SP-MP, 0)$. We used PW to weight the extent of insider

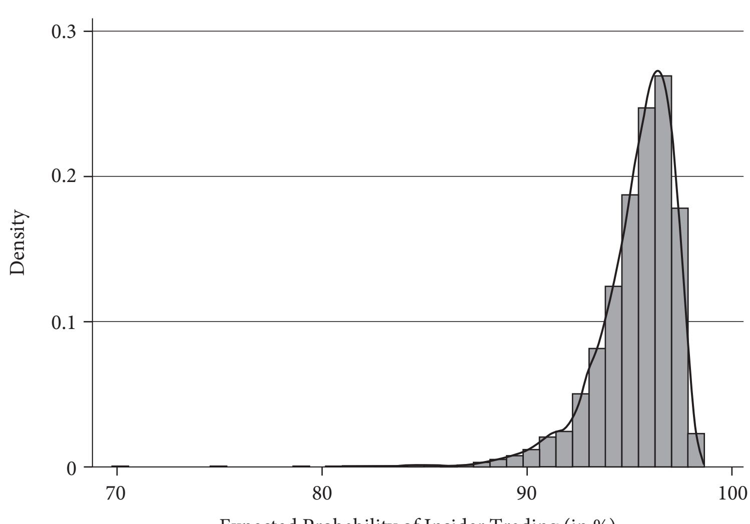


FIGURE 18.1 Expected probability of insider trading

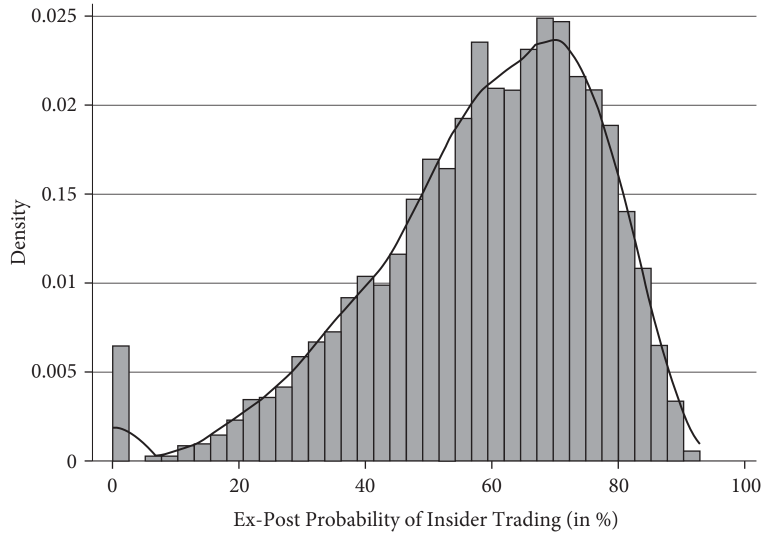


FIGURE 18.2 Ex post probability of insider trading

trading, q, per race. By defining a plunge as an upward movement of the price, we found that there have been 13,852 plunges in the dataset, mainly occurring between MP and SP. The 13,852 horses account for 30 percent of the total observations in the dataset. An additional benefit is that PW weights horses that have experienced a more severe plunge higher. However, a downside is that the estimate will be too high if part of the plunging is actually due to herding. The mean of the weighted extent of insider trading q_{pw} for the dataset is 2.20, and the average probability of insider trading, a_{pw}, is 59 percent. Figure 18.2 displays the distribution of the ex post probability of insider trading.

We can see that there are around 611 races in which no plunges occurred and hence no insider trading is observed. When insider trading does take place, the average probability is around 60 percent, though there is plenty of dispersion around the mean.

One final remark should be made with respect to the ex post probability of insider trading. It should be noted that the value depends on the bookmaker's expectation of the degree of insiders compared to outsiders. If we allow for a higher-than-zero expected profit, the bookmaker's expectation will be lower and we would find values of insider trading closer to the 20–30 percent range found by Schnytzer, Lamers, and Makropoulou (2010).

6 Conclusions

In this chapter, we modeled a fixed-odds horse betting market from a bookmaker's point of view under uncertainty. We relied on a model by Makropoulou and Markellos

(2011) and Schnytzer, Lamers, and Makropoulou (2010) that conceptualizes fixed-odds betting markets as option markets. Starting from a profit function we showed that a bookmaker offers an implicit call option to insiders when setting prices. The insiders in this chapter are assumed to know only the true winning probability of their horses, not the identity of the winning horse. Moreover, insiders bet only if their expected profit is positive. In the case in which both outsiders and insiders exist in the market, the bookmaker will set prices in such a way that the expected loss from dealing with insiders equals the expected gain from dealing with outsiders. When the bookmaker sets prices in this way, the latter will exhibit a favorite-longshot bias.

By allowing for betting in multiple time periods and making an assumption on how the insider money will arrive, the zero-profit condition of Schnytzer, Lamers, and Makropoulou (2010) has to hold only for the race and not for each individual horse. From this model it becomes possible to measure the expectations of the bookmaker regarding the ratio of insider money to outsider money. Using Monte Carlo simulations, we generated the implicit option values as quoted by the bookmaker and found that a zero-expected-profit bookmaker expects 95 percent of the money bet to be placed by insiders. However, these estimates are biased in the sense that we do not allow the bookmaker to make a positive profit. By keeping the expected profit equal to zero, we overestimated the expected degree of insider trading. When we used ex post plunging information, we concluded that the probability of insider trading in our dataset lies around 59 percent. Or to put it differently, for every Australian dollar bet by outsiders, the average amount bet by insiders is 2.20 dollars.

Appendix

Proof

Suppose first that only outsiders exist in the market and that their demand is inelastic, that is, $\frac{\partial W_n}{\partial OP_j} = 0$. A monopolistic bookmaker will set prices that maximize his expected profit:

$$\max E(\Pi) = W_n - \sum_{j=1}^{N} E\left[P_j(T)\right] \frac{w_j}{W_n} W_n \frac{1}{OP_j}$$

$$\frac{\partial E(\Pi)}{\partial OP_j} = 0 \Rightarrow -E\left[P_j(T)\right] \frac{w_j}{W_n} W_n \frac{-1}{OP_j^2} = 0$$

Since infinite prices do not make any sense, this leads to the conclusion that outsiders' demand should be elastic, that is, $\frac{\partial W_n}{\partial OP_j} < 0$. In this case we have the following solution:

$$\frac{\partial E(\Pi)}{\partial OP_j} = 0 \Rightarrow \frac{\partial W_n}{\partial OP_j} - E\left[P_j(T)\right]\frac{w_j}{W_n}W_n\frac{-1}{OP_j^2} - E\left[P_j(T)\right]\frac{w_j}{W_n}\frac{1}{OP_j}\frac{\partial W_n}{\partial OP_j} = 0.$$

Suppose now that insiders also exist in the market:

$$\max E(\Pi) = W_n\left(\sum_{j=1}^{N} OP_j\right) - \sum_{j=1}^{N} E\left[P_j(T)\right]\frac{w_j}{W_n}W_n\frac{1}{OP_j} - W_i \max_j C_j$$

$$\frac{\partial E(\Pi)}{\partial OP_j} = 0 \Rightarrow$$

$$\frac{\partial W_n}{\partial OP_j} - E\left[P_j(T)\right]\frac{w_j}{W_n}W_n\frac{-1}{OP_j^2} - E\left[P_j(T)\right]\frac{w_j}{W_n}\frac{1}{OP_j}\frac{\partial W_n}{\partial OP_j}$$

$$- \frac{\partial W_i}{\partial OP_j}\max_j C_j - W_i\frac{\partial C_i}{\partial OP_j} = 0.$$

In the above equation all terms except for the first one are positive. Specifically, $\frac{\partial W_i}{\partial OP_j}$ is negative given that insider demand drops as the price increases, and $\frac{\partial C_i}{\partial OP_j}$ is negative since the option price decreases as the strike price increases (or equivalently as the level of moneyness decreases).

Notes

1. The computational resources (Stevin Supercomputer Infrastructure) and services used in this work were provided by Ghent University, the Hercules Foundation, and the Flemish Government—Department EWI.
2. As we will see in section 4, knowing the marginal density function is equivalent to knowing the volatility of the jump size and the Poisson arrival rate.
3. One way to capture potential value erosion of the option due to other insiders would be to incorporate a dividend yield. According to the theory of options, it is never optimal to exercise an American option before maturity in the absence of dividends. This means that, in our context, insiders would always bet at the last minute. It is the presence of other insiders (dividends) that makes it optimal to bet before maturity.
4. If they bet at OP prices will exhibit a plunge, and therefore betting at MP would be worthless. This is true under the assumption that information regarding a certain horse can be revealed only once.
5. Races in which no plunges are visible in the data (odds at no point fell for any horse during the betting) are races in which inside trades were not observed. Of course, it could be that an insider placed a discreet bet with a single bookmaker and that this bet cannot be discerned in the average odds that rule in the market and are published. The greater the

extent of this phenomenon, the more our estimates of insider trading will underestimate its true extent.

6. The data were obtained from the CD-Rom *Australasian Racing Encyclopedia '98*, presented by John Russell.

References

Dowie, Jack. 1976. On the efficiency and equity of betting markets. *Economica* 43(170): 139–150.

Kahneman, Daniel, and Amos Tversky. 1979. Prospect theory: An analysis of decision under risk. *Econometrica* 47(2):263–292.

Makropoulou, Vasiliki, and Raphael N. Markellos. 2011. Optimal price setting in fixed-odds betting markets under information uncertainty. *Scottish Journal of Political Economy* 58(4):519–536.

Schnytzer, Adi, Martien Lamers, and Vasiliki Makropoulou. 2010. Measuring the extent of inside trading in a horse betting market. *Journal of Gambling and Business Economics* 4(2):21–41.

Schnytzer, Adi, and Yuval Shilony. 2002. On the timing of inside trades in a betting market. *European Journal of Finance* 8(2):176–186.

Shin, Hyun Song. 1991. Optimal betting odds against insider traders. *Economic Journal* 101(408):1179–1185.

——. 1992. Prices of state contingent claims with insider traders, and the favourite-longshot bias. *Economic Journal* 102(411):426–435.

——. 1993. Measuring the incidence of insider trading in a market for state-contingent claims. *Economic Journal* 103(42):1141–1153.

Quandt, Richard E. 1986. Betting and equilibrium. *Quarterly Journal of Economics* 101(1): 201–207.

SECTION IV

BETTING STRATEGY

This page intentionally left blank

CHAPTER 19

BETTING ON SIMULTANEOUS EVENTS AND ACCUMULATOR GAMBLES

ANDREW GRANT

1 Introduction

As bettors become more sophisticated, there is an increasing focus on bankroll risk management or how much to bet on certain opportunities. Bettors often face situations where multiple games are played simultaneously, and therefore the problem of allocating capital across different games or events must be considered, much the same as an investor allocating capital to different stocks in a portfolio. In this chapter, I explore the use of accumulator bets (parlays) as part of a portfolio betting strategy. Accumulators are a security offered by bookmakers (as opposed to casino games) that pay out on the joint outcome of multiple games. The optimal use of accumulators allows the bettor to replicate payoffs from sequential betting (with proportional strategies) by betting simultaneously (Grant, Johnstone, and Kwon 2008).

Using fixed-stake (e.g., £10 per outcome) betting strategies, betting simultaneously will not pose a problem as long as there are fewer stakes to be made than the bankroll permits. A proportional-stake betting system, as used by many sophisticated bettors, requires the bettor to wager a fixed proportion (e.g., 5%) of their bankroll at each betting opportunity. The most widely discussed proportional staking strategy with respect to fixed odds betting is the log-utility or Kelly (1956) strategy and its fractional variants (Maclean, Ziemba, and Blazenko, 1992), which approximate other power utility functions.

Proportional systems, as the name implies, require the bettor to wager some fraction, f, of their total wealth, W, at each point in time. In particular the Kelly (1956) criterion requires the bettor to risk the fraction, f, at each opportunity to maximize the logarithmic utility function, $U(W) = \ln(W)$. Bettors employing fractional variants of the Kelly criterion will wager some fraction λf, $0 \leq \lambda \leq 1$, of the "full" ($\lambda = 1$) Kelly fraction, which increases the security of the bankroll at the cost of reducing the bankroll growth. Fractional Kelly strategies represent the use of the power utility function $U(W) = \frac{1}{\theta} W^{\theta}$ for $\theta < 0$ and will see the bettor wager less than the full Kelly stake. Under the power utility function $\theta \to 0$ is the special case of log utility. The following work discusses the full-Kelly criterion only but can be easily applied to fractional Kelly strategies.

In this chapter I review the literature on betting on simultaneous games with particular focus on the Kelly (1956) criterion. In section 2 I discuss the general problem of finding the optimal amount to bet by considering the simultaneous tossing of identical, independent, biased coins with payoffs at even odds. I find characteristic functions for the tossing of up to 7 simultaneous coins and review some approximations to the optimal bet size. In section 3 I find analytical solutions to the optimal betting strategy for the special case of 2 games with symmetric payouts (the two games have the same payouts but not necessarily even odds). Section 3 discusses numerical techniques for finding the optimal amount to bet when there are a general number of games available and payoffs possible. In section 4 I provide a brief discussion of the relative merits of using accumulators over single-game bets and present an empirical comparison between Kelly betting strategies involving no accumulators, maximal level accumulators, and the Optimal M strategy of , Johnstone, and Kwon (2008). Section 6 discusses practical matters related to betting on simultaneous games and avenues for future research opportunities.

Readers interested in fixed-staking strategies may like to focus on the number of bets required to employ various approaches to betting. I conclude with a discussion of potential future research directions for simultaneous staking strategies.

2 Betting on the Simultaneous Tossing of Biased Coins

Consider the case of a single favorable coin, which pays out even odds. Refer to the gross payout per dollar bet as α (\$2 in this case per \$1 bet) and the net payout as β (\$1 per \$1 bet). Let the bettor hold probability $p > 0.5$ of heads occurring (and hence $q = 1 - p$) of tails. According to the Kelly criterion, the optimal bet is that which maximizes the growth rate, G, of the bettor's bankroll

$$G = p \ln(1 + f) + q \ln(1 - f). \tag{19.1}$$

The Kelly bettor hence wages a fraction $f^* = p - q = 2p - 1$ on the outcome heads. For example, if $p = 0.60$, then the Kelly bettor would wager $f^* = 20\%$ of his or her bankroll on the outcome heads. Clearly if five identical independent favorable coins are tossed simultaneously, and the bettor wagers 20 percent of bankroll on each toss, he or she faces the possibility of ruin if all coins happen to land on tails. Thus the proportional bettor needs to consider the fact that the coins are tossed simultaneously and adjust the bet size accordingly. This is best illustrated with the simultaneous tossing of two coins (see also Thorp 2000; Medo, Pis'mak, and Zhang 2008).

Suppose that there are $i = 1, 2, 3 \ldots, M$ independent and identical coins, to be tossed simultaneously. Assume that the payouts on offer are all even money ($\alpha = \$2$) and, because the coins are identical, that the bettor holds an identical probability $p_i = p$ for each of the coins. Thus the optimal bet size will be $f_i^* = f^*$, simplifying the process to a one-variable problem. For M simultaneous coin tosses there is a probability of $(1-p)^M > 0$ that all coins land on tails. It therefore follows that the optimal bet size takes the restriction $f^* < 1/M$. In other words, the simultaneous bettor should never risk more than half his wealth on any particular coin in a two-coin scenario, even if the probability of heads is greater than $p = 0.75$.

In the special case of $p_i = p$ and $\alpha = 2$ for every game, Medo, Pis'mak, and Zhang (2008) provide us with a simple method for finding the optimal wager for the Kelly bettor. Suppose there are w winning tosses and $M - w$ losing tosses. The bettor's return will be $(2w - M)f$ in this case (and will hence increase by a factor of $1 + (2w - M)f)$). For example, if $M = 3$ and $w = 2$, the bettor's bankroll will increase by a factor of $(1 + f)$, which would be economically equivalent to two coins landing on heads and one on tails.

The growth rate of the Kelly bettor's bank roll in the coin-tossing game is given by

$$G = \sum\nolimits_{w=0}^{M} P(w; M, p) \ln[1 + (2w - M)f], \tag{19.2}$$

where $P(w; M, p) = \binom{M}{w} p^w (1-p)^{M-w}$ is the binomial probability distribution function. We find the optimal investment fraction by using the first-order condition $\frac{\partial G}{\partial f} = 0$. Then, rewriting $2w - M = [f(2w - M) + 1 - 1]/f$ and using the normalization of $P(w; M, p)$, we obtain the resultant equation to find the optimal fraction for M independent coins

$$\sum\nolimits_{w=0}^{M} \frac{p(w; M, p)}{1 + (2w - M)f} = 1 \tag{19.3}$$

In this form the case $M = 1$ reduces to $(1 - p)/(1 - f) + p/(1 + f) = 1$, and so $f^* = (2p - 1)$

When $M = 2$, the optimal betting strategy may be found by using the generating function (19.3)

$$\frac{\binom{2}{0}(1-p)^2}{1-2f} + \frac{\binom{2}{1}p(1-p)}{1} + \frac{\binom{2}{2}p^2}{1+2f} = 1.$$

The first term provides the payoff when both coins land tails with probability $(1-p)^2$, and the bettor's bankroll is increased by a factor of $(1 - 2f)$. The second term provides the payoff when one coin lands heads and the other tails with probability $2p\ (1 - p)$ with the coefficient of two indicating order indifference. The bettor's bankroll would increase by a factor of $(1+f-f)=1$. The final term on the left-hand side is the payoff when both coins land heads (with probability p^2), which increases the bettor's bankroll by $(1+2f)$.

$$\frac{(1-p)^2(1+2f)+2p(1-p)(1-2f)(1+2f)+p^2(1-2f)}{(1-2f)(1+2f)}=1,$$

which, after some relatively simple algebra, reduces to

$$f=0 \text{ or } f(-8p^2+8p-4)+4p-2=0.$$

Ignoring the trivial solution we rearrange to obtain

$$f^*=\frac{2p-1}{4p^2-4p+2}=\frac{2p-1}{(2p-1)^2+1} \tag{19.4}$$

It is worth examining this solution more closely. The numerator of the expression is simply the optimal fraction for $M=1$ coin or the single-game Kelly fraction. The expression in the denominator is a parabola, which serves as a reduction factor for the bet from the nonsimultaneous case. The denominator takes a minimum value of 1 at $p=0.5$. For $p>0.5$ (the case of interest) the denominator increases (at an increasing rate) and approaches 2 as $p\to 1$, meaning that the bet size is maxed out at $f^*=1/2$, as required. As the edge gets larger, the impact of the restriction on the bet size becomes more binding; the optimal bet for each coin flip when $p=0.55$ is $f^*=0.1/1.01=0.099$, whereas the optimal bet when $p=0.8$ is $f^*=0.6/1.36=0.441$.

It is possible (but considerably more complicated) to construct an analytical solution for the simultaneous tossing of $M=3$ identical coins. Again the generating function (19.3) is used

$$\frac{\binom{3}{0}(1-p)^3}{1-3f}+\frac{\binom{3}{1}p(1-p)^2}{1-f}+\frac{\binom{3}{2}p^2(1-p)}{1+f}+\frac{\binom{3}{3}p^3}{1+3f}=1.$$

Note that there are three ways of obtaining two heads and one tail (which increases the bankroll by a factor of $(1+f)$ and three ways of obtaining two tails and one head. Expanding out this equation yields

$$\begin{aligned}&(1-P)^3(1-f)(1+f)(1+3f)+3p(1-p)^2(1-3f)(1+f)(1+3f)\\&+3p^2(1-p)(1-3f)(1-f)(1+3f)+3p^3(1-3f)(1+f)(1-f)\\&\quad=(1-3f)(1-f)(1+f)(1+3f),\end{aligned}$$

which we can simplify to

$$\begin{aligned}&(1-p)^3(1-3f-f^2-3f^3)+3p(1-p)^2(1+f-9f^2-9f^3)\\&+3p^2(1-p)(1-f-9f^2+9f^3)+3p^3(1-3f-f^2+3f^3)\\&\quad=1-10f^2+9f^4\end{aligned}$$

After some simplification, and ignoring a trivial solution $f = 0$, we are left with a characteristic cubic function

$$9f^3 + (6(2p-1)^3 - 9(2p-1))f^2 + (-6(2p-1)^2 - 9)f + 3(2p-1) = 0. \quad (19.5)$$

The coefficients of f in this equation are reduced to the form involving the original "edge" on the Kelly bet, $(2p-1)$, or constants. Examining these coefficients will allow us to glean some information about the optimal solution. From this we can clearly see that the coefficients of f^2 and f will be negative while the constant will be positive when $0.5 < p < 1$. We can find an analytical solution to this function using the cubic formula, but we gain little insight into the problem. It is very easily solved using numerical methods in this characteristic format.

When finding the characteristic equations for M simultaneous coin tosses it may useful to consider an adjustment to the generating function (19.3) to reduce the burden in simplification. Letting $X = (2w - M)f$ and using the identity $1/(1+X) = 1 - X/(1-X)$, we can rewrite (19.3) as

$$\sum_{w=0}^{M} \frac{\binom{M}{w} p^w (1-p)^{M-w} (2w-M)}{1+(2w-M)f} = 0. \quad (19.6)$$

The new generating function (19.6) is significantly simpler to work with because of the lack of constant on the right-hand side. When $w = M/2$ (the number of heads and tails is equal) the number of terms in the expression is also reduced by 1.

Using this new compact form, and following a similar process as in the case $M = 3$, we can find the characteristic equation for the $M = 4$ simultaneous coin case

$$\begin{aligned}(-24(2p-1)^4 &+ 48(2p-1)^2 + 40)f^3 + (24(2p-1)^3 - 40(2p-1))f^2 \\ &+ (-12(2p-1)^2 - 28)f + 4(2p-1) = 0 \quad (19.7)\end{aligned}$$

Figure 19.1 presents the optimal single coin betting fractions for $M = 1, 2, 3, 4$ simultaneous coins as found by using the expressions in (19.1), (19.4), (19.6), and (19.7). When the bettor has only a small edge, the optimal betting fractions are not reduced substantially from the single-coin case. As the edge becomes large it can be observed quite clearly that the optimal bet approaches $f^* = 1/M$.

Figure 19.2 shows the total amount bet Mf^* across all coins by bettor probability for $M = 1, 2, 3, 4$ coins. The bettor is constrained by the restriction $Mf^* < 1$ for simultaneous betting. The restriction becomes binding at smaller probability edges as the number of coins increases.

Figure 19.3 presents the expected bankroll growth rate (or expected log-utility) G for $M = 1, 2, 3, 4$ coins. The bettor's wealth is diversified across multiple coins as M increases, and there is a commensurate improvement in their expected bankroll

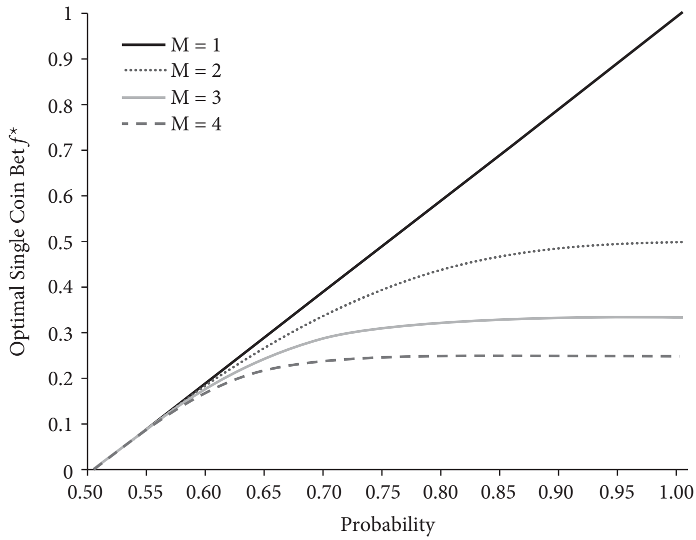


FIGURE 19.1 Optimal individual coin bets for $M = 1, 2, 3, 4$ coins by bettor probability

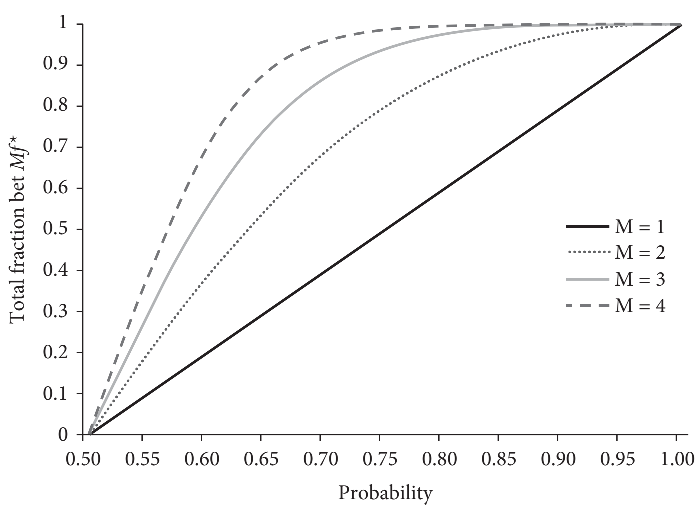


FIGURE 19.2 Total amount bet Mf^* for $M = 1, 2, 3, 4$ simultaneous coin tosses by bettor probability

growth rate. Thus it is advisable to bet optimally on $M > 1$ simultaneous coins than a single coin alone. There are diminishing marginal returns to diversifying one's bet by increasing the number of coins.

It is illustrative to find the characteristic functions for increasing numbers of coins, derived from the generator (19.6). Note that for $M > 4$ numerical methods must be

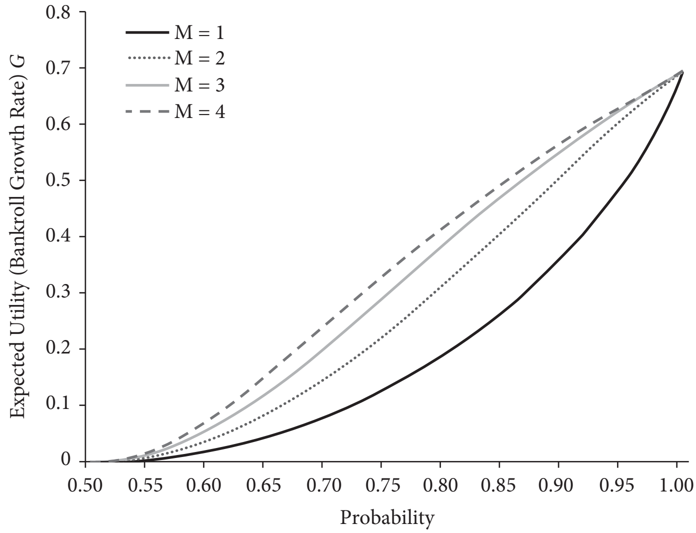


FIGURE 19.3 Expected Utility or bankroll growth rate for $M = 1, 2, 3, 4$ simultaneous coin tosses by bettor probability

used to find the optimal fractions, f^*. The case $M = 5$ has the characteristic function

$$\begin{aligned} &-225f^5 + (120(2p-1)^5 - 300(2p-1)^3 + 225(2p-1)f^4 + (-120(2p-1)^4 \\ &\quad + 260(2p-1)^2 + 110)f^3 + (60(2p-1)^3 - 110(2p-1)f^2 + (-20(2p-1)^2 - 5)f \\ &\quad + 5(2p-1) = 0 \end{aligned} \tag{19.8}$$

The characteristic function for $M = 6$ is

$$\begin{aligned} &(-720(2p-1)^6 + 2160(2p-1)^4 - 2160(2p-1)^2 - 1584)f^5 + (720(2p-1)^5 \\ &\quad - 1920(2p-1)^3 + 1584(2p-1)f^4 + (-360(2p-1)^4 + 840(2p-1)^2 \\ &\quad + 240(2p-1))f^3 + (120(2p-1)^3 - 240(2p-1))f^2 + (-30(2p-1)^2 f \\ &\quad + 6(2p-1) = 0 \end{aligned} \tag{19.9}$$

The characteristic function for the $M = 7$ coin case is

$$\begin{aligned} &11025f^7 + (5040(2p-1)^7 - 17640(2p-1)^5 + 22050(2p-1)^3 - 11025(2p-1)f^6 \\ &\quad + (-5040(2p-1)^6 + 15960(2p-1)^4 - 17178(2p-1)^2 - 6433)f^5 + (2520(2p-1)^5 \\ &\quad - 7140(2p-1)^3 + 6433(2p-1))f^4 + (-840(2p-1)^4 + 2100(2p-1)^2 + 455)f^3 \\ &\quad + (210(2p-1)^3 - 455(2p-1))f^2 + (-42(2p-1)^2 - 7)f + 7(2p-1) = 0 \end{aligned} \tag{19.10}$$

What we can see from this is that there is a common pattern among the coefficients of f. When the number of coins, M, is even the polynomial is of degree $(M-1)$, and when the number of coins is odd the polynomial is of degree M. In either case, note that the coefficient of $f^M = \Pi_{w=0}^{M}(2w-M)$, which will be negative for $M = 5,\ 9,\ 13,\ \ldots$, positive for $M = 3,\ 7,\ 11, \ldots$, and zero for M even. The degree of the polynomial is reduced by one for even M because if *M/2* coins land on heads, the bettor's return will be zero.

It is apparent that in the characteristic format the constant term (i.e., not involving f) is $M(2p-1)$ Each of the coefficients of f then follows a pattern of descending functions of even powers of $(2p-1)$ if the exponent of f is odd or odd powers of $(2p-1)$ if f is even. For example, the coefficient of f involves terms in $(2p-1)^2$ and $(2p-1)^0$, the coefficient of f^2 involves powers of $(2p-1)^3$ and $(2p-1)$, and so on. Therefore, if $=$ 1/2, and hence the bettor does not have an edge, the optimal bet size will be zero.

Moreover, the leading term in each of the coefficient is, beginning from the constant term, increasing by a factor of $-M$ for each additional power of f. For example, in the case M $=$ 7, the leading term in the coefficient of f is $(7 \times -6) = -42$, the leading term in the coefficient of f^2 is $(7 \times -6 \times -5) = 210$, and so on, all the way up to the f^6 term where the coefficient is $7! = 5040$. Hence even powers in f will have a positive leading coefficient while odd powers in f will have negative leading coefficients. The non-leading terms in the coefficients of f can be found mechanically through the comparison of powers with the direct expansion.

Medo, Pis'mak, and Zhang (2008) discussed methods for finding approximate solutions to the simultaneous coin-tossing problem. They found approximate solutions for an "unsaturated" portfolio, where the constraint of betting less than the total bankroll over all coins does not bind, ($Mf^* << 1$), and the "saturated" portfolio, where the total fraction constraint is binding ($1 - Mf^* << 1$). The approximate solution, (exact for $M = 1, 2$), for the unsaturated portfolio is

$$f^* = \frac{2p-1}{M(2p-1)^2 + 4p(1-p)}, \tag{19.11}$$

and the approximate solution for the saturated case is

$$f^* = \frac{1}{M}\left[1 - \frac{2p(1-p)^M}{2p-1}\right]. \tag{19.12}$$

The first approximation (19.11) should be used in the range $p \in \left[\frac{1}{2}, p_c\right]$ while the second (19.12) should be used in the range $p \in (p_c, 1]$. The critical value of the probability, p_c is at the point where the two approximations intersect. These approximations work well in practice for most values of p. Either there is a small reduction from the single game betting fraction or the bettor's optimal strategy is close to $f^* = 1/M$. The worst performance is typically around $p = p_c$. Figure 19.4 presents a comparison of the actual solution and its approximation for the $M = 3$ coin case.

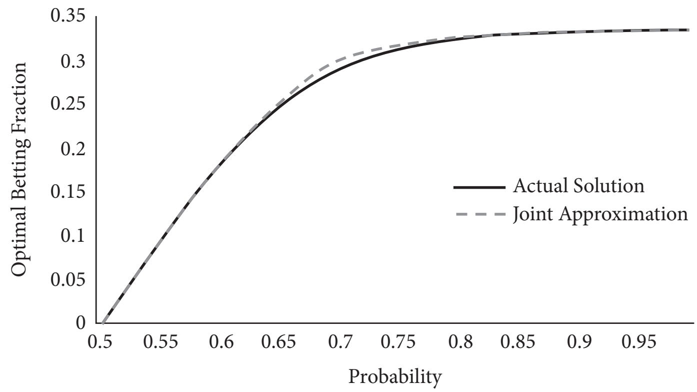


FIGURE 19.4 Optimal betting fraction and approximate solution by probability for $M = 3$ simultaneous coin tosses

3 Simultaneous Betting on Two Games with Symmetric Generalized Payouts

Up to this point I have considered only the special case of betting on identical biased independent coins, which meant that the payoff was of the "double or nothing" type at even odds and the bettor's subjective probability (or the coins' known mechanical probabilities) was identical in every case. In the more realistic case of betting with non-identical probabilities and more generalized payoffs, the problem of finding the optimal amount to bet changes in nature.

Edward Thorp (2000) presented the solution to the two-game, two-outcome case for which the Kelly bettor does not require identical probabilities (p_1 is not necessarily equal to p_2), but the payoff is restricted to even odds ($\alpha = \$2$). Here we follow the methodology but extend the problem to find the optimal amount to bet on two simultaneous two-outcome games, with non-identical probabilities and symmetric payouts ($\alpha_1 = \alpha_2 = \alpha \neq 2$).

The Kelly criterion for the two-simultaneous, two-outcome game maximizes, with respect to the Kelly fractions f_1, f_2, the expected log-utility

$$U(f_1,f_2) = p_1p_2\ln(b+\alpha f_1+\alpha f_2) + p_1q_2\ln(b+\alpha f_1) + q1p2\ln(b+\alpha f2) + q1q2\ln(b), \tag{19.13}$$

where $b = 1 - f_1 - f_2$ is the fraction not bet and α is the (assumed) common bookies payout for the two outcomes.

Figure 19.5 shows the contours of $U(f_1,f_2)$ for two cases ($\alpha > 2$ and $1 < \alpha < 2$) in the permissible domain: $f_i \geq 0, f_1 + f_2 \leq 1$. Specifically, the cases are $\alpha = [2.3, 1.8]$, $p_1 = [0.5217, 0.6667]; p_2 = [0.6087, 0.7778]$.

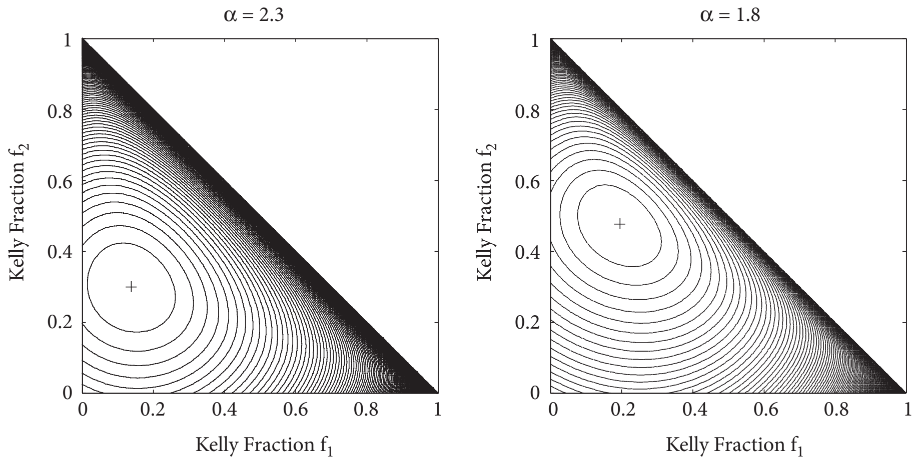


FIGURE 19.5 Growth function from equation (19.13) for two simultaneous betting opportunities. Left panel, $\alpha > 2$; Right panel, $1 < a < 2$.

A single clearly defined maximum is seen in both situations. The analysis that follows seeks to find this maximum analytically.

The master equation. Standard differential methods show that the optimal Kelly fractions are related to the "master" equation

$$\frac{k_1}{a_1 + b_1 f} + \frac{k_2}{a_2 + b_2 f} + \frac{k_3}{a_3 + b_3 f} = 0, \tag{19.14}$$

where the coefficients are given by

$$\begin{aligned} a_1 &= 1 + \beta c; \quad & b_1 &= \beta(1 + d); \quad & k_1 &= \beta p_1 p_2 \\ a_2 &= 1 - c; & b_2 &= \beta - d; & k_2 &= (\beta - 1) p_1 q_2 \\ q_3 &= 1 - c; & b_3 &= -(1 + d); & k_3 &= -q_1 q_2. \end{aligned}$$

The parameter $\beta = \alpha - 1$ (the odds against) and the constants c and d are defined by

$$c = \frac{q_1 p_2 - p_1 q_2}{\beta p_1 q_2 + q_1 p_2} \quad \text{and} \quad d = \frac{p_1 q_2 + \beta q_1 p_2}{\beta q_1 p_2 + q_1 p_2}.$$

Once f is determined from (19.14) the two optimal Kelly fractions are given respectively by

$$f_1 = f \quad \text{and} \quad f_2 = c + df.$$

The task before us therefore is to find solutions for f_i in the range $[0, 1]$ as functions of the parameters p_i and β.

The master quadratic. Observe that the master equation is actually equivalent to the quadratic

$$Af^2 + Bf + C = 0, \tag{19.15}$$

where

$$\left.\begin{aligned} A &= k_1 b_2 b_3 + k_2 b_1 b_3 + k_3 b_1 b_2 \\ B &= k_1(a_2 b_3 + a_3 b_2) + k_2(a_1 b_3 + a_3 b_1) + k_3(a_1 b_2 + a_2 b_1) \\ C &= k_1 a_2 a_3 + k_2 a_1 a_3 + k_3 a_1 a_2 \end{aligned}\right\}. \tag{19.16}$$

The discriminant $\Delta = B^2 - 4AC$ of the quadratic (19.15) can be written in the form

$$\Delta = X^2 + Y^2 + Z^2 - 2(XY + YZ + ZX),$$

where

$$X = k_1(a_2 b_3 - a_3 b_2); \quad Y = k_2(a_3 b_1 - a_1 b_3); \quad Z = k_3(a_1 b_2 - a_2 b_1).$$

Let **v** denote the vector (X, Y, Z)', then $\Delta =$ **v**'Q**v** where Q is the matrix

$$Q = \begin{bmatrix} 1 & -1 & -1 \\ -1 & 1 & -1 \\ -1 & -1 & 1 \end{bmatrix}.$$

Since Q has eigenvalues (-1, 2, 2) the quadratic form $\Delta =$ **v**'Q**v** is indefinite, which means that the discriminant Δ may be positive, negative, or zero. In practice, however, for the range of parameters involved, Δ is always positive.

The coefficients A, B, C. The coefficients A, B, C defined by (19.16) can be reduced to functions of the parameters p_1, p_2, and β. After some rather tedious algebra, and canceling the common factor

$$\frac{(\beta+1)^2 p_1 q_2}{(\beta p_1 q_2 + q_1 p_2)^2},$$

we obtain

$$A = -\beta(\beta - 1)D, \tag{19.17}$$

$$B = \beta(\beta-1)p_1 q_2(D + p_1 p_2) - (\beta - 1)q_1 p_2(D + q_1 q_2) - \beta D(p_1 p_2 + q_1 q_2), \tag{19.18}$$

$$C = p_2 q_2\left[(\beta+1)p_1 - 1\right], \tag{19.19}$$

with

$$D = p_1 q_2 + q_1 p_2 = p_1 + p_2 - 2p_1 p_2. \tag{19.20}$$

Two things immediately follow from these expressions. The first is that A is positive, negative, or zero according to β being respectively less than, greater than, and equal to one. The second is that since the subjective odds and bookies payout must in practice satisfy the condition

$$\alpha p_i = (\beta + 1)p_i > 1; \ (i = 1, 2),$$

C is always positive. Determining the sign of B is considerably more troublesome.

Since the product of the roots of (19.16) is C/A, we can conclude that the two roots (assuming them to be real) have opposite signs if $\beta < 1$ and have the same sign if $\beta > 1$. Of course if $\beta = 1$ we have the special case of even odds. In this situation $A = 0$ and hence the quadratic equation reduces to a linear equation with the single root $f = -C/B$. For this case we find that the corresponding Kelly fractions are

$$f_1 = \frac{p_2 q_2 (p_1 - q_2)}{D(1 - D)}, \quad f_2 = \frac{p_1 q_1 (p_2 - q_2)}{D(1 - D)}, \tag{19.21}$$

which are both positive, as $p_i > q_i$ since $p_i > 0.5$, and $D > 0$ since $p_i < 1$. The expression further reduces to the special coin-tossing example when $p_1 = p_2$, in which case

$$f = \frac{2p - 1}{(2p - 1)^2 + 1}$$

as obtained earlier.

Summary. Let us write the two roots of the master quadratic (19.16) as

$$f^{\pm} = \frac{-B \pm \Delta}{2A}; \quad \Delta = \sqrt{B^2 - 4AC}.$$

Then we are able to show the following:

Case $\alpha > 2$ ($\beta > 1$).

Here $A < 0, C > 0$ and $\Delta > |B|$. Note that B may be positive or negative. The roots $f^{\pm}$ have opposite signs with $f^{-} > 0$ and $f^{+} < 0$.

So the only feasible root is

$$f^{-} = \frac{\Delta + B}{2\,|A|},$$

and thus the optimal betting strategy f_1, f_2 is

$$f_1 = \frac{\Delta + B}{2\,|A|}, \ f_2 = c + d f_1 = \frac{q_1 p_2 - p_1 q_2}{\beta p_1 q_2 + q_1 p_2} + \frac{(p_1 q_2 + \beta q_1 p_2) f_1}{\beta q_1 p_2 + q_1 p_2} \tag{19.22}$$

Case $\alpha < 2 (\beta < 1)$.

Here $A > 0,\ C > 0,\ B < 0$ and $\Delta > |B|$. Both roots $f^{\pm}$ are positive with $f^{-} < f^{+}$. For all practical values of parameters involved in the Kelly betting strategy, $0 < f^{-} < 1$ and $f^{+} > 1$. In every case we need only take the single root $f = f^{-}$ of the master quadratic, and then the optimal betting strategy f_1, f_2 is

$$f_1 = \frac{\Delta + B}{2A},\ f_2 = c + df_1 = \frac{q_1 p_2 - p_1 q_2}{\beta p_1 q_2 + q_1 p_2} + \frac{(p_1 q_2 + \beta q_1 p_2) f_1}{\beta p_1 q_2 + q_1 p_2}. \tag{19.23}$$

4 Numerical Methods for Finding the Optimal Strategy

Finding the optimal bet size with a generalized problem is quite involved. The solution to the general problem for a two-game case with nonsymmetric payouts becomes intractable, as does any nonidentical coin-tossing problem for three or more games. First we will introduce some notation for the generalized simultaneous betting case.

Denote by $p_{i,j}$ the bettor's subjective probability of outcome j in game i, where $1 \le i \le n$ and $1 \le j \le m$. Let $a_{i,j}$ denote the corresponding bookmaker's or market payout, and let $f_{i,j}$ be the fraction of the bettor's wealth wagered on the game/outcome pair (i, j). We further denote by bi the fraction of wealth not wagered on game i, so that $b_i = 1 - \sum_{j=1}^{m} f_{i,j}$, and by b the total fraction not bet on all simultaneous games, so that $b = 1 - \sum_{i=1}^{m} \sum_{j=1}^{m} f_{i,j}$.

Let $S = \{j_1, j_2, \ldots, j_n\}$ denote a sequence of outcomes for games $\{1, 2, \ldots, n\}$ Over all possible outcomes the set S has dimension $\dim(S) = m^n$. Assuming that the games are independent, the (subjective) probability of obtaining this sequence of outcomes is given by

$$p(S) = p_{1,j1}\, p_{2,j2} \cdots p_{n,jn}\,. \tag{19.24}$$

Obviously we require that $\sum_{j=1}^{m} p_j = 1$ for all $i \in \{1, 2, \ldots, n\}$. We may then write the corresponding bettor's return as

$$R(S) = b + \sum\nolimits_{i=1}^{n} \alpha_{i,j_i} f_{i,j_i}, \tag{19.25}$$

where $b = 1 - \sum_{i=1}^{m} \sum_{j=1}^{m} f_{i,j}$ is the total fraction not wagered over all games and outcomes. The objective function a Kelly bettor maximizes is given by

$$U = E\{\log R(S)\} = \sum_{S} p(S) \log R(S). \tag{19.26}$$

Note that while each return $R(S)$ involves a sum over the n simultaneous games, the objective function U involves a sum over the m^n possible game outcomes in set S, where m is fixed across all games in S. This outer sum may include up to 3^n terms when there are n simultaneous soccer games and the bettor can wager on home win, draw, or away win (1×2 betting). To ensure non-negativity in the fractions and to incorporate the bankroll constraint, the optimization must be performed subject to the constraints $0 \leq f_{i,j} \leq 1$ and $b > 0$.

In the general setting it is more effective to use numerical rather than analytical methods to find the optimal bet size for more than two simultaneous games with generalized probabilities and payouts.

The contour diagrams of figure 19.2 present visualizations of the function to be maximized in the two-game case. Leo Breiman (1961) showed that the Kelly growth function has a global optimum from the properties of concavity of the logarithmic function: positive first partial derivatives and negative second partial derivatives provided the bettor has an edge in each game. This optimum may not be unique in the case of multiple-outcome simultaneous games (that is, there may be more than one set of f's that produce the maximum growth), but all the global optima will share the same value (Whitrow 2007). The maximum will be at an interior point because the bettor will never want to risk all his or her capital (b>0).

The numerical solution to the problem can be solved directly for a relatively small number of games (practically, less than 10) using iterative hill-climbing algorithms, such as the conjugate-gradient or active-set algorithms, which are used in the empirical analysis of Andrew Grant and Peter Buchen (2012). Robin Insley, Lucia Mok, and Tim Swartz (2004) proposed a derivative-based method for finding the optimal Kelly betting strategy based on a modified version of the Gauss-Seidel algorithm. Because the solution surface is typically flat in the neighborhood of the optimum (as seen in figure 19.5) the methods based on derivatives or iterative procedures may oscillate around the maximum value if the successive step sizes are not sufficiently small in magnitude.

The number of terms in the objective function (and derivative) increases exponentially with the number of games. Iterative numerical methods, such as the active-set algorithm (which is used by Matlab's fmincon routine), require successive calculations of these functions with a very large number of terms, which can make the computation very time-consuming. Because the solution surface is flat, it is practically useful to set the tolerance on the function maximum to be relatively low; optimization routines will often spend large amounts of time increasing the bettor's capital growth by an infinitesimal amount. It is imperative to consider that many betting websites have a minimum stake size, and as such increasing the accuracy of the betting fractions is likely unnecessary. Moreover, the precision of the optimal solution is likely to exceed that of the bettor's probabilities in the neighborhood of the maximum.

As the size of set of possible game outcomes S becomes large, either due to multiple-outcome bets or an increased number of games, it is efficient to use the Monte Carlo techniques of Chris Whitrow (2007) to find the optimal betting strategy. Because the solution is nonunique and the solution surface is quite flat, the Monte Carlo approach is able to find an approximate solution in the near neighborhood of the actual numerical solution quite efficiently by simulating game outcomes to estimate the objective function.

5 Optimal Betting When the Bookmaker Accepts Accumulator Gambles

Up to this point it has been assumed that the bettor's universe of possible stakes is limited to the outcomes of single games only. In practice it is possible to bet on the joint outcome of multiple games by using accumulator bets (also known as parlays). This allows the bettor to multiply his or her gross winnings in the case of a joint event occurring.

The bookmaker will typically offer these types of bets for uncorrelated events only. For example, if there were two football matches, with the two home teams paying $\alpha_1 = \$1.50$ and $\alpha_2 = \$1.80$, respectively, the bettor could place a wager that would pay off conditioned on both home teams winning, with a gross payout of $\alpha_1 \times \alpha_2 = 1.50 \times 1.80 = \$2.70$. The number of joint bets in an accumulator is usually referred to as the number of "legs" or "ways" or simply as an "n-fold" or "n-tuple" (e.g., a three-way accumulator or a quadruple).

Tim Kuypers (2000) mentioned that this particular strategy is quite popular in the United Kingdom for soccer betting because historically bettors could not bet on the outcome of a single football match but were restricted to betting on the joint outcome of at least three simultaneous games. This law has since been rescinded, but accumulators remain a popular instrument in betting markets: this may be in part a preference for skewness (Golec and Tamarkin 1998) or a means to prolong the consumption utility of the bet.

In the modern era of online bookmakers, accumulators are often used as incentive for new bettors to join websites or as part of special promotions. For example, the U.K.-based bookmaker Sportingbet offered a promotion in 2011—a bettor opening an account and depositing at least £10 was offered up to £100 of free bets consisting of matched bets of up to £10 on single game bets and £30 each of matched bets on doubles, triples, and quadruples or higher. Irish bookmaker Paddy Power held a promotion whereby bettors were offered a refund for losing four-plus-leg football accumulators if all bets were winning at halftime. Bet365 provided bonuses to a bettor's payout on successful football accumulator bets—winning three-way accumulators received a bonus 5 percent return, with increasing bonus returns (up to 100%) to accumulators with a higher number of legs.

It is interesting to note that many bookmakers use accumulators as a loss leader to generate revenue either through website advertising or, more likely, through increased market share of turnover. Accumulators are relatively profitable for the bookmakers, on average. This is because the bookmaker's transaction costs, as measured by the implicit overround, are multiplied for accumulators (see Ali 1979; Kuypers 2000). As a simple example, suppose that the bookmaker offers equal payouts on heads (H) and tails (T) of $\alpha_H = \alpha_T = \$1.90$, respectively, on a fair coin-tossing game. Observe that the overround, ε_1, for a single game bet is

$$\varepsilon_1 = 1/\alpha_H + 1/\alpha_T = 1/1.90 + 1/1.90 - 1 = 5.26\%. \tag{19.27}$$

Now, if the bookmaker accepts bets on the set of two-way accumulators $\{(HH), (HT), (TH), (TT)\}$ which each occur with probability 0.25, the bookmaker will offer payouts of $\alpha_{HH} = \alpha_{HT} = \alpha_{TH} = \alpha_{TT} = 1.90 \times 1.90 = \$3.61$. The set of doubles thus has overround ε_2

$$\varepsilon_2 = \frac{1}{\alpha_{HH}} + \frac{1}{\alpha_{HT}} + \frac{1}{\alpha_{TH}} + \frac{1}{\alpha_{TT}} = 4\left(\frac{1}{3.61}\right) - 1 = 10.80\% = (1+\varepsilon_1)^2 - 1 \tag{19.28}$$

In general, the bookmaker's overround will grow exponentially with the number of legs in the accumulator. It is therefore less costly for bookmakers to offer free bets in the form of accumulators and is potentially beneficial if it encourages bettors to use accumulators in future betting. Bookmakers do, however, face an increased level of difficulty in balancing their books or hedging risks when they accept these potentially unique, high-liability bets.[1]

5.1 Notation for Accumulators

Before we analyze the optimal use of accumulators in general, we introduce the following notation (adapted from Buchen and Grant 2012).

For each $1 \leq k \leq n$, define the set of k-leg accumulators by

$$M_k = \left\{\left(i_1, j_1\right), \left(i_2, j_2\right), \ldots, \left(i_k, j_k\right)\right\}, \tag{19.29}$$

where each $\left(i_s, j_s\right)$ denotes a specific (unique) game/outcome pair. This means that $i_s = i_t$ only if $s = t$, so that accumulators are made across games but not within games. The collection of all accumulators is denoted by the set-product $M = \coprod_{k=1}^{n} M_k$.

5.2 Optimal Betting with Accumulators

How might a Kelly bettor go about utilizing accumulators in a simultaneous betting strategy? Suppose there are n simultaneous games available to the bettor. An obvious strategy is to use bets of maximal level only, so the bettor would use three-way accumulators only for three simultaneous games. Alternatively, the bettor can use accumulator bets of all k-levels up to and including n, a strategy first discussed in Grant, Johnstone, and Kwon (2008). Thus the bettor would use triples, doubles, and single-game bets to wager on three simultaneous games. Thus the bettor would use triples, doubles, and single-game bets to bet on three simultaneous games.

5.3 Using Maximal Level Accumulators

If the bettor uses n-way accumulator bets only, then each bet is mutually exclusive (only one bet may pay off), and hence this strategy will work under the original Kelly (1956) framework. This strategy was discussed in Grant and Buchen (2012) and is as follows:

Let p_j be the bettor's subjective probability of outcome j in the set M_n of maximal accumulators, and let a_j be the corresponding bookmaker payout. Order the outcomes so that

$$p_1\alpha_1 \geq p_2\alpha_2 \geq \cdots \geq p_m\alpha_m. \tag{19.30}$$

Then let $k \in 1, 2, \ldots, m$ be the maximum integer with properties

$$\sigma_k = \sum\nolimits_{i=1}^{k} \frac{1}{\alpha_j} < 1 \text{ and } p_k\alpha_k > \frac{1-\pi_k}{1-\sigma_k}, \tag{19.31}$$

where $\pi_k = \sum_{i=1}^{k} p_k$. Then the optimal Kelly betting fractions f_j are given explicitly by the formula

$$f_j = \max[p_j - \frac{1}{\alpha_j}\left(\frac{1-\pi_k}{1-\sigma_k}\right), 0]. \tag{19.32}$$

In particular, when the outcomes are sorted as in (19.30), it is optimal to bet only on the first k outcomes.

Let's return to the example of two coins, each with payouts of $\alpha_H = \alpha_T = \$1.90$. Suppose that a bettor has subjective probability of heads on the first coin be $p_{(1,H)} = 0.55$ and heads on the second coin be $p_{(2,H)} = 0.64$. Each of the corresponding accumulators will have a joint payout of $\alpha_j = \$3.61$.

Following that strategy we order the payouts by expectation so that we have

$$p_{(1,H)} \times p_{(2,H)} \times \alpha_{HH} = 0.55 \times 0.64 \times 3.61 = 1.2707$$
$$p_{(1,T)} \times p_{(2,H)} \times \alpha_{TH} = 0.45 \times 0.64 \times 3.61 = 1.0397$$
$$p_{(1,H)} \times p_{(2,T)} \times \alpha_{HT} = 0.55 \times 0.36 \times 3.61 = 0.7149$$
$$p_{(1,T)} \times p_{(2,T)} \times \alpha_{TT} = 0.45 \times 0.36 \times 3.61 = 0.5848$$

Then find the payout reciprocals, $\frac{1}{\alpha_j} = \frac{1}{3.61} = 0.277$, so we can find that $\sigma = (0.277, 0.554, 0.831, 1.108)$ and so the bettor will never bet on more than three outcomes from the first condition. From the second condition, $\frac{1-\pi_k}{1-\sigma_k} = (0.8963, 0.8072, 0.9587, 0)$, and thus only the first two outcomes ($HH$and TH) provide positive betting fractions due to being greater than their respective expectations. The optimal maximal accumulator strategy is, from (19.32),

$$f_{HH} = 0.1037,\ f_{TH} = 0.0644,\ f_{HT=0},\ f_{TT} = 0.$$

The total amount bet is therefore $f_{HH} + f_{TH} = 0.1037 + 0.0644 = 0.1681$, and hence $b = 1 - 0.1681 = 0.8319$. This bet provides expected utility (bankroll growth rate) of

$$\begin{aligned} E(U) = G = {} & 0.55 \times 0.64 \ln(0.8319 + 0.1037(3.61)) + 0.45 \times 0.36 \ln(0.8319 \\ & + 0.0644(3.61)) + 0.55 \times 0.36 \ln(0.8319) \\ & + 0.45 \times 0.36 \ln(0.8319) = 0.0177. \end{aligned} \tag{19.33}$$

5.4 Using All Levels of Accumulators—The Optimal M Strategy

Grant, Johnstone, and Kwon (2008) developed the optimal M strategy for simultaneous games (see their paper for a detailed proof; this chapter provides a detailed illustration). In the optimal M strategy the Kelly bettor uses all levels of accumulator up to and including the maximal level n to maximize utility. The optimal amount to bet on each accumulator is simply derived from the product of the union and complements of each of the individual game betting fractions, as calculated from (19.32).

Let's return to the example of two coins, each with payouts of $\alpha_H = \alpha_T = \$1.90$. Suppose that a bettor has subjective probability of heads on the first coin be $p_{(1,H)} = 0.55$ and heads on the second coin be $p_{(2,H)} = 0.64$. The optimal bet on the first coin individually would be

$$f_{(1,H)} = p_{(1,H)} - \frac{1 - p_{(1,H)}}{\alpha_{(1,H)} - 1} = 0.55 - \frac{0.45}{0.9} = 0.55 - 0.50 = 0.05,$$

and the corresponding optimal bet for the second coin would be $f_{(2,H)} = 0.64 - \frac{0.36}{0.9} = 0.24$. The corresponding amounts withheld from betting on the coins are $b_1 = 0.95$ and $b_2 = 0.76$. The optimal M strategy involves the placement of a two-leg accumulator bet of $\{HH\}$on $f_{\{(1,H)(2,H)\}} = f_{(1,H)} \times f_{(2,H)} = 0.05 \times 0.24 = 0.012$ and two separate single-game bets of $f_{\{(1,H)\}} = f_{1,H} \times b_2 = 0.05 \times 0.76 = 0.038$ on H on the first coin and $f_{\{(2,H)\}} = f_{(2,H)} \times b_1 = 0.24 \times 0.95 = 0.228$ on H on the second coin. Note that the amount remaining in the bank, b, for this strategy is

$$b = 1 - f_{\{(1,H)(2,H)\}} - f_{\{(1,H)\}} - f_{\{(2,H)\}} = 1 - 0.012 - 0.038 - 0.228 = 0.722,$$

which is, in fact, $b_1 \times b_2 = 0.95 \times 0.76 = 0.722$

It is interesting to see that, had the games been played sequentially, the bettor's expected utility from the construction of the optimal M strategy corresponds exactly to the bettor's utility. Let's consider the two-game example again. The expected utility from betting sequentially may be written as

$$E\left[U\left(f_1,f_2\right)\right]=p_1p_2\ln\left(b_1b_2+\alpha_1f_1b_2+\alpha_2f_2b_1+\alpha_1\alpha_2f_1f_2\right)+p_1q_2\ln\left(b_1b_2+\alpha_1f_1b_2\right)$$
$$+q_1p_2\ln\left(b_1b_2+\alpha_2f_2b_1\right)+q_1q_2\ln(b_1b_2), \quad (19.34)$$

where f_1 and f_2 are the individual game betting fractions. For the two-coin example, the expected utility would be

$$E\left[U\left(f_1,f_2\right)\right]=0.55\times0.64\ln(0.722+1.90(0.05\times0.76)+1.90\times(0.24\times0.95)$$
$$+3.61\times(0.05\times0.24))+0.55\times0.36\ln(0.722+1.90(0.05\times0.76))$$
$$+0.45\times0.64\ln(0.722+1.90(0.95\times0.24))+0.45\times0.36\ln(0.722)$$
$$=0.02749 \quad (19.35)$$

We can compare this solution to the expected utility of betting on two games with symmetric payouts using the analytical solution in (19.23). By calculations, we have $A=0.04374, B=-0.22109, C=0.01037, D=0.486, c=0.19305$ and $d=0.98070$. Solving the quadratic gives the solution $f_1=0.04734$ and $f_2=c+df_1=0.23947$. This gives expected utility (bankroll growth rate) of $U=0.02743$, which is slightly lower than the value obtained from the optimal strategy with accumulators (but higher than the value from using maximal accumulators only).

Figures 19.6 to 19.9 present simple comparisons of the wealth distribution that a bettor would realize from the examples just considered. Figure 19.6 shows the bettor's wealth distribution had the two coins been tossed sequentially. Figure 19.7 shows the wealth distribution for the bettor that uses the optimal M strategy. The bettor's wealth

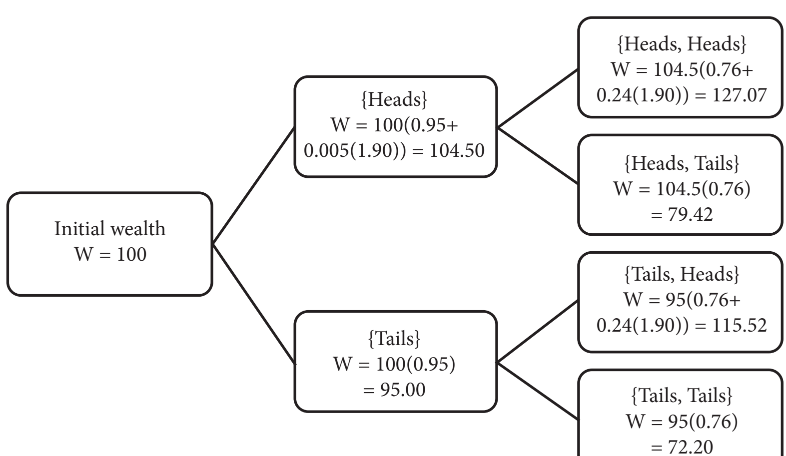


FIGURE 19.6 Outcomes of betting optimally on the tossing of two biased coins, sequential betting; $p_1=0.55$, $p_2=0.64$, $\alpha_1=\alpha_2=1.90$

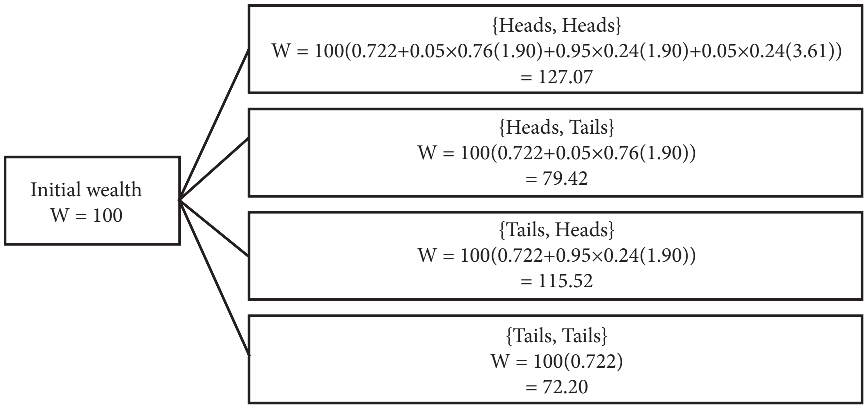


FIGURE 19.7 Outcomes from the optimal use of accumulators to bet on two simultaneous games; $p_1 = 0.55$, $p_2 = 0.64$, $\alpha_1 = \alpha_2 = 1.90$

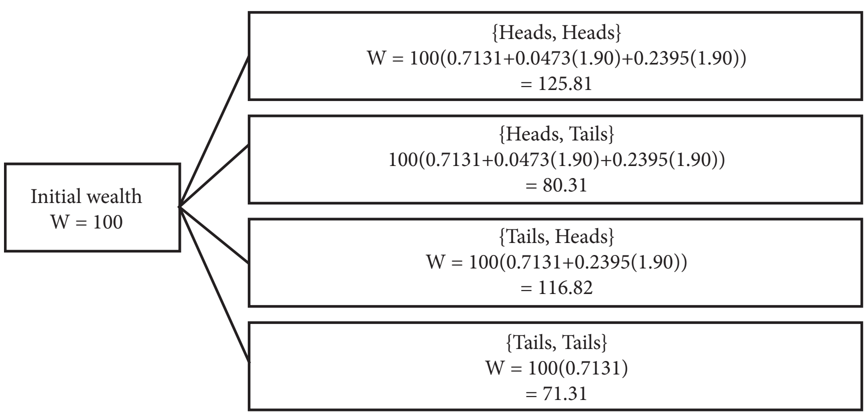


FIGURE 19.8 Outcomes from the optimal betting strategy without accumulators; $p_1 = 0.55$, $p_2 = 0.64$, $\alpha_1 = \alpha_2 = 1.90$

in each of the four states is the same if they bet optimally using sequential bets or bet simultaneously using the optimal M strategy. Figures 19.8 and 19.9 present the analogues for the simultaneous betting strategy without accumulators and using only two-way accumulators.

5.5 Comparing Typical Distributions from Simultaneous Kelly Betting Strategies

The three approaches to simultaneous Kelly betting produce quite different expected return distributions for the bettor. Expanding the example to actual betting markets

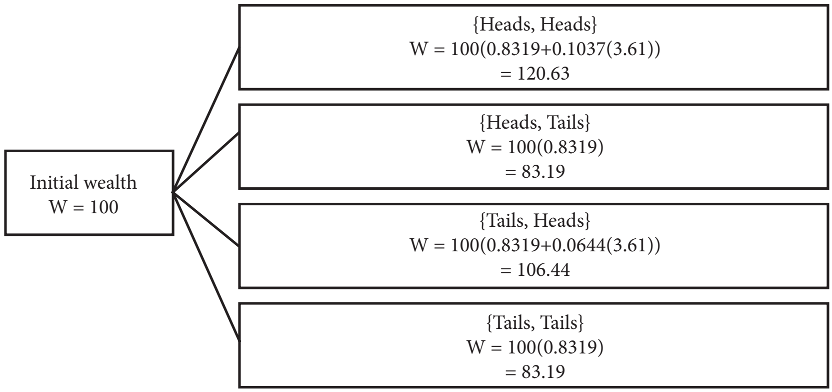


FIGURE 19.9 Using two-way accumulators only to bet optimally simultaneous coin tosses; $p_1 = 0.55$, $p_2 = 0.64$, $\alpha_1 = \alpha_2 = 1.90$

and odds will help explore the differences. Suppose the bettor has the opportunity to bet on five simultaneous English Premier League games, as shown in table 19.1.

Each cell in table 19.1 contains the bookmaker's payouts, the bettor's assumed subjective probability, and the optimal individual game bet for, respectively, the home team,

Table 19.1 Bookmaker 1X2 Payouts, Bettor Subjective Probabilities, and Individual Game Betting Fractions for Five Simultaneous English Premier League Matches

Home Team	Away Team	Home Payoff / Prob. Home / $f^*_{i,1}$	Draw Payoff / Prob. Draw / $f^*_{i,2}$	Away Payoff / Prob. Away / $f^*_{i,3}$	Bankroll Growth Rate, G_i / Fraction not bet, b_i
Chelsea	Bolton	1.29	5.50	11.00	$G_1 = 0.001815$
		0.80	0.12	0.08	
		0.1103	0	0	$b_1 = 0.889655$
Newcastle	Wolves	1.57	4.00	6.00	$G_2 = 0.008847$
		0.70	0.20	0.10	
		0.1737	0	0	$b_2 = 0.826316$
QPR	Fulham	2.63	3.25	2.75	$G_3 = 0.002137$
		0.35	0.25	0.40	
		0	0	0.0571	$b_3 = 0.942857$
West Brom	Sunderland	2.40	3.25	3.00	$G_4 = 0.0155688$
		0.45	0.35	0.20	
		0.1477	0.1267	0	$b_4 = 0.725582$
Wigan	Aston Villa	2.75	3.25	2.63	$G_5 = 0.058984$
		0.25	0.20	0.55	
		0	0	0.2739	$b_5 = 0.726074$

Note: Matches were played on Feb. 25, 2012. Bookmaker prices are from Bet 365.

Table 19.2 Optimal Non-Accumulator Betting Fractions for the Five-Game Betting Opportunity Presented in Table 19.1

Match	Home	Draw	Away
Chelsea v Bolton	0.088692	0	0
Newcastle v Wolves	0.144736	0	0
QPR v Fulham	0	0	0.047596
West Brom v Sunderland	0.12401	0.107394	0
Wigan v Aston Villa	0	0	0.261187

Note: Sum of betting fractions: 0.7736; expected logarithmic utility: 0.08411; total number of bets made: 6.

draw, and away team. The optimal fractions for each game (individually) are shown in the third row of each cell. The probabilities are such that the bettor would prefer to bet on a single outcome in four of the five games and bet on both West Brom and the draw in the fourth game. The fifth column shows the expected utility G_i (bankroll growth rate) and the fraction of wealth withheld b_i for each individual game i. The sum of the growth rate across the five games is $\sum_{i=1}^{5} G_i = 0.0088033$, which would be the bettor's expected utility if the games were played sequentially rather than simultaneously.

A bettor who does not use accumulators will find the numerical solution using either the Monte Carlo method or the direct optimization method. In this case, using the direct optimization through Matlab's "active-set" algorithm, the optimal non-accumulator strategy (M_1 strategy) is as in table 19.2.

In each case the betting fractions are slightly reduced from the individual game case, as expected. The overall expected utility from the five-game betting opportunity is $E(U_1) = 0.08411$, and the total amount wagered is 77.36 percent of the bettor's bankroll. The bettor makes a total of 6 bets in the strategy without accumulators. Figure 19.10 presents the distribution of potential gross returns that the bettor could realize from following the optimal strategy without accumulators.

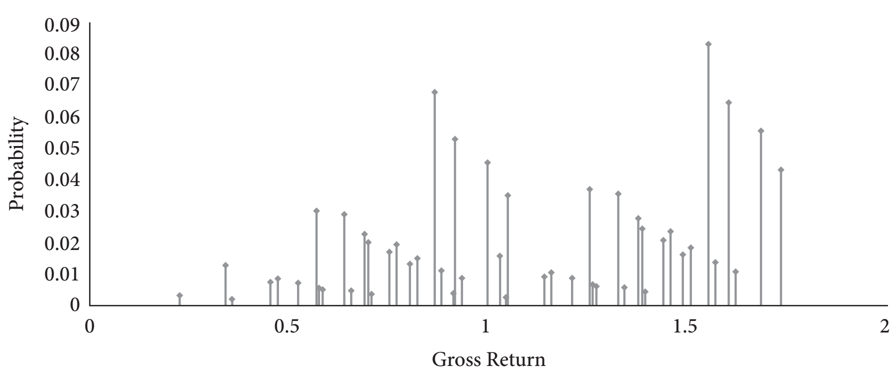


FIGURE 19.10 Distribution of bettor's gross return from employing the optimal non-accumulator betting strategy as shown in table 19.2.

The bettor who uses n-way accumulators (in this case five-way accumulators) only, would place 54 total bets on the $3^5 = 243$ mutually exclusive outcomes. These fractions were computed using the method from the original Kelly (1956) paper. The bets with the highest expectation were those on sets (Chelsea, Newcastle, Fulham, Draw, Aston Villa) and (Chelsea, Newcastle, Fulham, West Brom, Aston Villa), to which the bettor allocated 2.26 percent and 2.8 percent of the bankroll, respectively. The non-zero betting fractions are shown in the table 19.3.

The total amount wagered across the 54 bets was 18.53 percent of the bankroll, and the expected utility was $E(U_n) = 0.050838$. This is substantially lower than the expected utility from single-game only strategy (and in fact lower than simply betting optimally on Aston Villa alone). This is partially due to the multiplicative nature of the transaction costs, as previously discussed. The distribution of return outcomes from employing the five-way only strategy was capped from below because of the relatively low total amount bet, as seen in figure 19.11.

The optimal strategy using all levels of accumulators (the optimal M strategy) is shown in table 19.4. The strategy consists of a total of 47 bets: 2 five-ways, 9 four-ways, 16 triples, 14 doubles, and 6 singles. The total amount bet is 63.48 percent of the bettor's bankroll; 42 percent of the bankroll is invested in the single-game bets and 28 percent of the bankroll is in the two-way accumulators. The expensive, higher level accumulators make up only 5.81 percent of the total wager. This strategy produces expected utility of $E(U) = 0.088033$, which is the same as betting sequentially on the five games as shown earlier.

Comparing the outcomes between the optimal M strategy and the non-accumulator strategy provides us with some interesting insights, as demonstrated by the distribution of gross return differential shown in figures 19.12 and 19.13.

The mean of the distribution of differentials is 0.022, and the median is -0.1825, so the distribution is positively skewed. The optimal strategy without accumulators outperforms the optimal M strategy 60.8 percent of the time, but in these cases, the average outperformance is 3.24 percent. In the cases where the optimal M strategy outperforms, the average outperformance is 5.51 percent. Much of this is made up of the extreme right of the distribution, where either one of the 2 five-way accumulators (and all the corresponding lower level accumulators) or 4 of the 5 preferred bettor outcomes pay off. A small proportion of the outperformance of the optimal M strategy also comes from the bettor's worst-case scenario, where either 1 or 0 of the single-game bets pays off (because the bettor does not risk as much of the bankroll, as explained earlier.)

The intermediate cases, where most of the bets were made (two-ways and three-ways accounted for 30 of the 47 total bets under the optimal M strategy) provide the scenarios where the single-game only strategy tends to outperform. Bettors spend a large proportion of their bankroll on expensive accumulators that do not pay off and, subsequently, would have been better off without them.

Table 19.3 Non-Zero Betting Fractions from Maximal Kelly Betting on the Five-Game Betting Opportunity Presented in Table 19.1

Game 1	Game 2	Game 3	Game 4	Game 5	Probability	Payoff	Expectation	Fraction
Chelsea	Newcastle	Fulham	Draw	Aston Villa	0.0431	$ 47.61	2.052768	0.0225887
Chelsea	Newcastle	Fulham	West Brom	Aston Villa	0.0554	$ 35.16	1.949002	0.0284643
Bolton	Newcastle	Fulham	Draw	Aston Villa	0.0043	$ 405.94	1.750423	0.0019810
Chelsea	Newcastle	QPR	Draw	Aston Villa	0.0377	$ 45.53	1.717794	0.0173482
Bolton	Newcastle	Fulham	West Brom	Aston Villa	0.0055	$ 299.77	1.66194	0.0024573
Chelsea	Newcastle	QPR	West Brom	Aston Villa	0.0485	$ 33.62	1.63096	0.0216876
Chelsea	Newcastle	Draw	Draw	Aston Villa	0.0270	$ 56.26	1.516249	0.0111431
Chelsea	Draw	Fulham	Draw	Aston Villa	0.0123	$ 121.29	1.494281	0.0050352
Bolton	Newcastle	QPR	Draw	Aston Villa	0.0038	$ 388.23	1.464785	0.0015016
Chelsea	Newcastle	Draw	West Brom	Aston Villa	0.0347	$ 41.55	1.439604	0.0138095
Chelsea	Draw	Fulham	West Brom	Aston Villa	0.0158	$ 89.57	1.418746	0.0062560
Bolton	Newcastle	QPR	West Brom	Aston Villa	0.0049	$ 286.69	1.390741	0.0018646
Draw	Newcastle	Fulham	Draw	Aston Villa	0.0065	$ 202.97	1.312817	0.0022633
Bolton	Newcastle	Draw	Draw	Aston Villa	0.0027	$ 479.75	1.292926	0.0009184
Bolton	Draw	Fulham	Draw	Aston Villa	0.0012	$ 1,034.25	1.274193	0.0004084
Chelsea	Draw	QPR	Draw	Aston Villa	0.0108	$ 116.00	1.250441	0.0034732
Draw	Newcastle	Fulham	West Brom	Aston Villa	0.0083	$ 149.89	1.246455	0.0026834
Bolton	Newcastle	Draw	West Brom	Aston Villa	0.0035	$ 354.28	1.227569	0.0010858
Bolton	Draw	Fulham	West Brom	Aston Villa	0.0016	$ 763.75	1.209783	0.0004812
Chelsea	Draw	QPR	West Brom	Aston Villa	0.0139	$ 85.66	1.187232	0.0040867
Chelsea	Wolves	Fulham	Draw	Aston Villa	0.0062	$ 181.93	1.120711	0.0015694
Chelsea	Draw	Draw	Draw	Aston Villa	0.0077	$ 143.34	1.10373	0.0018904
Draw	Newcastle	QPR	Draw	Aston Villa	0.0057	$ 194.11	1.098589	0.0013786
Chelsea	Newcastle	Fulham	Sunderland	Aston Villa	0.0246	$ 43.94	1.082779	0.0059048
Bolton	Draw	QPR	Draw	Aston Villa	0.0011	$ 989.12	1.066268	0.0002460
Chelsea	Wolves	Fulham	West Brom	Aston Villa	0.0079	$ 134.35	1.064059	0.0018126
Chelsea	Draw	Draw	West Brom	Aston Villa	0.0099	$ 105.85	1.047937	0.0021762
Draw	Newcastle	QPR	West Brom	Aston Villa	0.0073	$ 143.35	1.043056	0.0015881
Bolton	Draw	QPR	West Brom	Aston Villa	0.0014	$ 730.42	1.012369	0.0002702
Chelsea	Newcastle	Fulham	Draw	Wigan	0.0196	$ 49.78	0.97565	0.0033193
Draw	Newcastle	Draw	Draw	Aston Villa	0.0040	$ 239.87	0.969694	0.0006680
Draw	Draw	Fulham	Draw	Aston Villa	0.0018	$ 517.12	0.955645	0.0002835
Bolton	Wolves	Fulham	Draw	Aston Villa	0.0006	$ 1,551.37	0.955645	0.0000946
Bolton	Draw	Draw	Draw	Aston Villa	0.0008	$ 1,222.29	0.941165	0.0001083
Chelsea	Wolves	QPR	Draw	Aston Villa	0.0054	$ 173.99	0.937831	0.0007480
Chelsea	Newcastle	Fulham	West Brom	Wigan	0.0252	$ 36.76	0.926332	0.0033608
Bolton	Newcastle	Fulham	Sunderland	Aston Villa	0.0025	$ 374.72	0.9233	0.0003229
Chelsea	Newcastle	Fulham	Draw	Draw	0.0157	$ 58.83	0.922433	0.0020963
Draw	Newcastle	Draw	West Brom	Aston Villa	0.0052	$ 177.14	0.920677	0.0006924
Draw	Draw	Fulham	West Brom	Aston Villa	0.0024	$ 381.88	0.907337	0.0002874
Bolton	Wolves	Fulham	West Brom	Aston Villa	0.0008	$ 1,145.63	0.907337	0.0000959
Chelsea	Newcastle	QPR	Sunderland	Aston Villa	0.0216	$ 42.03	0.906089	0.0026871
Bolton	Draw	Draw	West Brom	Aston Villa	0.0010	$ 902.62	0.89359	0.0001115
Chelsea	Wolves	QPR	West Brom	Aston Villa	0.0069	$ 128.49	0.890424	0.0007683
Chelsea	Newcastle	Fulham	West Brom	Draw	0.0202	$ 43.44	0.875805	0.0020133

(*Continued*)

Table 19.3 (*Continued*)

Game 1	Game 2	Game 3	Game 4	Game 5	Probability	Payoff	Expectation	Fraction
Bolton	Newcastle	Fulham	Draw	Wigan	0.0020	$ 424.46	0.83195	0.0001032
Chelsea	Wolves	Draw	Draw	Aston Villa	0.0039	$ 215.01	0.827798	0.0001858
Chelsea	Newcastle	QPR	Draw	Wigan	0.0172	$ 47.61	0.816442	0.0006239
Chelsea	Newcastle	Draw	Sunderland	Aston Villa	0.0154	$ 51.93	0.79978	0.0002602
Draw	Draw	QPR	Draw	Aston Villa	0.0016	$ 494.56	0.799701	0.0000273
Bolton	Wolves	QPR	Draw	Aston Villa	0.0005	$ 1,483.68	0.799701	0.0000091
Bolton	Newcastle	Fulham	West Brom	Wigan	0.0025	$ 313.45	0.789895	0.0000119
Chelsea	Draw	Fulham	Sunderland	Aston Villa	0.0070	$ 111.96	0.788192	0.0000184
Bolton	Newcastle	Fulham	Draw	Draw	0.0016	$ 501.64	0.786571	0.0000009

Note: Only five-way accumulators were used. Sum of betting fractions: 0.1853; expected logarithmic utility: 0.05838; total number of bets made: 54.

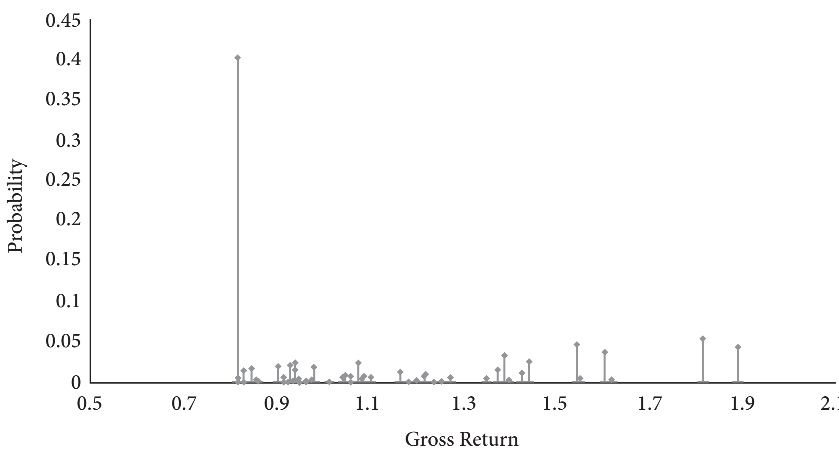


FIGURE 19.11 Distribution of gross returns from using the five-way accumulators only strategy as shown in table 19.3

6 Practical Aspects of Betting on Simultaneous Games

The illustrative five-game example demonstrated quite clearly a number of the potential difficulties of implementing sophisticated Kelly betting strategies across a relatively large number of games.

First, some of the betting fractions produced were very small (such as the amount bet on the five-way accumulators in the optimal M strategy or the amount bet on non-preferred diversifying outcomes in the five-way only strategy), and bookmakers

Table 19.4 Non-Zero betting fractions from Optimal *M* Betting Strategy on Five-Game Betting Opportunity Presented in Table 19.1

Game 1	Game 2	Game 3	Game 4	Game 5	Probability	Payoff	Fraction
Chelsea	Newcastle	Fulham	West Brom	Aston Villa	0.0554	$ 35.16	0.000044
Chelsea	Newcastle	Fulham	Draw	Aston Villa	0.0431	$ 47.61	0.000038
Chelsea	Newcastle	Fulham	West Brom		0.1008	$ 13.37	0.000117
Chelsea	Newcastle	Fulham	Draw		0.0784	$ 18.10	0.000101
Chelsea	Newcastle	Fulham		Aston Villa	0.1232	$ 14.65	0.000218
Chelsea	Newcastle		West Brom	Aston Villa	0.1386	$ 12.78	0.000731
Chelsea	Newcastle		Draw	Aston Villa	0.1078	$ 17.31	0.000627
Chelsea		Fulham	West Brom	Aston Villa	0.0792	$ 22.39	0.000211
Chelsea		Fulham	Draw	Aston Villa	0.0385	$ 35.84	0.000181
	Newcastle	Fulham	West Brom	Aston Villa	0.0693	$ 27.25	0.000357
	Newcastle	Fulham	Draw	Aston Villa	0.0539	$ 36.90	0.000307
Chelsea	Newcastle	Fulham			0.224	$ 5.57	0.00058
Chelsea	Newcastle		West Brom		0.252	$ 4.86	0.00194
Chelsea	Newcastle		Draw		0.196	$ 6.58	0.00166
Chelsea	Newcastle			Aston Villa	0.308	$ 5.33	0.00359
Chelsea		Fulham	West Brom		0.144	$ 8.51	0.00056
Chelsea		Fulham	Draw		0.112	$ 11.53	0.00048
Chelsea		Fulham		Aston Villa	0.176	$ 9.33	0.00104
Chelsea			West Brom	Aston Villa	0.198	$ 8.14	0.00348
Chelsea			Draw	Aston Villa	0.154	$ 11.03	0.00298
	Newcastle	Fulham	West Brom		0.126	$ 10.36	0.00095
	Newcastle	Fulham	Draw		0.098	$ 14.03	0.00081
	Newcastle	Fulham		Aston Villa	0.154	$ 11.36	0.00175
	Newcastle		West Brom	Aston Villa	0.173	$ 9.91	0.00589
	Newcastle		Draw	Aston Villa	0.135	$ 13.42	0.00506
		Fulham	West Brom	Aston Villa	0.099	$ 17.36	0.00170
		Fulham	Draw	Aston Villa	0.077	$ 23.51	0.00146
Chelsea	Newcastle				0.560	$ 2.03	0.00952
Chelsea		Fulham			0.320	$ 3.55	0.00274
Chelsea			West Brom		0.360	$ 3.10	0.00922
Chelsea			Draw		0.280	$ 4.19	0.00791
Chelsea				Aston Villa	0.440	$ 3.39	0.01709
	Newcastle	Fulham			0.280	$ 4.32	0.00465
	Newcastle		West Brom		0.315	$ 3.77	0.01562
	Newcastle		Draw		0.245	$ 5.10	0.01341
	Newcastle			Aston Villa	0.385	$ 4.13	0.02896
		Fulham	West Brom		0.180	$ 6.60	0.00450
		Fulham	Draw		0.140	$ 8.94	0.00387
		Fulham		Aston Villa	0.220	$ 7.23	0.00835
			West Brom	Aston Villa	0.248	$ 6.31	0.02804
			Draw	Aston Villa	0.193	$ 8.55	0.02406
Chelsea					0.80	$ 1.29	0.04529
	Newcastle				0.70	$ 1.57	0.07675
		Fulham			0.40	$ 2.75	0.02213
			West Brom		0.45	$ 2.40	0.07432
			Draw		0.35	$ 3.25	0.06379
				Aston Villa	0.55	$ 2.63	0.13776

Note: Five-way, four-way, three-way, and two-way accumulators plus single game bets were used. Sum of betting fractions: 0.6348; expected logarithmic utility: 0.088038; total number of bets made: 47.

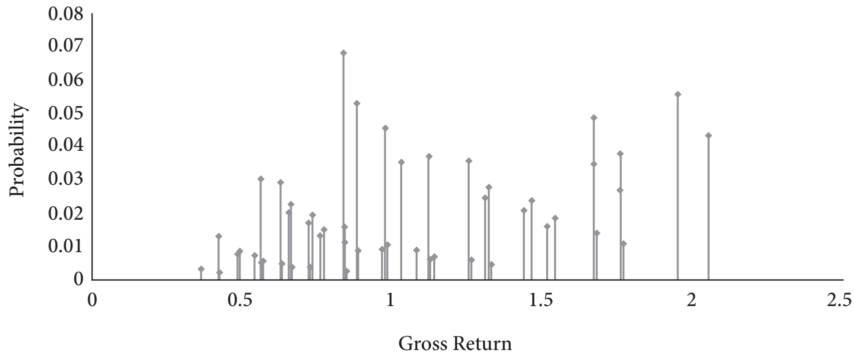


FIGURE 19.12 Distribution of gross returns from employing the optimal M strategy as shown in table 19.4.

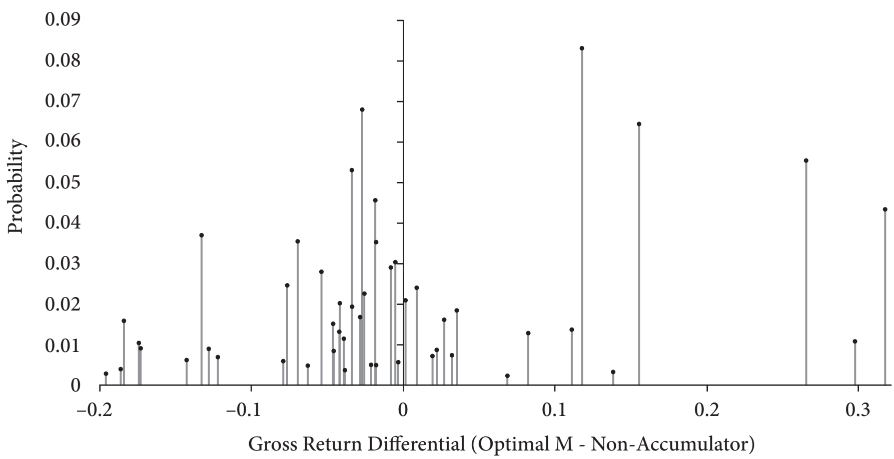


FIGURE 19.13 Distribution of gross return differential between optimal M and non-accumulator betting strategies

typically accept only bets of a minimum size. The impact of non-divisibility of a bettor's capital on betting strategies is an area that requires further exploration. Consideration should be given also to the number of bets that are required to implement a strategy, which is discussed in Grant and Buchen (2012). For a fixed-stake gambler, products exist in many betting markets that place equal-sized bets on, say, a three-way accumulator and the three corresponding two-way accumulators for a total of four equal-sized bets (a Trixie). These products are known as "full cover" bets and are referred to by nicknames related to the maximum-sized accumulator.[2] These full cover bets do not include single-game bets. An interesting project would be to see the relative improvement of the single-game only strategy when augmented with full cover products to minimize the drawback of having to place a large number of bets to implement the optimal M strategy.

Second, we have assumed that prices are fixed for the bettor or that only a single bookmaker is available for the different strategies. Bettors are restricted to the use of a single bookmaker when employing accumulators but not with single-game outcome bets. Hence the advantage of the optimal M strategy might be reduced for a savvy bettor by shopping around for the best odds across a set of bookmakers. Grant and Buchen (2012) compared the impact of a bettor using the optimal M strategy with a single bookmaker and the non-accumulator strategy using the best odds across a set of bookmakers. They found, using simulated results of English Premier League matches, that a bettor who shops around and bets optimally without accumulators can outperform a bettor who takes the average market odds (proxying for the odds offered by a single large bookmaker), provided that the bettor's probability edge is sufficient. A potential downside with shopping around increases the Kelly bettor's stake size, and thus there is an associated increase in risk.

Third, the strategies discussed herein (from both the bettor's and the bookmaker's perspective) consider only simultaneous bets on the outcomes of independent games. Bookmakers very occasionally offer correlated accumulator bets (such as betting on point spreads and over-under totals in the same National Football League game), but these are mainly for promotional purposes or on a limited number of outcomes. Bet365 offers the Scorecast product, which pays on the combination of a limited number of correlated outcomes. For example, Tottenham to win 2–0 pays \$6.50 and Jermain Defoe as first goal scorer pays \$4.50; the Scorecast "combination" of these two events pays \$15.00.

It would be interesting to explore the implications of a bettor's belief that game outcomes were correlated to a different degree from that of the bookmaker. Thorp (2000) discussed betting on simultaneous blackjack hands at the same table, which have a correlation of approximately 0.5 (Griffin 1999) but do not allow accumulators. As one might expect, if outcomes exhibit positive correlation, Kelly bettors should reduce their overall wager relative to the identical case without correlation. It is difficult to estimate correlations across multiple outcomes for non-casino games, and the provision of markets would likely require more sophisticated modeling from the bookmaker's perspective than presently exists. Rainbow derivatives, which are usually call options or put options on the best or worst of a set of assets, provide a useful starting analogy from finance.

Finally, it would be useful to explore the use of simultaneous betting strategies from the bookmaker's perspective. Bookmakers find themselves exposed to potentially disastrous outcomes from bettors placing small wagers on large accumulators. It would be interesting to analyze whether the additional profits generated from offering accumulator bets exceed the costs of potentially large payouts. Given the strong marketing of accumulator bets, it appears that the occasional large bookmaker payout is more than made up for by the large number of small, unsuccessful bets made by punters.

Notes

1. The *Financial Times* (Thompson 2012) reported that the U.K. bookmaker Ladbrokes blamed a series of victories in December for heavily backed Premier League teams, combined in accumulator bets, for a 12.3 percent decline in annual profits. Ladbrokes had ironically advertised on their website the big win of a punter earlier that year (Kemp 2011).
2. These are known (in betting parlance)as Trixies (4 total bets, three-way accumulator maximum), Yankees (11 total bets, four-way maximum), Canadians (26 total bets, five-way maximum), Heinz (57 total bets, six-way maximum), Super Heinz (120 total bets, seven-way maximum), and Goliaths (247 total bets, eight-way accumulator maximum). Separate products exist for a bettor to add the fixed-stake bet on the single outcome to the full cover bets (a patent, for example, is the same as the Trixie but with three single-game bets included).

References

Ali, Mukhtar M. 1979. Some evidence on the efficiency of a speculative market. *Econometrica* 47(2):387–392.

Breiman, L. 1961. Optimal gambling systems for favorable games. *Proceedings of the fourth Berkeley Symposium of Probability and Statistics* 1:65–78.

Golec, Joseph, and Maurry Tamarkin. 1998. Bettors love skewness, not risk, at the horse track. *Journal of Political Economy* 106(1):205–225.

Grant, Andrew, and Peter W. Buchen. 2012.A comparison of simultaneous Kelly betting strategies. *Journal of Gambling Business and Economics* 6(2):1–28.

Grant, Andrew, David Johnstone, and Oh Kang Kwon. 2008. Optimal betting strategies for simultaneous games. *Decision Analysis* 5(1):10–18.

Griffin, Peter A. 1999. *The theory of blackjack: The compleat card counter's guide to the casino game of 21*. 6th ed. Las Vegas, Nev.: Huntington.

Insley, Robin, Lucia Mok, and Tim Swartz. 2004. Issues related to sports gambling. *Australian & New Zealand Journal of Statistics* 46(2):219–232.

Kelly, John L. (Jr) 1956. A new interpretation of information rate. *Bell Systems Technical Journal* 35(4):917–926. http://en.wikipedia.org/wiki/John_Larry_Kelly,_Jr.

Kuypers, Tim. 2000. Information and efficiency: An empirical study of a fixed odds betting market. *Applied Economics* 32(11):1353–1363.

Maclean, Leonard C. William T. Ziemba, and George Blazenko 1992. Growth versus security in dynamic investment analysis. *Management Science* 38(11):1562–1585.

Medo, Matus, Yury M. Pis'mak, and Yi-Cheng Zhang. 2008. Diversification and limited information in the Kelly game. *Physica A* 387:6151–6158.

Thompson, Christopher. 2012. Online profits down at Ladbrokes. *Financial Times*, Feb 17; http://www.ft.com/cms/s/0/af12bb14-58ab-11e1-9f28-00144feabdc0.html#axzz2SXCUu9Sj.

Thorp, Edward O. 2000. The Kelly criterion in blackjack sports betting, and the stock market. In *Finding the edge: Mathematical analysis of casino games*, edited by Olaf Vancura, Judy A. Cornelius, and William R. Eadington. Reno, Nev.: Institute for the Study of Gambling and Commercial Gaming, 163–213.

Whitrow, Chris. 2007. Algorithms for optimal allocation of bets on many simultaneous events. *Journal of the Royal Statistical Society, Series C (Applied Statistics)* 56(5):607–623.

CHAPTER 20

A PRIMER ON THE MATHEMATICS OF GAMBLING

ROBERT C. HANNUM

Introduction

The mathematics underlying gambling can explain why, over time , one player wins while another loses. In the context of commercial gaming, the "player" who wins is typically the house—the operators of the casino, racetrack, card room, or lottery. Simply put, commercial gaming enterprises make money because of the mathematics behind the games. As many experienced gamblers have learned—for better or worse—and some knowledgeable insiders have voiced, "There is no such thing as luck; it is all mathematics." This chapter presents the fundamental mathematics of gambling and shows how the math behind the games generates revenues and drives the economics of gambling.

Win Rate Metrics

In the gambling business a variety of terms are used to refer to the rate at which a player wins (or loses) money. These include

- house advantage,
- house edge,
- theoretical win percentage,
- expected win percentage,
- win percentage,
- hold percentage.

Some of these win rate metrics refer to the same concept, but others are distinctly different measures of win rate. The *house advantage* is (under the usual formulation) the long-term percentage of money wagered that is retained by the house. The *house edge*, *theoretical win percentage*, and *expected win percentage* are just different names for the house advantage. *Win percentage* refers to the actual percentage of money won during a finite gambling session or series of bets. In the short term the actual win percentage will differ from the theoretical win percentage; the magnitude of this deviation can be predicted from statistical theory. The win percentage is equal to the observed win divided by the *handle*, the total amount wagered. Because of the law of large numbers, the actual win percentage should get closer to the theoretical win percentage as the number of trials gets larger: As $n \to \infty$, $Win\% \to HA$.

In the casino business handle can be difficult to measure for table games (this may well change when smart table/chip technology becomes cost-effective), and game performance is often measured by *hold percentage*. Hold percentage is equal to win divided by *drop*. Drop is the total amount of the currency and chips in the table's drop box—a locked box affixed to the underside of the table—plus the value of credit instruments issued or redeemed at the table.

For slot machines, where each bet is recorded, actual hold percentage and actual win percentage are in principle equivalent, and the house advantage for a slot is sometimes referred to as the theoretical hold percentage. For table games, however, hold percentage is sometimes confused with house advantage and/or win percentage. To illustrate how different these metrics are, consider that in Nevada in 2010 the hold percentage for roulette was about 17 percent, whereas the house advantage is about 5 percent. Similar large differences between hold percentage and house advantage exist for other games. Nevada's hold percentages in 2010 for blackjack, baccarat, three-card poker, Caribbean stud, and let it ride were 11 percent, 11 percent, 28 percent, 26 percent, and 23 percent, respectively, compared to house advantages of roughly 1 percent for blackjack and baccarat and 3–5 percent for the latter three games. It would be a mistake to think that the house wins more than 10 percent of all monies wagered at any of these games.

In summary, house advantage, house edge, theoretical win percentage, and expected win percentage are different names for the same metric. Hold percentage is win divided by drop; (actual) win percentage is win divided by handle. Win percentage approaches the house advantage as the number of plays increases.

House Advantage—A Caveat

The house advantage can be subject to varying formulations and interpretations. In the table game let it ride, for example, the casino advantage can be expressed as 3.51 percent of the base bet or 2.86 percent of the average bet. Those familiar with the game know that the player begins with three equal base bets but may withdraw one or two of these initial units. The final amount put at risk, then, can be one (84.6% of the time assuming proper strategy), two (8.5%), or three units (6.9%), making the average bet size 1.224 units. In the long run the casino will win 3.51 percent of the hands, which equates to 2.86 percent of the money wagered. So it is correct to say that the house edge is 3.51 percent per hand; it is also correct that it is 2.86 percent per unit wagered. Regardless, the bottom line is that with three $1 base bets, the casino can expect to earn 3.5¢ per hand ($1.224 \times 0.0286 = 0.035$). The question of whether to express the house advantage relative to a base bet or average bet also arises in the casino games Caribbean stud poker (5.22% vs. 2.56%), three-card poker (3.37% vs. 2.01%), casino war (2.88% vs. 2.68%), and red dog (2.80% vs. 2.37%).

Another situation that admits varying formulations of the house advantage is a bet that can result in a tie. For such bets the house edge can be stated either including or excluding ties. Examples include baccarat's player bet (1.24% if ties are included vs. 1.37% if ties are excluded), banker bet (1.06% vs. 1.17%), and the don't pass bet (1.36% vs. 1.40%) in craps. Again, regardless of which view is taken, proper interpretation leads to the same expected revenue in dollars and cents.

The Basic Mathematics of Gambling

The language of gambling math revolves around such terms as *probability*, *odds*, *expectation*, *house advantage*, and *volatility*. This section presents an overview of these terms.

Probability and Odds

Probability is a number between zero and one that represents the relative likelihood that an event will occur. The closer the probability is to one, the more likely the event is to occur—an event with a probability of one is certain to occur; the closer to zero, the less likely it is to occur—if the probability is zero, it is impossible for the event to occur. A probability of one-half, or 50 percent, means that the event in question is just as likely to occur as not occur. Probability can be viewed as the long-run ratio of {# of times an outcome occurs}to {# of times experiment is conducted}. The *odds* against an event represent the long-run ratio of {# of times an outcome does not occur}to {# of times an outcome occurs}. For example, the probability of rolling a seven with a pair of honest dice is 1/6 and the odds against are 5 to 1. If one card is randomly selected from a standard deck of 52 playing cards, the probability that it is a diamond is 1/4 and the odds against it being a diamond are 3 to 1; the probability that it is an ace is 1/13 and the odds against it being an ace are 12 to 1.

The odds against an event represent the payoff that would make a bet on that event fair. A bet on black in single-zero roulette has a probability of winning equal to 18/37, so the payoff necessary to make it a fair bet is 19 to 18 (the actual payoff is 18 to 18);

for the same game a bet on a single number has a probability of winning equal to 1/37, so to break even in the long run a player winning this bet would need to be paid 36 to 1 (the actual payoff is 35 to 1).

Laws of Probability

The three major laws of probability are the Complement Law, the Addition Law, and the Multiplication Law.

Complement Law: The probability of an event not occurring is equal to one minus the probability of that event occurring.

Addition Law: The probability of at least one of two events occurring equals the sum of their individual probabilities, minus the probability they both occur.

Multiplication Law (independent events): For independent events, the probability of all of them occurring equals the product of their individual probabilities. (Events are independent if the occurrence of one has no effect on the probability of the occurrence of the others.)

Multiplication Law (dependent events): For nonindependent events, the probability of all of them occurring equals the product of their conditional probabilities, where the conditional probability of one event is affected by the event(s) that came before it.

Examples illustrating the use of these basic probability laws are given below.

Complement Law: When selecting a card at random from a standard deck of playing cards, the probability it is a spade is 13/52 or .25. The probability it is not a spade is $(1 - .25) = .75$.

Addition Law: When a single card is randomly selected from a deck, the probability it is an ace or a spade is $(4/52 + 13/52 - 1/52) = 16/52$ or .308. When rolling a pair of dice, the probability of getting a total of 7 or 11 is $(6/36 + 2/36) = 8/36$, and the probability of rolling a 2, 3, or 12 is $(1/36 + 2/36 + 1/36) = 4/36$.

Multiplication Law (independent events): If you flip a coin two times, whether you flip heads or tails on the first flip has no influence on whether you will flip heads or tails on the second flip. The two flips of the coin are independent events. Thus the probability of two heads in two tosses is $(1/2) \times (1/2) = 1/4$. Similarly, the probability of getting five heads in five tosses of a fair coin is $(1/2)^5 = 1/32$ or about .031. The probability of getting five consecutive 12's in five tosses of a pair of dice is $(1/36)^5 = .000000017$ or about 1 in 60,466,176. The probability of 10 consecutive red numbers in 10 spins of a roulette wheel is $(18/38)^{10} = .0005687$ or about 1 in 1,758. The calculations in these examples are valid because the events involved are independent trials. The outcome on any given trial has no effect on the outcome of any other trial.

Multiplication Law (dependent events): If three cards are randomly selected (without replacement) from a standard deck, the probability all three are aces is (4/52)(3/51)(2/50) or .000181 (1 in 5,525). Notice that the trials here are not independent since the outcome of the first draw will affect the outcome on the second, and the outcome of the second will affect the outcome on the third.

The latter examples highlight a key difference between games like roulette and craps and the game of blackjack. Roulette and craps involve independent trials, whereas blackjack involves dependent trials. In roulette and craps the probabilities remain the same from trial to trial (it matters not if there are 10 consecutive red numbers; the probability of red is still 18/38 on the 11th trial). In blackjack the trials are dependent since the composition of the cards remaining to be played depends on which cards have already been played. It is this feature, dependent trials, which causes the advantage in blackjack to flow back and forth between the house and the player and why "card-counting" can be effective in blackjack.

Mathematical Expectation

The amount of money a player can expect to win or lose in the long run on a given wager at a given bet size—if the same bet is made over and over again—is called the wager *expectation* or *expected value* (EV). When the EV for a bet is negative, the player will lose money over time on this bet.

The general formula for the expectation, or EV, of a bet, is

$$EV = \sum (Win_i \times p_i), \tag{20.1}$$

where Win_i is the net win associated with outcome i and p_i is the probability of Win_i. The EV for a bet represents the amount of money the bettor will win or lose on average (in the long run) from making the specified bet.

For example, a \$5 bet on the color red in double-zero roulette (38 total numbers, of which 18 are red, 18 are black, and 2—the 0 and 00—are green) has a probability of winning equal to 18/38, a probability of losing equal to 20/38, and pays even money if won. The wager expectation is

$$EV = (+5)(18/38) + (-5)(20/38) = -0.263.$$

On the average the player will lose just over \$0.26 for each \$5 bet on red.

House Advantage

As Nicholas Pileggi phrased it in his novel *Casino* (1995), "A casino is a mathematics palace set up to separate players from their money. Every bet made in a casino has been calibrated within a fraction of its life to maximize profit while still giving the players the illusion that they have a chance." The fundamental reason that casino games make money is the house advantage, a mathematical edge built into virtually every bet in the casino. This house edge, coupled with the famous mathematical result called the law of large numbers, virtually guarantees that the casino will win in the long run.

The house advantage (HA) is the negative of the expected value of the bet, expressed as a percentage of the amount bet

$$HA = \frac{-EV}{Bet} \times 100\%. \tag{20.2}$$

Note that if the EV calculation in (20.1) is performed for a one-unit amount, the negative of the resulting value is the house edge, expressed as a decimal. For example, for a one-unit bet on a single number (payoff equal to 35 to 1) in double-zero roulette

$$EV = (+35)(1/38) + (-1)(37/38) = -0.053.$$

Thus the house advantage is 5.3 percent.

The house advantage, or house edge, represents the long-run percentage of the wagered money that will be retained by the house. In economics terms, the house advantage is the price to the player for playing the game. Although the house edge can be computed easily for some games—for example, roulette and craps—for others it requires more sophisticated mathematical analysis and/or computer simulations. Table 20.1 shows typical prices for some popular casino games.

Because this positive house edge exists for virtually all bets in a casino, in these commercial gambling halls players are faced with an uphill and, in the long run, losing battle. There are some exceptions, however. The odds bet in craps has zero house edge (although this bet cannot be made without making another negative expectation wager); a select few video poker machines return greater than 100 percent if played with optimal strategy, and skilled blackjack card counters, poker players, and sports bettors can make money playing their games. But these are small groups—very few

Table 20.1 Typical Prices for Popular Casino Games

Game	House Advantage
Roulette (double-zero)	5.3%
Craps (pass/come)	1.4%
Craps (pass/come with double odds)	0.6%
Blackjack* (6 decks)	0.5%
Blackjack—average player	2.0%
Baccarat (no tie bets)	1.2%
Caribbean Stud*	5.2%
Let It Ride*	3.5%
Three-Card Poker*	3.4%
Pai Gow Poker (ante/play)*	2.5%
Slots	5% – 10%
Video Poker*	0.5% – 3%
Keno (average)	27.0%

*optimal strategy

have what it takes to win money in the long run at any of these games. Occasionally the casino will offer a promotion that gives the astute player a positive expectation. These promotions are often mistakes—sometimes casinos fail to check the math—and are terminated once the casino realizes the player has the edge. But by and large the player will lose money in the long run, and the house edge is a measure of how fast the money will be lost.

Gambling Economics

In the gambling business the standard revenue-price-quantity economics relationship, $R = p \times q$, becomes

$$Expected\ Win = HA \times handle. \tag{20.3}$$

The left side of expression (20.3) is "Expected Win" rather than "Win" because of the statistical nature of the gambling business. While the house advantage (*HA*) will be realized in the long run, the actual win is subject to statistical fluctuations, and in the short term the win percentage may differ from the house advantage. The size of these fluctuations can be predicted using statistical theory. The subject of volatility is discussed in more detail below.

Game Pricing

Unlike most other consumer purchases where the price of a product is fixed, the price of a gambling game depends on the specific rules of the game, the payoffs for winning wagers, the amount of time the player plays, and possibly the skill of the player. In large part, then, pricing in the gaming business is made by conscious or unintentional decisions on setting the house advantage (informally, "setting the odds") for the games. This can be done through rule variations that are more or less favorable to the player, by altering payoffs, or by offering a different type of game product.

Rule Variations

Rule variations can affect the house advantage in all games. Two of the more notable in which casinos vary the rules most often are blackjack and craps. In some blackjack games, the dealer must hit a soft seventeen (a soft hand is one that includes an ace that is being counted as 11), which increases the house advantage 0.2 percent over a comparable game where the dealer stands on soft seventeen. Other examples of rule variations in blackjack that are house-favorable include multiple decks, no soft doubling, no re-splitting of pairs, and naturals paying 6 to 5 (usually 3 to 2). Player-favorable blackjack rule variations include double down after splitting pairs, late and early surrender, and re-splitting aces.

In craps the amount of "free odds" that can be taken by a player will change the price of the game product. A player who bets the pass line and takes single odds (i.e., the amount of the odds bet is equal to the amount of the pass line bet) is at a 0.85 percent disadvantage but only at 0.61 percent with double odds and 0.47 percent with triple odds. This means that for every $100 combined bet on the pass line with single odds (i.e., $50 pass line plus $50 odds) the player, on average, will pay a price of about $0.85, while a $100 combined bet on the pass line with triple odds ($25 pass line plus $75 odds) will cost about $0.47. A casino allowing triple odds offers a better priced craps game than one that permits only single odds. Some casinos have offered as high as 100X odds—a player taking full 100X odds will face only a 0.02 percent house advantage on the combined pass line (or come) plus odds wagers.

Altering Payoffs

Another way to set prices on games is to alter the payoffs. The field bet in craps, for example, typically offers even money on the 3, 4, 9, 10, and 11, and a 2 to 1 payoff for the 2 and 12. The house advantage for this payoff structure is 5.56 percent. If either the 2 or the 12 (but not both) were to pay triple, the house advantage would be reduced to 2.78 percent. If both the 2 and the 12 pay 3 to 1, the field bet is "free"—the house advantage is equal to zero.

In baccarat, the usual casino commission on winning banker bets is 5 percent. Some casinos attempt to attract and keep players by lowering this commission to 4 percent or even 3 percent (setting the commission at only 2% would give the player the advantage). Such a reduction is equivalent to raising the payoff on this bet from 95 to 96 or 97 percent, with the effect of reducing the house advantage from the usual 1.06 to 0.60 percent (with the 4% fee) or 0.14 percent (3% fee).

Keno and video poker are other examples where payoffs can be easily manipulated to raise or lower the price of the game. Payoffs for various keno wagers vary widely across casinos. A full-pay Jacks-or-Better video poker machine pays 9 for 1 and 6 for 1 for a full house and flush, respectively (hence the moniker "9/6" in reference to this game), but machines paying only 8 for 1 and 5 for 1 for these hands ("8/5" machines) are common. The former returns 99.54 percent with perfect play while the maximum payback on the latter is 97.29 percent. Some casinos have offered a version of Jacks-or-Better with a 7 for 1 payoff on flushes, giving the player a greater than 100 percent return with perfect play.

Offering a Different Game Product

Introducing new games, or variations on standard games, is a way to diversify both the product mix and the range of prices offered. Games like Double Exposure and Spanish 21 are just twists on blackjack, and three-card poker, let it ride, and Caribbean stud are variations on poker. These differing game products offer players a wide variety of prices to pay for their gaming experience. Single-zero roulette costs the player about \$0.27 for every \$10 wagered compared to the double-zero wheel with a price of \$0.53 per \$10 bet.

Other Factors Affecting Game Prices

For games with a skill component, the effective house advantage depends on the skill level of the player. In a typical six-deck blackjack game, for example, the average player gives about a 1.5 percent edge to the house while a perfect basic strategy player is at a 0.5 percent disadvantage. Game speed (number of decisions per hour) affects the cost per hour to the player, though it does not alter the base price per unit wagered. In roulette, for example, a \$20-per-spin player betting at a double-zero table making 40 spins per hour can expect to pay about \$42 per hour (\$20 × 40 × .0526). At a table completing 60 spins each hour the same player would spend an average of \$63 per hour. Similarly, at a base price of 1.15 percent, a \$1,000-per-hand baccarat player will pay, on average, about \$920 per hour if playing 80 hands per hour, but it will cost this same player \$1,380 per hour if dealt 120 hands per hour.

Pricing Gaffes

Commercial gaming establishments occasionally offer novel wagers, side bets, increased payoffs, or rule variations in an effort to entice players and increase business. These promotions are intended to lower the house advantage and the effective price of the game for the player. While this might be sound reasoning from a marketing standpoint, it can have negative consequences for the gaming business enterprise if care is not taken to ensure that the math behind the promotion is sound.

An Illinois riverboat casino reportedly lost \$200,000 in one day with a "2 to 1 Tuesdays" promotion that paid players 2 to 1 on blackjack naturals (the usual payoff is 3 to 2). Without other compensating rule changes, a 2-to-1 payoff on naturals in a typical blackjack game can increase the players' expectation enough to give them a 1.5 to 2 percent advantage over the house. Other casinos also fell victim to this faux pas, and one casino in California went so far as to pay out 3 to 1 on naturals during a "happy hour" offered three times a day, two days a week, for over two weeks, giving the player a healthy 6 percent edge.

A casino in Mississippi ran a promotion in which it offered 80 to 1 payoffs instead of the usual 60 to 1 for bets on (a total of) 4 and 17 in sic bo, a game in which players bet on the outcome of the roll of three dice. The effect of this change on the player's expectation is easy to determine. Since the probability of rolling a total of 4 (or 17) with three dice is $1/72(1/6 \times 1/6 \times 1/6 \times 3)$, the expected values of this bet under the usual and the promotional payoffs are, respectively,

$$EV = (+60)(1/72) + (-1)(71/72) = -0.153$$

and

$$EV = (+80)(1/72) + (-1)(71/72) = +0.125.$$

The increase in payout changed the advantage from 15.3 percent in favor of the house to 12.5 percent in favor of the player. With this advantage, a player betting \$100 per hand at 50 hands per hour can expect to win \$625 per hour; a \$500 per hand bettor would win an average of \$3,125 per hour.

In still other examples of pricing blunders in the gambling business:

- A major Las Vegas casino suffered a \$230,000 loss in three-and-a-half days on a "50/50 Split" blackjack side bet that allowed the player to stand on an initial holding of 12–16 and begin a new hand for equal stakes against the same dealer up card. Although the game marketers claimed the variation was to the advantage of the casino, it turned out that those players who exercised the 50/50 Split only against a dealer's 2–6 enjoyed a 2 percent advantage.
- A small Las Vegas casino lost an estimated \$17,000 in eight hours when it offered a blackjack rule variation called the "Free Ride" in which players were given a free right-to-surrender token every time they received a natural—proper use of the token led to a player edge of 1.3 percent.
- Another casino reduced the commission on winning baccarat banker bets from 5 to 2 percent, resulting in a 0.32 percent player advantage.

The pricing mistakes above serve to highlight how important it is to be vigilant when altering payoffs or varying rules to offer a more or less attractive game. Payoffs that are too large or rules that are too player-favorable could give the players an advantage and cost the house dearly. On the other hand, if payoffs are too low or rules too house-friendly, sales volume may suffer because players will not want to play. A proper mathematical analysis is needed on all proposed side bets, payoff changes, rules variations, and marketing promotions to ensure that the economics of the game is both attractive to players and acceptable to the house.

Price and Product Demand

In most businesses, if an operator lowers the price of goods or services the volume of sales will increase, provided that the operator's competitors also do not lower their prices. A commercial gaming establishment may consider lowering the price (house advantage) to the player to increase the volume of play. In some cases the motivation for this may be to keep the tables full—the gaming operators may see idle tables and come to the conclusion that they would be better off filling the empty tables at worse odds since they are already paying for the dealers and other game supervisors. Ultimately the gaming establishment needs to consider the impacts of lowering the price on operating profits. Unfortunately this is not a simple question—whether the reduced price will result in sufficiently increased volume of play is often unknown. That is, it is difficult for gaming establishments to accurately estimate price elasticities of demand.

Player Value

The most valuable assets to a commercial gaming establishment are its players—the player base is the economic engine that drives revenues. An important problem for management is to determine the value of each player. Assessing a player's value to the gambling establishment allows it to develop typical player profiles and decide game mixes that appeal to its players. Player value is also necessary for effective marketing programs such as complimentaries, rebates on theoretical loss, discounts on actual loss, and dead chip programs that attempt to attract and retain players.

Player value is simply how much the gambling establishment can expect to win from that player. It is often called earning potential, player worth, or theoretical win. Player value can be computed using the house advantage (*HA*), bet size, duration of play, and pace of the game as follows:

$$\text{Player Value} = \text{Average Bet} \times \text{Hours Played} \times \text{Decisions per Hour} \times \text{HA}.$$

For example, suppose a roulette player bets \$25 per spin for four hours at 50 spins per hour. Using a house advantage of 5.3 percent (double-zero roulette), this player would be expected to lose, and would be rated at a player value of, \$265$(25 \times 4 \times 50 \times .053)$. If this player were betting \$100 per spin for 12 hours, the player value rating would be \$3,180.

Complimentaries, or simply "comps," can include anything from free drinks to payment of every conceivable expense, including room, food, and beverage (RFB) and airfare. A typical comp policy is to award a player with comps equal to a set percentage of the player's earning potential. Rebates and discounts are similar to comps but offer a monetary return rather than a meal or a hotel room in exchange for a certain level of play.

Risk and Volatility

The risk associated with a gamble is due to the uncertainty of the outcome. Although the house advantage can be used to predict the amount that will be won (or lost) in the long run, in the short term the actual win will deviate from the expected win. How much variation from the theoretical win can be expected? What might be considered a normal fluctuation and what is considered unusual? The magnitude of the difference between actual and theoretical wins for a given number of wagers can be predicted using statistical theory. The basis for such a volatility analysis is the standard deviation. This statistical measure of variation is, roughly speaking, the average deviation of all possible outcomes from the expected. Together with the central limit theorem—one of the most important results in statistics—the standard deviation can provide confidence limits on the amount a player should win or lose over a series of wagers. The mathematical formulation of the standard deviation (SD) of a bet is

$$SD = \sqrt{\sum [(Win_i - EV)^2 \times p_i]} \tag{20.4}$$

where Win_i is the net win associated with outcome i, EV is the expected value of the bet given by (20.1), and p_i is the probability of Win_i.

As an example, consider a \$5 bet on a single number in double-zero roulette. With a payoff of 35 to 1 the base expected value and standard deviation of this bet are

$$EV = (+175)(1/38) + (-5)(37/38) = -0.263;$$

$$SD = \sqrt{(175 - (-0.263))^2(1/38) + (-5 - (-0.263))^2(37/38)} = 28.813.$$

The single-number bet has the greatest volatility in roulette. Analogous calculations for a \$5 bet on red, for example, yield a bet standard deviation of 4.993 (the bet EV is the same –0.263 as the single-number bet).

Confidence Limits

While expressions (20.1) and (20.4) can be used to calculate the base EV and SD of a bet, it is the expected value and standard deviation of a series of n bets that are needed to obtain confidence limits for the amount won over a series of wagers. For a series of n identical, independent bets of one unit each, the expected value and standard deviation of the total number of units won are given by

$$EV(units\ won) = n \times EV; \tag{20.5}$$

$$SD(units\ won) = \sqrt{n} \times SD, \tag{20.6}$$

where EV and SD are the bet expected value and bet standard deviation given by (20.1) and (20.4), respectively.

Assuming n is large, the central limit theorem guarantees that the distribution of the observed units won in the n trials will be approximately normal with mean and standard deviation equal to (20.5) and (20.6), respectively. Thus standard normal distribution theory can be invoked to obtain upper and lower limits for the actual number of units won for any desired level of confidence. The following example illustrates how this is done.

Consider a series of 1,000 one-unit bets on a single number—payoff equal to 35 to 1—in double-zero roulette. The bet expected value and bet standard deviation from (20.1) and (20.4) are –0.05263 and 5.76262, respectively. From (20.5) and (20.6), then, the expected value and standard deviation of the number of units won in a series of 1,000 bets are

$$EV(\textit{units won}) = 1{,}000 \times -0.05263 = -52.63; \tag{20.7}$$

$$SD(\textit{units won}) = \sqrt{1{,}000} \times 5.76262 = 182.23. \tag{20.8}$$

Using (20.7), (20.8), and the appropriate standard normal probability value (z) for the desired level of confidence, limits for the number of units won in the series of 1,000 bets can be derived. The z value can be obtained from a table of standard normal probabilities or a software package, such as MS Excel. For example, the z value for 95 percent confidence is 1.96, so the 95 percent confidence limits for the number of units won are

$$-52.63 \pm 1.96(182.23) = -409.8 \textit{ and } 304.5. \tag{20.9}$$

The interpretation of (20.9) is that in a series of 1,000 independent single-number (double-zero) roulette bets of one unit each, 95percent of the time the player win will be between -409.8 and $+304.5$ units.

It is easy to convert limits for the number of units won to limits for the amount of money won or the win percentage. If the bets in the above roulette example were \$5 each, the 95 percent limits for the amount of money won in the series would be $-$\$2,049 and $+$\$1,523 (obtained by merely multiplying the limits in (20.9) by \$5), and the 95 percent limits for the win percentage would be -41.0 percent and 30.5 percent (dividing the limits in (20.9) by 1,000).

Appropriately modifying the z value will produce limits for other desired levels of confidence. For example, using $z = 2.58$ for the above series of \$5 single-number roulette bets produces 99 percent confidence limits for number of units won, amount of money won, and win percentage: -522.8 and 417.5, $-$\$2,614 and \$2,088, and -52.3 percent and 41.8 percent, respectively.

Note that the confidence limits for the win percentage will converge to the house advantage as the number of bets increases. This is the result of the law of large numbers—as the number of trials gets larger the actual win percentage should get closer to the theoretical win percentage.

Volatility Benchmarks

An easy rule of thumb set of benchmarks for volatility can be fashioned around the percentages of times an outcome will be more than a various number of standard deviations (*SD*s) from the expected outcome.

Volatility Benchmarks

- Outcomes will be more than 2 SDs from the expected about 5 percent of the time.
- Outcomes will be more than 3 SDs from the expected about 0.3 percent of the time.
- Outcomes will be more than 4 SDs from the expected about 0.006 percent of the time.
- Outcomes will be more than 5 SDs from the expected about 0.00006 percent of the time.

For the number of units won in the series of 1,000 one-unit bets on a single number in double-zero roulette considered above, the volatility benchmarks translate to

- About 5 percent of the time the player will either win more than 312 units or lose more than 417 units.
- About 0.3 percent of the time the player will either win more than 494 units or lose more than 599 units.
- About 0.006 percent of the time the player will either win more than 676 units or lose more than 782 units.
- About 0.00006 percent of the time the player will either win more than 859 units or lose more than 964 units.

One-Sided Limits and Benchmarks

The confidence limits and volatility benchmarks described above are two-sided, providing probabilities that an observed win will deviate from the expected win by more or less than a certain amount *in either direction.* If one is concerned with a deviation in only one direction, one-sided limits and benchmarks are appropriate. This would be the case, for example, when a player in a commercial gaming establishment wins what appears to be an unusually large amount and management is interested in the odds of such a large win, assuming an honest game.

For the roulette example considered above—a series of 1,000 bets of $5 each on a single number in double-zero roulette—one can say that 95 percent of the time the player win should be no more than 248.0 units or $1,240. These values are the 95th percentiles of their respective distributions, obtained by using $z = 1.65$, the 95th percentile of the standard normal distribution. (Note that from the two-sided, 95% confidence limits one can say there is a 97.5% chance that the player win will be no more than 304.5 units or $1,523.)

Volatility benchmarks are easily modified to one-sided

- Outcomes will be more than 2 *SDs above* the expected about 2.5 percent of the time.

- Outcomes will be more than 3 *SDs above* the expected about 0.15 percent of the time.
- Outcomes will be more than 4 *SDs above* the expected about 0.003 percent of the time.
- Outcomes will be more than 5 *SDs above* the expected about 0.00003 percent of the time.

One-sided benchmarks for outcomes more than a certain number of standard deviations below the expected are analogous to those above, replacing the word "above" with "below."

Likelihood of an Observed Win

Although confidence limits and benchmarks provide useful tools for analyzing risk and volatility, it often is helpful to compute the likelihood of a particular observed win. This can be accomplished by calculating the z score associated with the observed win and then finding the relevant probability referring to a standard normal probability table (Z table). The observed z score is calculated as follows:

$$z = \frac{Observed\ Win - Expected\ Win}{SD(Win)}. \tag{20.10}$$

The *Expected Win* and $SD(Win)$ in (20.10) are the expected value and standard deviation of the series win, measured in the same units as the *Observed Win* (units, dollars, or percent). For example, suppose after a series of 1,000 bets of $5 each on a single number in double-zero roulette a player has won $3,000. What are the chances of this happening (assuming an honest game)? From (20.7) and (20.8), the expected value and standard deviation of the number of units won are −52.63 and 182.23, respectively, or −$263.16 and $911.15. Thus, the z score associated with the observed $3,000 win is

$$z = \frac{\$3,000 - (-\$263.16)}{\$911.15} = 3.58.$$

This means that the player's $3,000 win was 3.58 standard deviations more than what was expected. A lookup in a standard normal probability table shows the likelihood of such a win or greater to be .00017 or about 1 in 5,851. Had this player won $5,000, the odds against such a win or greater would be about 1 in 262 million (probability equal to .0000000038).

Note that had the player in the above example been making bets on red rather than a single number, a $3,000 win or more would have been practically impossible ($z = 20.67$) due to the much smaller volatility associated with a bet on red—the $SD = 4.993$ for a $5 bet on red versus $SD = 28.813$ for a $5 bet on a single number.

A Final Example

Consider a series of 1,000 pass line wagers in craps. It can be shown from (20.1) and (20.4) that the expected value and standard deviation of a single pass line bet are $EV = -0.014$ and $SD = 1.000$, respectively. For a series of 1,000 pass line wagers, then, from (20.5) and (20.6)

$$EV(units\ won) = 1,000 \times -0.014 = -14.0;$$

$$SD(units\ won) = \sqrt{1,000} \times 1.000 = 31.6.$$

The expected value results mean that the house advantage for a pass line bet is 1.4 percent and after 1,000 bets on average the player will be behind by 14 units. Applying the two-sided volatility benchmarks, there is roughly a 95 percent chance that the player's actual win will be between 49 units ahead and 77 units behind and about a 0.15 percent chance that the player win will be between 81 units ahead and 109 units behind. If each of the 1,000 bets was \$50, the player would be expected to lose \$700, with two and three standard deviations benchmark limits of –\$3,860 and +\$2,460 and –\$5,440 and \$4,040, respectively. That is, approximately 5 percent of the time the player will either win more than \$2,460 or lose more than \$3,860, and about 99.85 percent of the time the player will either win more than \$4,040 or lose more than \$5,440. If, for example, the player actually won \$10,000, management might want to scrutinize this player and game more closely, as such a win would represent an event with one-sided odds of 157.6 billion to 1 ($z = 6.77$), assuming an honest game.

Additional Sources

Ethier (2010) provides the most comprehensive coverage of the mathematics of gambling; a more practical, somewhat less mathematical treatment can be found in Hannum and Cabot (2005). Other general references on the mathematics of gambling include Epstein (2009), Griffin (1991), Packel (1981), and Thorp (1984). David (1998) gives an excellent historical perspective on the subject. Two volumes produced by the University of Nevada at Reno's Institute for the Study of Gambling and Commercial Gaming, Ethier and Eadington, eds. (2007) and Vancura, Cornelius, and Eadington, eds. (2000), contain scholarly research articles on the subject. Kilby, Fox, and Lucas (2005) cover casino operations management issues related to the mathematics of gambling; Cabot and Hannum (2002) explore the relationship between gaming regulation and mathematics. Two classic and excellent books on probability theory more generally are Feller (1968) and Weaver (1982).

Summary

Mathematics lies at the heart of the gambling business. It helps us understand how a commercial gaming establishment generates a profit and why players usually lose, in the long run. A familiarity with probability is necessary in order to understand the principles of gambling mathematics, principles that include such gambling theory fundamentals as mathematical expectation, house advantage, and volatility. From these topics flow a plethora of related economics issues, such as game pricing, player value, and risk assessment and management. From the assurance of revenues due to the game advantage to the earning potential of players and the risk inherent in making a bet, mathematics is the key to the economics of gambling.

References

Cabot, Anthony N., and Robert C. Hannum. 2002. Gaming regulation and mathematics: A marriage of necessity. *John Marshall Law Review* 35:333–358.

David, F. N. 1998. *Games, gods, and gambling: A history of probability and statistical ideas.* Mineola, N.Y.: Dover Publications.

Epstein, Richard A. 2009. *The theory of gambling and statistical logic.* 2nd ed. Burlington, Mass.: Elsevier.

Ethier, Stewart N. 2010. *The doctrine of chances: Probabilistic aspects of gambling.* Berlin: Springer-Verlag.

Ethier, Stewart N., and William R. Eadington, eds. 2007. *Optimal play: Mathematical studies of games and gambling.* Reno: Institute for the Study of Gambling and Commercial Gaming, University of Nevada.

Feller, William. 1968. *An introduction to probability theory and its applications.* 3rd ed. New York: Wiley.

Griffin, Peter. 1991. *Extra stuff: Gambling ramblings.* Las Vegas, Nev.: Huntington Press.

Hannum, Robert C., and Anthony N. Cabot. 2005. *Practical casino math.* 2nd ed. Reno: Institute for the Study of Gambling and Commercial Gaming, University of Nevada.

Kilby, Jim, Jim Fox, and Anthony F. Lucas. 2005. *Casino operations management.* 2nd ed. New York: Wiley.

Packel, Edward. 1981. *The mathematics of games and gambling.* Washington, D.C.: Mathematical Association of America.

Thorp, Edward O. 1984. *The mathematics of gambling.* Hollywood, Calif.: Gambling Times.

Vancura, Olaf, Judy A. Cornelius, and William R. Eadington, eds. 2000. *Finding the edge: Mathematical analysis of casino games.* Reno: Institute for the Study of Gambling and Commercial Gaming, University of Nevada.

Weaver, Warren. 1982. *Lady luck: The theory of probability.* New York: Dover Publications.

CHAPTER 21

THE SCIENCE AND ECONOMICS OF POKER

ROBERT C. HANNUM

Introduction

Although poker has roots in card games played in Europe and the Middle East as far back as the fifteenth century, the modern game of poker originated in the early nineteenth century in the gambling saloons of New Orleans. It has long been the most popular card game among Americans and recently has reached unprecedented popularity due to a variety of factors, including television and Internet exposure. Poker is somewhat unique among gambling activities due to the fact that it is not house banked and because of the considerable elements of skill required and not present in other casino-style and lottery-type games. Those interested in the history of card games generally may wish to consult Parlett (1991). For more on the history of poker in particular see McManus (2009) and Parlett (2005).

Unlike such games as roulette or blackjack, poker is a player-versus-player game—the house does not wager against its players. Instead of making money through the house edge present in house-banked casino games, poker typically generates revenues for the card room operating the game through a mechanism called the *rake*, a scaled commission fee taken by the card room operators. Generally the rake is 5 to 10 percent of the pot in each poker hand, up to a predetermined maximum amount, but there are also other non-percentage ways for a card room to collect a rake, such as charging an hourly fee or collecting a flat amount for every hand.

Since there is a large element of skill in poker—most experienced players would attest to this, and there is also a growing body of scientific evidence that supports this notion—and players are competing against each other and not the house, poker can be a positive expectation game for players who are skilled. Good poker players can win money over time, and the game attracts professional and semiprofessional players.

For most casino games—roulette, craps, baccarat, and keno—it is not mathematically possible to win in the long run. Blackjack and some video poker machines offer the opportunity for a slight player edge—one to two percent—and a few players can make money at these games with perfect play (and counting cards in the case of blackjack). It is not easy, however, and the earnings are small. Skill does play a part in these games, but luck is a large factor. In card room poker, skill is a much greater factor, and the better players will win out over the weaker players in the long run. Few people earn a living or make money consistently playing casino games, but the majority of those who do are poker players.

Because players in a poker game compete against each other and not the house, money won or lost is merely transferred from one player to another. The basic principles of gambling math still apply, however, and are at the core of the science and economics of poker, albeit intertwined with other significant elements of skill necessary for success. The sections that follow examine the various facets of the science and economics behind poker.

Types of Poker

Poker is a vying game played with standard playing cards where players bet as to who holds the best card combination by progressively raising the stakes until either there is a showdown, when the best hand wins all the stakes ("the pot"), or all but one player has given up betting and dropped out of play. In the latter case, the last person to raise the bet wins the pot without a showdown.

The game of poker can be played almost anywhere and in many different forms. Poker is really a generic name for a collection of literally hundreds of games, varying widely with respect to specific rules of play and betting structure. The variations include high games (the highest hand wins), low games (lowest hand wins), high-low split, fixed limit, spread limit, pot limit, no limit, stud, draw, games in which the hands are closed and others in which some of the players' cards are exposed. Jokers, wild cards, and special rules can be used to produce even more variations. The most popular version of poker today is Texas hold 'em, by far the most common type of poker found in card rooms and the variation played most often in tournaments, including the main event that determines the poker world champion at the World Series of Poker. Other common variants played in card rooms include seven-card stud, Omaha, lowball, and Razz.

Mathematics of Poker

The science and economics aspects of poker are, in large part, grounded in the mathematics behind the game. Although there are numerous skills necessary for excellence in poker play, understanding and using the mathematics is crucial for success in poker. Knowledge of the relevant mathematical probabilities and odds will not necessarily make a poker player more effective, but a total disregard for them will make a player bad. The following sections show how poker hand probabilities are calculated and used, with a focus on various odds in Texas hold 'em, the most popular form of poker today.

Combinatorics

Informally, combinatorics is the branch of mathematics concerned with counting methods. The two main techniques used in combinatorics are permutations and combinations. A *permutation* of a set of objects is an arrangement of those objects into a particular order. The other technique, *combinations*, refers to the number of ways that a set of objects can be selected from a group without regard to the order in which the members of the set are arranged. Both permutations and combinations are useful for computing probabilities in poker.

Permutations: The number of permutations of a set of n distinct objects is given by

$$n! = n \times (n-1) \times (n-2) \times \kappa \times 2 \times 1 \tag{21.1}$$

for integer $n > 1$. The left side of expression (21.1) is read "n factorial."

For example, there are six permutations of the set {A,B,C}, namely (A,B,C), (A,C,B), (B,A,C), (B,C,A), (C,A,B), and (C,B,A).

Combinations: The number of combinations of n objects taken x at a time (each object can appear only once, and order is not relevant) is given by

$$\binom{n}{x} = \frac{n!}{x!(n-x)!}, \tag{21.2}$$

where $n! = n \times (n-1) \times (n-2) \times \kappa \times 2 \times 1$, for integer $n >1$, and by definition $0! = 1$. The left-side of expression (21.2) is usually read "n choose x."

As an example of combinations, the number of different five-card poker hands that can be dealt from a standard deck of 52 playing cards can be found using expression (21.2) with $n = 52$ and $x = 5$, that is, "52 choose 5," which evaluates to 2,598,960. Since four of these are royal flushes, the probability of being dealt a royal flush is 4 divided by 2,598,960, or 1 in 649,740.

Poker Hand Rankings and Probabilities

Most variants of poker are based on a five-card hand-ranking system, from the strongest (least likely) to the weakest (most likely). The ranking of the types of hands in a standard 52-card deck is as follows:

1. Straight flush—any five-card sequence in the same suit.
2. Four of a kind—all four cards of the same value.

3. Full house—any three cards of the same value combined with a pair.
4. Flush—any five cards of the same suit that are not in sequence.
5. Straight—any five cards in sequence but not in the same suit.
6. Three of a kind—any three cards of the same value.
7. Two pair—any two separate pairs.
8. One pair—any two cards of the same value.
9. High card—a hand with none of the above combinations (valued by its highest card).

In standard poker suits are not ranked. If two hands are identical apart from the suits, they are considered to be equal. If there are two highest equal hands in a showdown, the pot is split between them.

Poker hand rankings are determined by their likelihood of occurrence when five cards are dealt at random from a shuffled deck of 52 cards. The highest-ranking hand is the least likely; the second highest-ranking hand is the second least likely, and so on. The probabilities of the various five-card poker hands can be determined using combinations, permutations, and the usual probability laws.

First note since there are $\binom{52}{5} = 2{,}598{,}960$ possible five-card hands that can be dealt from a single deck of 52 cards, the probability associated with a particular type of poker hand can be computed by counting the possible number of such hands and dividing by 2,598,960. Table 21.1 shows the five-card poker hand rankings and associated probabilities of occurring in a five-card hand dealt from a deck of 52 cards.

Further Probability Considerations

The basic five-card poker hand probabilities presented in table 21.1 serve as a reference point, but further knowledge of the probabilities and odds associated with the play of the hands is required for excellence in playing poker. To illustrate, consider the most popular form of poker today, Texas hold ’em.

Texas Hold ’em

Typically Texas hold ’em is played with 8 to 10 players, though it is not unusual to have fewer than 8. Before any cards are dealt in a hold ’em hand, two players post “blind” bets—a “small blind” by the player to the dealer’s immediate left and a “big blind” by the next player to the left of the small blind. These blinds are forced bets made before the players see their cards and are used to start the pot and stimulate action. The small blind is usually equal to one-half the amount of the big blind. Since the deal rotates around the table (in a casino where the dealer is not a player, a *button* used to signify the nominal dealer rotates after each hand), all players participate equally in the posting of any forced blind bets.

Table 21.1 Poker Hand Rankings and Probabilities

Hand	Number of Ways	Number	Probability	≈1 in
1. Royal flush	$\binom{4}{1}$	4	0.00000154	649,740
2. Straight flush*	$\binom{9}{1}\binom{4}{1}$	36	0.00001385	72,193
3. Four of a kind	$13 \cdot \binom{48}{1}$	624	0.00024010	4,165
4. Full house	$13 \cdot \binom{4}{3} \times 12 \cdot \binom{4}{2}$	3,744	0.00144058	694
5. Flush	$4 \cdot \left[\binom{13}{5} - 10\right]$	5,108	0.00196540	509
6. Straight	$10 \times \left(\frac{20\cdot16\cdot12\cdot8\cdot4}{5!}\right) - 40$	10,200	0.00392465	255
7. Three of a kind	$13 \times \left[\binom{4}{3} \cdot \frac{48\cdot44}{2!}\right]$	54,912	0.02112845	47
8. Two pairs	$13 \cdot 12 \times \left\{\left[\binom{4}{2} \cdot \binom{4}{2} \div 2!\right] \times 44\right\}$	123,552	0.04753902	21
9. One pair	$13 \times \binom{4}{2} \times \frac{48\cdot44\cdot40}{3!}$	1,098,240	0.42256903	2.4
10. High card	$\left[\binom{13}{5} - 10\right]\left[\binom{4}{1}^5 - 4\right]$	1,302,540	0.50117739	2.0
Total	$\binom{52}{5}$	2,598,960	1.00000000	

* Excluding royal flushes.

After the blinds are posted, each player is dealt two cards face down, followed by a round of betting. After this first round of betting, three community cards (the *flop*) are exposed, followed by a second round of betting. After the second betting round, a fourth community card (the *turn*) is exposed, followed by another round of betting, and then the fifth and final community card (the *river*) is exposed, followed by a final round of betting.

Each betting round except for the first begins with the first active player to the left of the dealer (or in a game dealt by a house dealer, the first active player to the left of the button used to indicate dealer position). Because the first two players to the left of the dealer (or button) have already acted by putting in blind bets, the player one to the left of the big blind is the first with any choices on the first betting round. In general, when it is a player's turn to act, that player may choose to (a) *bet*—if there has been no previous bet made in the round, (b) *call*—match the current outstanding bet, (c) *raise*—match the current outstanding bet and simultaneously make an additional bet, (d) *check*—if there is no current outstanding bet, or (e) *fold*—withdraw from the hand, leaving all previously wagered money in the pot. Subject to the usual maximum of three (or sometimes four) raises per round, betting rotates clockwise until each player who has not folded has put the same amount of money into the pot for the current round or until one player remains. In the latter case, this player is the winner and is awarded the pot without having to reveal the cards.

If more than one player remains in the hand after the final round of betting, there is a *showdown* and the player with the best five-card poker hand—formed by using any combination of the five community cards and the player's own two hole cards—wins the pot. If a tie occurs, the pot is split.

Hold 'em Probabilities

Suppose in a Texas hold 'em game you held the ace and seven of diamonds and the flop contained the two of diamonds, the jack of diamonds, and the five of spades. At this point you have four to a diamond flush and would make an ace-high flush (and likely win the hand) if a diamond fell on the turn or the river (or both). What is the probability you will make your flush?

The easiest way to compute the probability you will make your flush is to first compute the probability you don't make the flush and then subtract this value from one. Since you have seen five cards—your two hole cards and the three flop cards—there are 47 remaining unseen cards, of which 9 are diamonds and 38 are non-diamonds. Thus the probability you will not make your flush with the turn card is 38/47, and the probability you will not make the flush with the river card, given that you missed the flush with the turn card, is 37/46. Using the multiplication law (dependent events), the probability you will not make your flush is, then, $(38/47)(37/46) = .650$. Thus the probability you will make your flush (on either the turn or the river or both) is $1 - .650 = .350$. Put another way, the odds against making your flush are 1.86 to 1.

Note that in the above flush draw example, the probability that the flush is made with one card to come depends on whether we look at making the flush on the turn card or, having not made the flush on the turn, the river card. The probability of making the flush on the turn is $9/47 = .191$, for odds against of about 4.2 to 1; if you don't make

Table 21.2 Texas Hold 'Em Probabilities and Odds

		Two Cards to Come		One Card to Come*	
Outs	Drawing to (Example)	Probability	Odds Against	Probability	Odds Against
21		69.9%	.43 to 1	45.7%	1.19 to 1
20		67.5%	.48 to 1	43.5%	1.30 to 1
19		65.0%	.54 to 1	41.3%	1.42 to 1
18		62.4%	.60 to 1	39.1%	1.56 to 1
17		59.8%	.67 to 1	37.0%	1.71 to 1
16		57.0%	.75 to 1	34.8%	1.88 to 1
15	Open straight flush draw	54.1%	.85 to 1	32.6%	2.07 to 1
14		51.2%	.95 to 1	30.4%	2.29 to 1
13		48.1%	1.08 to 1	28.3%	2.54 to 1
12		45.0%	1.22 to 1	26.1%	2.83 to 1
11		41.7%	1.40 to 1	23.9%	3.18 to 1
10		38.4%	1.60 to 1	21.7%	3.60 to 1
9	Four flush (flush)	35.0%	1.86 to 1	19.6%	4.11 to 1
8	Open straight draw (straight)	31.5%	2.18 to 1	17.4%	4.75 to 1
7		27.8%	2.59 to 1	15.2%	5.57 to 1
6		24.1%	3.14 to 1	13.0%	6.67 to 1
5		20.4%	3.91 to 1	10.9%	8.20 to 1
4	Gutshot straight (straight)	16.5%	5.07 to 1	8.7%	10.50 to 1
3		12.5%	7.01 to 1	6.5%	14.33 to 1
2	Pocket pair (three of a kind)	8.4%	10.88 to 1	4.3%	22.00 to 1
1	Three of a kind (four of a kind)	4.3%	22.50 to 1	2.2%	45.00 to 1

*River (fifth community card) to come.

your flush on the turn, the probability you make it on the river is 9/46 = .196, for odds against of 4.1 to 1.

The flush draw described above is an example of a post-flop hold 'em situation in which a player has nine *outs*—cards that will make the desired hand (in this case a flush). A similar situation but drawing to an inside straight offers the player only four outs, while an outside straight draw offer eight outs (ignoring the possibility of making other hands, such as a pair, three of a kind, etc.). Table 21.2 shows probabilities and odds for making hands with two cards to come and one card to come for specified numbers of outs, with the one-card-to-come probabilities and odds computed assuming that the one card is the final river card.

Expected Value

To fully utilize the mathematics of gambling in poker play, the odds and/or probabilities in table 21.2 (or analogous values for other poker games) would need to be balanced against the amount of money that would be won or lost. This comparison of the winning odds and the pot odds is at the heart of mathematical expectation, or, expected value.

Proper poker play requires the ability to correctly evaluate the expected value of the various alternative decisions (bet, fold, call, raise, re-raise, etc.). The expected value of a decision is a function of both the probability of the possible outcomes of the action and the values of these outcomes. In poker, the value of an outcome is the amount of money won or lost.

The expected value (EV) of a wager can be computed by multiplying the possible payoffs by their probabilities and then summing the resulting terms. Mathematically

$$EV = \sum (Win_i \times p_i) \tag{21.3}$$

where Win_i is the net win associated with outcome i and p_i is the probability of Win_i. The EV for a bet represents the amount of money the bettor will win or lose on average, or in the long run, from making the specified bet.

As a simple example, suppose you pay \$1 to play a game where a single card is drawn at random from a standard deck of playing cards and if the selected card is a spade you will win even money. That is, you will be given \$1 in addition to the \$1 you paid to play the game. It should be clear that this is not a smart bet, as you will win only once every four times, and therefore you will be, on average, down two dollars for every four times you play this game. Your expected value for this wager is negative \$0.50.

$$EV = (+\$1)(1/4) + (-\$1)(3/4) = -\$0.50.$$

That the EV is negative means you will lose \$0.50 on average each time you make this wager. If, on the other hand, the net payoff in this game of spades is \$4.00, the EV would be \$0.25—you will win \$0.25 per time you make this bet on average—and the wager is now favorable.

The EV is a function of both the probability of winning and the amount you will win. Even if the probability of winning is small—if the payoff is large enough—the EV will be positive, making the bet favorable. The section below describes how expected value is used in poker.

EV and Poker

If we consider only wagers where the outcome is either a single winning value with probability p or a loss equal to the amount bet with probability $(1 - p)$, then (21.3) simplifies to

$$EV = (Net\ Win)p - (Bet)(1 - p) = (Net\ Win + Bet)p - Bet. \tag{21.4}$$

Expression (21.4) essentially says that the EV in this situation is equal to the expected *return* of the bet, where the return is the amount paid back to the player *for* (rather than in addition to) the cost of playing (i.e., the amount bet), minus the cost of playing.

When contemplating the efficacy of a making a bet in poker, it can be viewed it in the above outlined framework—that is, the bet can result in either the player winning the pot if he makes his hand or the player losing the amount bet. If p is the probability the player makes his hand, then (21.4) becomes

$$EV = (Pot + Bet)p - (Bet) = (Pot)p - Bet(1 - p). \tag{21.5}$$

From (21.5), the EV of the contemplated bet is positive when

$$\frac{Pot}{Bet} > \frac{1-p}{p}.$$

The left side of (21.6) is the ratio of the size of the pot to the amount of the bet needed—the *pot odds*; the right side is the ratio of the probability of losing to the probability of winning—the odds against winning. In other words, the EV is positive when the pot odds are greater than the odds against winning.

To illustrate, consider the following generic example. Suppose you need to call a \$10 bet to stay in the hand, and the current size of the pot is \$30. Further suppose that the probability you will win if you call the bet is .20. Then, $Pot = \$30$, $Bet = \$10$, $p = .20$, and $(1 - p) = .80$. The odds against winning, $(1 - p)/p$, are 4 to 1, and the pot odds, Pot/Bet, are 3 to 1. Because the odds against winning are greater than the pot odds the EV is negative ($EV = -\$2$), and from a purely mathematical perspective you will lose money in the long run by calling this bet. One in five times ($p = .20$) you make this bet you'll win \$30, and the other four out of five times you'll lose \$10, for a total loss of \$10, an average of \$2 per bet. Other considerations aside, you should fold in this situation. On the other hand, if in the same situation the current size of the pot is \$50.00, the pot odds (5 to 1) would be greater than the odds against winning, and EV analysis would recommend making the bet.

The comparison between what the pot offers and the chances of winning or losing lies at the heart of the mathematics of poker. Although the outcome of any given hand is uncertain, and a bet with positive expectation could lose while one with a negative expectation could win, when faced with a decision to bet, fold, or raise, a player who consistently chooses the action with the largest EV will come out a winner. A player who consistently makes decisions with negative expectations will, over the long run, lose money.

As another example, consider the hold 'em flush draw situation mentioned previously: You have four to a flush after the flop and know the odds against making your flush (on the turn or the river) are 1.86 to 1. If you are contemplating calling a bet, the simple EV analysis outlined above would tell you to call the bet if the pot is offering you more than 1.86 to 1. The EV analysis presented here is, of course, a simplification. Highly skilled players will also be familiar with and consider *effective odds*, *implied odds*, *reverse-implied odds*, and other more advanced concepts.

Game Theory and Poker

In their seminal book on game theory, *The Theory of Games and Economic Behavior* (1944), John von Neumann and Oskar Morgenstern devote an entire chapter to poker. As one author asserts (Nasr 1998, 13–14), von Neumann believed that poker and economics share a key connection.

> A seemingly trivial and playful pursuit like poker, von Neumann argued, might hold the key to more serious affairs for two reasons. Both poker and economic competition require a certain type of reasoning, namely the rational calculation of advantage and disadvantage based on some internally consistent system of values ("more is better than less"). And in both, the outcome for any individual actor depends not only on his own actions, but on the independent actions of others.

Game theory is a field of mathematics that models conflicts and strategic interactions between competing agents. It attempts to solve games, help players avoid mistakes, and discover mathematically best strategies. A game in the everyday sense might be defined as a competitive activity in which players contend with each other according to a set of rules, a pastime or diversion. In game theoretic sense a game is any rule-governed situation with a well-defined outcome characterized by strategic independence. Firms competing in a market are players in a game; countries involved in international trade or finance negotiations are players in a game; so are poker players seated around a poker table. Though first applied to the theoretical study of economics, game theory has broadened in scope to include applications in such diverse fields as international relations, social science, political science, military science, psychology, biology, and poker.

A particularly important and informative example of how game theory can be used in poker is the determination of an optimal bluffing strategy, the strategy through which you will average the same amount won regardless of whether your opponent calls or folds. Game theory shows that the optimal bluffing strategy is a *mixed strategy*—one where you *randomly* choose between multiple possible actions (here, bluffing or folding) a predetermined percentage of the time. As David Sklansky (1999, 190) summarized

> When using game theory to decide whether to bluff, you must determine the pot odds your opponent is getting if you bet and then randomly bluff in such a way that the odds against your bluffing are identical or almost identical to your opponent's pot odds. If your opponent is getting 5-to-1, the odds against your bluffing should be 5-to-1. By playing this way, you give your opponent no correct decision. He does just as well—or badly—in the long run by calling or folding.

Game theory can also be used to decide whether to call a possible bluff (assuming your hand can beat only a bluff and assuming your judgment doesn't give you a hint). The appropriate mixed strategy here is to make the ratio of your calls to your folds the same as the odds your opponent is getting on a bluff. For example, if your opponent is getting 3-to-1 odds on a bluff, then randomly calling three out of four times will make that bluff unprofitable.

Many scientists have used a game theoretic approach to study poker. These approaches typically use game theory to carefully analyze simplified models, or specific aspects or situations, of real poker in order to try to understand the fundamental nature of the game. The interested reader is referred to chapter 19 of von Neumann and Morgenstern (1944), chapter 19 of Sklansky (1999), chapter 22 of Ethier (2010), and Ferguson and Ferguson (2007) for illustrative examples of such analyses. Chen and Ankenman (2006) combines mathematical analysis with a heavy dose of game theory in an attempt to create "near-optimal" strategies and enhance the player's ability to win money at real poker.

Artificial Intelligence and Poker

As a game of imperfect information and nondeterministic dynamics where multiple competing agents deal with probabilistic knowledge, risk assessment, and possible deception, poker is both fertile ground and a challenging problem for artificial intelligence research. By modeling various elements of skill and appealing to advanced mathematical and statistical techniques, such as game theory and Bayesian probabilistic models, scientists have had some success at designing champion-level artificial intelligence poker agents.

One of the leading research groups in this area is the Computer Poker Research Group (CPRG) at the University of Alberta. In designing intelligent poker bots to play Texas hold 'em, the CPRG has specifically attempted to incorporate six skill-related aspects:

- Hand strength assesses how strong your hand is in comparison to what your opponents may hold and is computed on the flop, turn, and river.
  - Minimum skill—a function of your cards and the community cards.
  - Moderate skill—also takes into account the number of players still in the game, position at the table, and the history of betting in the hand.
  - Maximum skill—also considers the different probabilities for each possible opponent hand, based on the likelihood of each hand being played to the current point in the game. Skill levels can be improved even further by varying hidden hand probabilities for each player depending on that player's model of play.
- Hand potential assesses the probability of the hand improving (or being overtaken) as additional community cards appear.
  - Minimum skill—a function of your cards and the community cards.

  - Maximum skill—also considers the number of players still in the game, position at the table, history of betting in the hand, different probabilities for each possible opponent hand, and each player's model of play.
- Betting strategy determines whether to fold, call/check, or bet/raise.
  - Minimum skill level—based on hand strength.
  - Maximum skill level—also considers hand potential, pot odds, bluffing, opponent modeling, and unpredictability.
- Bluffing makes it possible to profit from weak hands and can create a false impression about your play that can improve the chances of winning subsequent hands.
  - Minimum skill level—bluffing a fixed percentage of all hands.
  - Maximum skill level—incorporate opponent modeling.
- Opponent modeling determines a likely probability distribution for opponent's hidden cards or betting strategy.
  - Minimum skill level—uses a single model for all opponents in a given hand.
  - Maximum skill level—modifies the probabilities based on a classification of each opponent (e.g., weak or strong, passive or aggressive), betting history, and collected statistics.
- Unpredictability makes it difficult for opponents to form an accurate model of your strategy by varying playing style over time to induce opponents to make mistakes based on inaccurate models.

In 2008 a CPRG-designed poker robot dubbed Polaris played six matches of two-player limit Texas hold 'em against some of the best players in the world and managed to outperform them by $200,000, winning three matches, losing two, and tying one. The previous year Polaris played a duplicate match consisting of four 500-hand sessions against two top professional poker players and finished with a record of 1–2–1.

A Caveat—House-Banked Poker Games

The size of the skill component in Texas hold 'em is such that it can be fairly characterized as a game of skill—that is, a game in which skill predominates over chance. This is true for many other player-versus-player poker games, such as seven-card stud and Omaha. But the same cannot be said of house-banked poker games, such as Caribbean stud, three-card poker, and video poker machines. The player of these games is playing against the house and an associated house advantage, and while a certain element of skill is necessary to minimize losses—except for a few types of video poker the player in these games is always at a disadvantage—mastering optimal strategy in these house-banked poker derivatives does not require the level of skill as in such player-versus-player poker games as Texas hold 'em.

For example, the reported advantages or payback percentages for a video poker machine assume perfect optimal strategy and maximum coin played, but the effective house advantage depends on the skill of the player. The reported house advantage can be achieved only if playing optimal strategy, which varies depending on the particular machine or game. Although playing video poker can be challenging, the skill levels of video poker and Texas hold 'em are significantly different. Skills such as psychological insight, assessment of competition, the ability to read hands, recognize tells, and exploit position as well as money management are conspicuously absent in video poker. Another notable difference is that almost all video poker games, even when played at optimum strategy, result in the player losing over time.

Skill in Poker

Scientific studies on the relative roles of skill versus chance in poker have used a variety of approaches to examine the issue, and the body of research on the subject points to the conclusion that poker is a game predominately of skill. The following summarizes the conclusions from some of the key studies:

- In an analysis of more than one billion hands of real poker, Robert Hannum, Matthew Rutherford, and Teresa Dalton (2012) presented a method for isolating the effect of systemic chance by taking advantage of knowledge of players' hole cards to predict their return on investment in a Texas hold 'em game devoid of all elements of skill. Using a regression approach to compare players' actual returns on investment (ROIs) to their expected returns on investment in the chance-only hold 'em game, they found that only 0.03 percent of the variation in players' actual ROIs could be attributed to systemic chance and so 99.97 percent could be attributable to skill. They further found that 85.2 percent of all hands were resolved without a showdown, and of the 14.8 percent that did go to showdown, nearly half (46.8%) were won by a player at the table who did not have the best hand (because the player with the best hand folded prior to showdown).
- Steven Levitt and Thomas Miles (2011) used World Series of Poker data to address the issue of skill versus chance. Having designated skill of players based on prior year performance, they considered the average rate of return on investment in the poker tournament. Their findings revealed that skilled players had, on average, a 30 percent ROI while all other players had a loss of 15 percent. They used this difference in ROI for skilled players over all other players as support for the proposition that poker is a game of skill.
- Presenting evidence of skill differentials among 899 poker players finishing in one of the final two tables in 81 high-stakes Texas hold 'em tournaments between 2001 and 2005, Rachel Croson, Peter Fishman, and Devin Pope (2008) concluded (a) that there appears to be a significant skill component to poker: previous finishes

in tournaments predict current finishes; and (b) that skill differences among top poker players are similar to skill differences across top golfers.

- Comparing players who were taught strategies based on expert opinion with other players who were taught no strategies, Michael Dedonno and Douglas Detterman (2008) found that participants who were instructed outperformed those who were not instructed. They concluded that the skill involved in poker is complex and that luck (random factors) disguises the fact that poker is a game of skill and also declared the unequivocal finding that poker is a game of skill.
- Based on the rationale that chance elements cancel out in the long run while skill elements do not, Ingo Fiedler and Jan-Philipp Rock (2009) proposed a critical repetition frequency—the threshold of repetitions at which a game becomes predominantly influenced by skill rather than by chance. Analyzing data from an empirical survey of 51,761 poker players, they concluded that poker lies in the continuum between being a game of chance and being a game of skill and that for their sample, poker is a game of skill.
- In a mathematical analysis supported by computer simulations of a random player versus a skilled player in Texas hold 'em, Hannum Anthony Cabot (2009) found that the skilled player won nearly 97 percent of all hands and 1.6 big blinds per hand. Additional full-table simulation games of Texas hold 'em and seven-card stud with highly skilled and lesser skilled players showed that the highly skilled players convincingly beat unskilled players. The authors concluded that poker is a game predominantly of skill with the skill elements expressed through the player's betting strategy—that is, the decisions on whether to check, bet, call, raise, or fold.
- In computerized experiments in a simplified version of stud poker comparing 12 different strategies, ranging from a simple strategy that plays randomly (a player with no skill) through progressively more sophisticated strategies utilizing varying degrees of skill, Patrick Larkey et al. (1997) found a wide range of outcomes across the different strategies, with better outcomes generally associated with higher degrees of skill. They also determined that simple random strategy (zero skill) is the worst overall performer in terms of winnings.
- Using a game theoretic approach to study the relative roles of skill and chance in games, Peter Borm and Ben van der Genugten (2001) concluded that the level of skill in the three popular variants of poker, seven-card-stud, Texas hold 'em, and draw poker, is greater than that in roulette, craps, and blackjack.

Summary

The science and economics of poker derive from the mathematics underlying the game and the various and considerable elements of skill needed to be successful over the long term. The mathematics of probability, odds, and expected value are intimately associated with proper decision-making, a key objective in poker. Principles of game theory shed further light on optimal or near-optimal strategies for winning money. A spate of recent scientific studies on the subject provides compelling evidence that skill is the major factor, and predominates over chance, in determining the outcome in poker.

References

Borm, Peter, and Ben van der Genugten. 2001. On a relative measure of skill for games with chance elements. *Trabajos de Investigacion Operative* [TOP—Official Journal of the Spanish Society of Statistics and Operations Research] 9(1):91–114.

Chen, Bill, and Jerrod Ankenman. 2006. *The mathematics of poker*. Pittsburgh, Pa.: ConJelCo LLC.

Croson, Rachel, Peter Fishman, Devin G. Pope. 2008. Poker superstars: Skill or luck? *Chance* 21(4):25–28.

Dedonno, Michael A., and Douglas K. Detterman. 2008. Poker is a skill. *Gaming Law Review* 12(1):31–36.

Ethier, Stewart N. 2010. *The doctrine of chances: Probabilistic aspects of gambling*. Berlin: Springer-Verlag.

Ferguson, Chris, and Tom Ferguson. 2007. The endgame in poker. In *Optimal play: Mathematical studies of games and gambling*, edited by Stewart N. Ethier and William R. Eadington. Reno: Institute for the Study of Gambling and Commercial Gaming, University of Nevada, Reno.

Fiedler, Ingo, and Jan-Philipp Rock. 2009. Quantifying skill in games—Theory and empirical evidence for poker. *Gaming Law Review and Economics* 1(3):50–57.

Hannum, Robert C., and Anthony N. Cabot, A. (2009). Toward legalization of poker: The skill vs. chance debate. *UNLV Gaming Research & Review Journal* 13(1):1–20.

Hannum, Robert, Matthew Rutherford, and Teresa Dalton. 2012. Economics of poker: The effect of systemic chance. *Journal of Gambling Business & Economics* 6(1):25–48.

Larkey, Patrick, Joseph B. Kadane, Robert Austin, and Shmuel Zamir. 1997. Skill in games. *Management Science* 43(5):596–609.

Levitt, Steven D., and Thomas J. Miles. 2011. The role of skill versus luck in poker: Evidence from the World Series of Poker. National Bureau of Economic Research Working Paper No. 17023. Cambridge, Mass.: National Bureau of Economic Research; www.nber.org/papers/w17023.

McManus, James. 2009. *Cowboys full: The story of poker*. New York: Farrar, Straus, and Giroux.

Nasar, Sylvia. 1998. *A beautiful mind*. New York: Simon & Schuster.

Parlett, David. 1991. *A history of card games*. Oxford: Oxford University Press.

——. 2005. *A history of poker*; www.pagat.com/vying/pokerhistory.html.

Sklansky, David. 1999. *The theory of poker*. Henderson, Nev.: Two Plus Two Publishing.

von Neumann, John, and Oskar Morgenstern. 1944. *The theory of games and economic behavior*. Princeton, N.J.: Princeton University Press.

CHAPTER 22

THE KELLY CRITERION WITH GAMES OF CHANCE

LEONARD C. MACLEAN AND WILLIAM T. ZIEMBA

1 Introduction

In games of chance, a bettor must determine how much capital to bet in risky games at each play, with a focus on the accumulation of maximum capital after a sequence of plays. An important betting strategy is based on the Kelly criterion, through which the expected logarithm of wealth is maximized. (Kelly 1956.) Log utility dates to 1738, when Daniel Bernoulli postulated that marginal utility was monotone increasing but declining with wealth and, specifically, was equal to the reciprocal of wealth, w, which yields the utility of wealth, $u(w) = log(w)$. Prior to this it was assumed that decisions were made on an expected value or linear utility basis. This idea ushered in declining marginal utility or risk aversion or concavity which is crucial in investment decision-making. In his paper, Bernoulli also discussed the St. Petersburg paradox and how it might be analyzed using $log(w)$. This problem concerns how much to pay for the following gamble:

> A fair coin with $\frac{1}{2}$ probability of heads is repeatedly tossed until heads occurs, ending the game. The investor pays c dollars and receives in return 2^{k-1} with probability 2^{-k} for $k = 1, 2, \ldots$ should a head occur. Thus after each succeeding loss, assuming a head does not appear, the bet is doubled to $2, 4, 8, \ldots$ and so on. Clearly the expected value is $\frac{1}{2} + \frac{1}{2} + \frac{1}{2} + \cdots$ or infinity with linear utility.

Robert Bell and Thomas Cover (1980) argued that the St. Petersburg gamble is attractive at any price c but that the investor wants less of it as $c \to \infty$. The proportion of the investor's wealth invested in the St. Petersburg gamble is always positive but decreases with increasing cost c. The rest of the wealth is in cash.

Bernoulli offered two solutions since he felt that this gamble is worth a lot less than infinity. In the first solution, he arbitrarily set a limit to the utility of very large payoffs. Specifically any amount over 10 million was assumed to be equal to 2^{24}. Under that bounded utility assumption the expected value is

$$\frac{1}{2}(1)+\frac{1}{4}(2)+\frac{1}{8}(4)+\cdots+\left(\frac{1}{2}\right)^{25}\left(2^{24}\right)+\cdots=12+\text{the original } 1=13$$

When utility is log the expected value is

$$\frac{1}{2}\log 1+\frac{1}{4}\log 2+\frac{1}{2}\log 4+\cdots=\log 2=0.6915.$$

As Karl Menger (1934) pointed out, the log, the square root, and many other, but not all, concave utility functions eliminate the original St. Petersburg paradox, but it does not solve one where the payoffs grow faster than 2^n. So if log is the utility function, one can create a new paradox by having the payoffs increase at least as fast as log reduces them so one still has an infinite sum for the expected utility. With exponentially growing payoffs one has

$$\frac{1}{2}\log(e)+\frac{1}{4}\log\left(e^2\right)+\cdots=\infty.$$

The super St. Petersburg paradox, in which even $ElogX=\infty$, is examined in Cover and Thomas (2006, 181, 182), where a satisfactory resolution is reached by looking at relative growth rates of wealth.

J. L. Kelly Jr. (1956) is credited with using log utility in gambling and repeated games. His analysis uses Bernouli trials. He established that log is the utility function which maximizes the long run growth rate, and is myopic in the sense that period by period maximization is optimal. Latané (1959) introduced log utility as an investment criterion to finance independent of Kelly's work. Leo Breiman (1961) established the basic mathematical properties of the expected log criterion: (i) wealth for the Kelly strategy overtakes almost surely that of any other essentially different strategy as the horizon becomes infinitely distant, (ii) the strategy attains arbitrarily large wealth goals faster than any other strategy, and (iii) with a fixed opportunity set a fixed Kelly strategy is optimal.

In an economy with one log bettor and with all other bettors having essentially different strategies, the log bettor will eventually get all of the economy's wealth (Hens and Schenk-Hoppé 2005). The drawback of log, with its essentially zero Arrow–Pratt absolute risk aversion, is that in the short run it is the most risky utility function one would ever consider. Since there is essentially no risk aversion, the wagers it suggests are very large and typically undiversified. Simulations show that log investors have much more final wealth most of the time than do those using other strategies, but those investors can essentially go bankrupt a small percentage of the time, even facing very favorable choices (Ziemba and Hausch 1986). One way to modify the growth-security profile is to use either ad hoc or scientifically computed fractional Kelly strategies that blend the log optimal portfolio with cash. For instance, a fractional Kelly strategy will keep accumulated capital above a specified wealth path with high probability given log normally distributed payoffs. This is equivalent to using a negative power utility function whose coefficient (analogous to a risk aversion index) is determined by the fraction and vice versa. Thus one moves the risk aversion away from zero to a higher level. This results in a smoother wealth path but has less growth. For non-lognormal payoff distributions, the fractional Kelly is an approximate solution to the optimal risk-return trade-off, but the approximation may be inaccurate (MacLean, Thorp, and Ziemba 2011).

2 Games of Chance and the Kelly Strategy

A game of chance is a game whose outcome depends on a random process. Consider a set of K games whose outcomes are stochastic, that is, they are defined on the probability space (Ω, B, P). Assume there is a payoff or return from a play of game i, given the outcome $\omega \in \Omega$, defined by by $r_i(\omega), i = 1, \ldots, K$. A gamble is a bet placed on the outcome of the game. For a unit of capital bet on game i the return is

$$R_i(\omega) = 1 + r_i(\omega),\ i = 1, \ldots, K. \tag{22.1}$$

In the gambling market games are played at points in time and the return on bets leads to the accumulation of capital for a bettor. In the analysis of betting strategies, the following structure is assumed:

(a) All games have limited liability.
(b) There are no playing costs, taxes, or problems with indivisibility of capital.
(c) Capital can be borrowed or lent at a risk-free interest rate r.
(d) Negative betting (borrowing against the game outcome) is allowed.

The returns on outcomes from the K games generate random vector $R' = (R_0, \ldots, R_K)$, where $R_0 = 1 + r$. Suppose a bettor has w_t units of capital at time t, with the proportions bet in each game given by and $\tilde{X} = (x_1(t), \ldots, x_K(t))'$. If the fraction of capital not bet is $x_0(t)$ and earns the risk-free rate, then a betting strategy at time t is the vector process

$$X(t) = (x_0(t), \tilde{X}'(t))'. \tag{22.2}$$

Given the bets $w_t X(t)$ at time t, the accumulated capital at time $t + 1$ is

$$W(t + 1) = w_t R' X(t). \tag{22.3}$$

For a sequence of T plays of the set of games starting with capital w_0, wealth at time T is

$$W(T) = w_0 \prod_{t=1}^{T} R' X(t). \tag{22.4}$$

Alternatively wealth is

$$W(T) = w_0 \left(\exp \left[\frac{1}{T} \sum_{t=1}^{T} \ln(R' X(t)) \right] \right)^T. \tag{22.5}$$

The exponential form highlights the *growth rate* with the strategy $X = (X(1), \ldots, X(T))$,

$$G_T(X) = \frac{1}{T} \sum_{t=1}^{T} \ln(R' X(t)). \tag{22.6}$$

If the distribution of accumulated capital (wealth) at the horizon is the criterion for deciding on an investment strategy, then the rate of growth of capital becomes the determining factor when the horizon is distant. Consider then the average growth rate

$$EG_T(X) = \frac{1}{T} \sum_{i=1}^{T} E \ln(R'(t) X(t)). \tag{22.7}$$

The case usually discussed is when the incremental returns are serially independent. So the maximization of $EG_T(X)$ is

$$\max \left\{ E \ln \left(R'(t)\, X(t) \right) \right\}, \tag{22.8}$$

separately for each t. If the returns distribution is the same for each t, a fixed strategy holds over time. The strategy that solves (22.8) is called the Kelly or optimal growth strategy.

2.1 Kelly Formulas

The Kelly strategy is optimal for the asymptotic growth rate criterion, but there are other measures of performance of interest. In table 22.1, some measures are defined, with a classification by wealth and time. In the definitions in the table, the notation $\tau(W(X) \geq u)$ is the first passage time to the set $[u, \infty)$.

The wealth and time dimensions of the stochastic process $\{W(X), t \geq 0\}$ are alternative perspectives, with the time component emphasizing growth *speed* and wealth emphasizing growth *magnitude*. The distributions for the random quantities are illustrated in figure 22.1.

Table 22.1 Performance Measures

Criterion	Mean	Percentile
Wealth	$\phi_T(X) = E\left\{in[W_T(X)]^{\frac{1}{T}}\right\}$	$\gamma_T(X) = pr[ln(W_T(X)) \geq 1n(w_T)]$
Time	$\eta_u(X) = E\{\tau(W(X)) \geq u\}$	$\beta_{l,u}(X) = pr[\tau(W(X)) \geq u < \tau(W(X)) \leq l]$

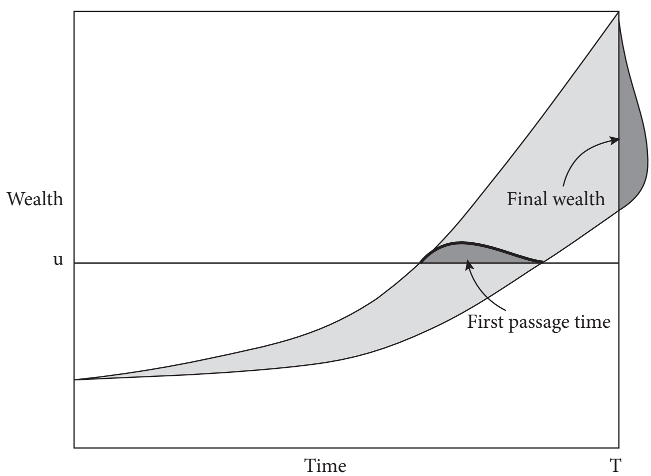


FIGURE 22.1 Wealth and time dimensions of wealth process.

The Kelly strategy is defined by the expected value, but the risk characteristics are significant as well. Risk is defined by the chance of falling short of targets. By the wealth criterion, a value-at-risk W_T is specified at the horizon T, so risk is defined by a VaR condition. In the case of first passage time, the chance of achieving a desired wealth target before falling to an undesirable level is assessed.

The Kelly strategy is optimal for the expected values φ and η but may not do well on risk measures γ and β. These performance measures have been considered in several papers in the finance and probability literature (see, for example, Hanoch and Levy (1969), Hakansson (1970), Kallberg and Ziemba (1981, 1983, 1984), Grauer and Hakansson (1985, 1987), Jorion (1986, 1987), Browne (1997), Dohi et al. (1995), Ethier and Tavaré (1983), Luenberger (1993, 2009), MacLean and Ziemba (1991, 1999, 2000), MacLean, Ziemba, and Blazenko (1992), Hakansson and Ziemba (1995), Stutzer (2003) and MacLean, Ziemba and Li (2005)). The evaluation of these measures is the basis for selecting an appropriate betting strategy that satisfies our preferences. If the strategy has constant investment proportions, which includes the Kelly, then standard results on random walks (discrete time) and diffusions (continuous time) can be adapted to provide computational formulas for the measures.

Discrete Time

The discrete time formulas for the measures were developed in MacLean, Ziemba, and Blazenko (1992). They apply to any constant proportions strategy. They are presented in table 22.2 using the following notation:

(i) $X = (x_0, x_1, \ldots, x_K)'$ is the constant betting policy.
(ii) $J(X) = log(R'X)$.
(iii) $Z_T(X) = \sum_{t=1}^{T} J_t(X)$, where $J_0(X) = log\ W_0$.
(iv) $J_M(X) = inf\ J(X), J_M(X) = sup\ J(X)$.
(v) $l^* = \log(l),\ u^* = \log(u)$, for lower and upper targets l, u.
(vi) $L^*(X) = l^* + J_m(X),\ U^*(X) = u^* + J_M(X),\ U_*(X) = u^* - J_m(X)$.
(vii) θ is the non-unit root of $E\theta^{J(x)} = 1$, where E is expectation.
(viii) $\Phi[\bullet]$ is the cumulative normal distribution.

Continuous Time

The performance measures for repeated plays of games can be approximated by continuous time formulas. That is, the random walk of log wealth is approximated by a diffusion. For the evaluation of measures in the continuous time formulation of the log wealth process we refer to Dohi et al. (1995). Consider the following notation:

(i) The total bets in risky games is $\lambda(X) = \sum_{i=1}^{K} x_i$.
(ii) The risky games are combined into a single game fund with instantaneous rate of return $R = \sum_{i=1}^{k} a_i R_i$, where $a_i = \frac{x_i}{\lambda(X)}$.
(iii) The mean and variance of the instantaneous rate of return on the game fund are $\mu\left(\tilde{X}\right)$ and $\sigma^2\left(\tilde{X}\right)$, respectively.
(iv) $D(X) = (\mu(X) - r)\lambda - \frac{\sigma^2(\widehat{X})\lambda^2(x)}{2}$, where it is assumed that $D(X)\ > 0$.
(v) Φ [• "•"] is the cumulative normal distribution.

The expressions in tables 22.2 and 22.3 are similar and usually yield similar values. One advantage of the discrete time formulas is the flexibility over the rate of return distributions on risky assets, since we can work with any distribution with finite support. For the exact continuous time model, the return distributions are assumed to be lognormal, and that may not give an acceptable approximation. However, in the continuous time approximation, where the payoffs on risky games have a lognormal distribution, the decision is determined from

$$\max\left\{\left\{(\mu - re)'X\right\} + r - \frac{1}{2}\widetilde{X}\Delta\widetilde{X}\right\}, \tag{22.9}$$

with mean returns $\mu_i,\ i = 1, \ldots, K$, and covariance of returns $\Delta = (\delta_{ij})$. Also, $\widetilde{X}' = \left(x_1, \ldots, x_k\right)$ are the bets in risky games. For this continuous time problem, the Kelly strategy has the closed form

$$\widetilde{X}^* = \Delta^{-1}(\mu - re). \tag{22.10}$$

Table 22.2 Discrete Time Formulas

Measure	Formula
$\varphi_T(X)$	$E\log(R'X)$
$\gamma_T(X)$	$\Phi\left(\dfrac{TEJ(X)-\log\left(\frac{W_T}{w_0}\right)}{\sqrt{T\sigma(J)}}\right)$
$\eta_u(X)$	$\dfrac{\theta}{2EJ(X)}\left[\dfrac{U_*(X)}{\theta^{u^*}(x)}\right]-\dfrac{z_0}{EJ(X)}$
$\beta_{l,u}(X)$	$\dfrac{\left(\theta^{z_0}-\theta^{u^*}\right)\left(\theta^{u^*}-\theta^{L^*(X)}+\left(\theta^{z_0}-\theta^{L^*(X)}\right)\theta^{u^*(X)}-\theta^{l^*}\right)}{2\left(\theta^{u^*(X)}-\theta^{l^*}\right)\left(\theta^{u^*}-\theta^{L^*(X)}\right)}$

Table 22.3 Continuous Time Formulas

Measure	Formula
$\varphi_T(X)$	$((\mu(X)-r)\lambda(X)+r)T$
$\gamma_T(X)$	$\Phi\left(\dfrac{T.D(X)-\log\frac{W_T}{w_0}}{\sqrt{T\sigma(\widetilde{X})\lambda(X)}}\right)$
$\eta_u(X)$	$\dfrac{1}{D(X)}\log\left(\dfrac{u}{W_0}\right)$
$\beta_{l,u}(X)$	$\dfrac{1-\left(\frac{u}{W_0}\right)^{\frac{2D(X)}{\sigma^2(\widetilde{X})\lambda^2(X)}}}{1-\left(\frac{u}{T}\right)^{\frac{2D(X)}{\sigma^2(\widetilde{X})\lambda^2(X)}}}$

The Kelly or log optimal portfolio is $X^{*'}=\left(x_0^*,\widetilde{X}^{*'}\right)$, where $x_0^*=1-\sum_i^k=_1 x_i^*$. The continuous time formula can be viewed as an approximate solution to the discrete time betting problem. The Kelly strategy is a *fixed mix*. That is, the fraction of wealth bet in games is determined by X^* but rebalancing as wealth varies is required to maintain the fractions.

2.2 Some Important Properties of the Kelly Strategy

The considerable interest in the Kelly strategy, particularly in the gambling, investing, and probability literature, has produced a number of important good and bad properties. Some are listed below; for more details refer to MacLean, Thorp, Zhao, and Ziemba (2010, 2011).

Good: Maximizing $ElogW(X)$ asymptotically maximizes the rate of asset growth (Breiman (1960, 1961); Algeot and Cover 1988).

Good: The expected time to reach a preassigned goal u is asymptotically as u increases least with a strategy maximizing $ElogW(X)$ (Breiman 1960; Algeot and Cover 1988; Browne 1997).

Good: Under general conditions Maximizing $ElogW(X)$ also asymptotically maximizes the median final wealth (Ethier 2004).

Good: The $ElogW(X)$ bettor never risks ruin (Hakansson and Miller 1975).

Good: The absolute amount bet is monotone increasing in wealth (Thorp 2006).

Good: The $ElogW(X)$ bettor has a myopic policy (Kelly 1956; Hakansson 1971).

Good: The $ElogW(X)$ bettor's fortune pulls ahead of other "essentially different" strategies' wealth for long sequences of plays (Thorp 1975/1971; Ziemba and Hausch 1986; Aucamp 1993; Browne 1997).

Good: Kelly gambling yields wealth W^* such that $E\frac{W}{W^*} \leq 1$ for all other strategies (Bell and Cover 1980, 1988; Griffin 1984).

Good: The chance that the Kelly bettor is ahead of any other bettor after the first play is at least 50 percent (Bell and Cover 1980).

Good: The Kelly bettor is never behind any other bettor on average in $1, 2, \ldots$ trials. (Finkelstein and Whitley 1981).

Bad: The bets are extremely large when the wager is favorable and the volatility of returns is very low (Ziemba and Hausch 1986).

Bad: One over bets when the probabilities for events are in error (Chopra and Ziemba 1993, MacLean, Foster, and Ziemba 2007).

Bad: For coin tossing, if the number of wins equals the number of losses then the bettor is behind (MacLean and Ziemba 1999).

Bad: The average Kelly return converges to half the optimal arithmetic return (Ethier and Tavaré 1983).

Bad: The total amount bet swamps the winnings—that is, there is much churning (Ethier and Tavaré 1983).

3 Fractional Kelly Strategies

A challenging aspect concerning the performance of the Kelly strategy is that it is good on some measures and poor on others. If we refer back to table 22.1, the good properties of the Kelly are with the means (or long-term growth) and the less desirable properties are with percentiles (or the security of the acceptable results).

A variation on the Kelly strategy is the fractional Kelly strategy defined as $\widetilde{X}_f = f\widetilde{X}^*, 0 \leq f \leq 1$, so it is a blend of the Kelly strategy and cash. The fractional Kelly strategy has the same distribution of wealth across risky gambles as the Kelly but varies the fraction of wealth invested in those risky assets. The strategy $\widetilde{X}_f$ has fixed fractions, and therefore the formulas in tables 22.2 and 22.3 apply. That is, the fractional Kelly strategies can be evaluated based on performance on the measures of expected growth rate and security of growth (or low risk).

3.1 Bi-Criteria Problems

The problem of determining the optimal trade-off of expectation and risk can be formalized by using a utility function over these two attributes similar to that in mean-variance analysis; see, for example, Markowitz (1952, 1987). Leonard MacLean et al. (2004) did this another way by adding probability constraints on the drawdown; then, using scenario analysis, an optimal fractionalized Kelly strategy can be determined. The examples in section 4 illustrate this trade-off choice.

Analogous to static mean-variance analysis (see Markowitz 1952, 1987), a growth–security combination pair is inefficient if another pair in that combination has either a higher mean growth and no lower security level or a higher security level and no lower growth rate. A strategy is inefficient if its growth–security combination is inefficient. Efficient growth–security combinations are those that are not inefficient. The efficient growth–security frontier is the set of all efficient growth–security pairs. An efficient trade-off between growth and security occurs along the efficient frontier. The efficiency problems based on the performance criteria in table 22.1 are given in table 22.4. The constraint on security sets an acceptable risk α for not meeting the performance target.

In addition to the growth–security problems, an expected utility problem M_1 is defined. The power utility plays an important role in the Kelly and fractional Kelly strategies. The log is the power utility as $p \to 0$. The coefficient ρ is a risk aversion

Table 22.4 Alternative Decision Models

Model	Criterion	Problem
M_1	Expected power utility, with risk aversion index: $1-\rho$, ρ <1..	$\mathrm{Max}\left\{E\left[\frac{1}{p}\left(W(T)^p - 1\right)\right]\right\}$
M_2	Optimal growth subject to a VaR constraint.	$\mathrm{Max}\left\{\phi_T(X)\,\vert\,\gamma_T(X) \geq 1-\alpha\right\}$
M_3	Optimal growth subject to wealth goals constraint.	$\mathrm{Min}\left\{\eta_u(X)\,\vert\,\beta_{l,u}(X) \geq 1-\alpha\right\}$

parameter, and the Arrow–Pratt relative risk aversion index is $(1-\rho)$. If $\widetilde{X}^*$ is the Kelly strategy for risky gambles, then $\widetilde{X}_\rho = \frac{1}{1-\rho}\widetilde{X}^*$, $\rho < 0$, is the optimal solution to problem M_1 in the continuous time case where returns are log-normally distributed. That formula gives an indication of the risk aversion property of fractional Kelly strategies.

The optimality property applies to the other problems. The subclass of fractional Kelly strategies is defined as

$$\widetilde{x}^* = \left\{ \widetilde{X}_f^* | \widetilde{X}_f^* = f.\widetilde{X}^*,\ 0 \le f \le 1 \right\}. \tag{22.11}$$

The significance of the fractional Kelly strategies lies in their optimality for the problems in table 22.4, assuming the geometric Brownian model for returns is correct.

Theorem

Let X_{M_j} *be the optimal solution to growth problem* M_j, $j = 1,2,3$ *defined in* table 22.6. *Then* $X_{M_j} \in \widetilde{x}^*$, *that is, the solution is fractional Kelly.*

For proof see MacLean, Zhao, and Ziemba (2005).

In the continuous time formulation, the optimal investment strategies for the various problems have the same form. However, the actual fraction in each problem, which controls the allocation of capital to risky and risk-free instruments, depends on the decision model and parameters. The formulas for the fractions for different models are in table 22.5. The notation $\widetilde{\mu} = (\mu - re)'\widetilde{X}^* + r$, and $\widetilde{\sigma}^2 = \widetilde{X}^{*'}\Delta\widetilde{X}^*$ is used for the mean and variance of the rate of return on the Kelly strategy. Also at time t, y_t^* is the minimum positive root of the equation $\gamma y^{c_t+1} - y + (1-\gamma) = 0$, for $c_t = \frac{\log(u)-\log(w_t)}{\log(w_t)-\log(l)}$. Coefficients are $\Big\{A_1 = (\mu - re)'\Delta^{-1}(\mu - re),\ B_{1t} = A_1 + z_a\sqrt{\frac{A_1}{T-t}}, C_{1t} = r - (T-t)^{-1}\log\left(\frac{w_T}{w_t}\right)\Big\}$, and $\Big\{H = \frac{\widetilde{\mu}-r}{\widetilde{\sigma}^2},\ h_t = \frac{\log(w_t)-\log(l)}{\log(w_t)-\log(y_t^*.l)}\Big\}$.

Table 22.5 Investment Fractions

Model	Parameters	Kelly Fraction
M_1	ρ	$f_1 = \dfrac{1}{1-\rho}$
M_2	(w_T, α)	$f_{2t}\dfrac{B_{1t} + \sqrt{B_{1t}^2 + 2A_1C_{1t}}}{A_1}$
M_3	(l, u, α)	$f_{3t} = h_t.H + \sqrt{[h_t.H]^2 + \dfrac{2rh_t}{\widetilde{\sigma}^2}}$

The solutions displayed in table 22.5 are derived from the continuous time wealth equation, though the strategies are calculated at discrete decision points in time. The alternative problems in table 22.4 can be based on the discrete time wealth equation, but the optimal solution is not necessarily fractional Kelly. However, the fractional Kelly solution may be near-optimal. If the feasible strategies for the discrete time problem are restricted to the class of fractional strategies, the solutions are *effective* (MacLean, Ziemba, and Blazenko 1992). That is, as the fraction changes, the growth (objective) and security (constraint) move in opposite directions so that growth is monotone nonincreasing in security. Specifically, $0 \le f \le 1$,

$$\left\{\left(\frac{d}{df}\phi_T\left(X_f^*\right) \ge 0, \frac{d}{df}\gamma_T\left(X_f^*\right) \le 0\right), \left(\frac{d}{df}\eta_u\left(X_f^*\right) \ge 0, \frac{d}{df}\beta_{1,u}\left(X_f^*\right) \le 0\right)\right\}. \tag{22.12}$$

The implication of this monotonicity is that growth can be traded for security using the fraction allocated to the optimal growth portfolio. So the growth–security trade-off can be observed for various fractional Kelly strategies and suitable fractions (meeting investor preferences) can be determined. This will be demonstrated with some applications.

4 Applications

4.1 Blackjack

The game of blackjack, or 21, evolved from several related card games in the nineteenth century. It became fashionable during World War I and has since become enormously popular, played by millions of people in casinos around the world. Billions of dollars are lost each year by people playing the game in Las Vegas alone. A small number of professionals and advanced amateurs, using various methods such as card counting, are able to beat the game (Thorp 1962). See Janacek (1998) for a computer program that evaluates game statistics, such as advantage depending on the card-counting system used and casino rules. The software is also available online (see Statistical Blackjack Analyzer 5.5). The object is to reach, or be close to, 21 with two or more cards. Scores above 21 are said to bust or lose. Cards 2 to 10 are worth their face value: Jacks, queens, and kings are worth 10 points, and aces are worth 1 or 11 points, at the player's choice. The game is called blackjack because an ace and a 10-valued card was paid three for two and an additional bonus accrued if the two cards were the ace of spades and the jack of spades or clubs. While this extra bonus has been dropped by current casinos, the name has stuck. Dealers normally play a fixed strategy of drawing cards until the total reaches seventeen or more, at which point they stop. A variation is when a soft 17 (an ace with cards totaling six) is hit. It is better for the player if the dealer stands on soft 17. The house has an edge of 1–10 percent against typical players. The strategy of mimicking the dealer loses about 8 percent because the player must hit first and busts about 28 percent of the time ($0.28^2 \simeq 0.08$). However, in Las Vegas the average player loses only about 1.5 percent. The edge for a successful card counter varies from about -5 percent to $+10$ percent, depending on the favorability of the deck. By wagering more in favorable situations and less or nothing when the deck is unfavorable, an average weighted edge is about ½–2 percent.

An approximation to provide insight into the long-run behavior of a player's fortune is to assume that the game is a Bernoulli trial with a probability of success $p = 0.51$ and a probability of loss $1 - p = 0.491 - p = 0.49'' - = 0.49$. Then with outcomes $\Omega = \{0, 1\}$, the payoffs are $K(0, x) = x$ with probability p, and $K(1,x) = -x$ with probability $1 - p1 - p$ "–". The mean growth rate is $E\,log\,(1{+}K\,(w, x)) = p\,log\,(1 + x) + \big(1 - p\big)\,log\,(1 - x)$. The optimal fixed fraction strategy is $x^* = 2p - 1$ *if* $EK > 0$; $x^* = 0$ *if* $EK \leq 0$. This optimal strategy may be interpreted as the edge divided by the odds (1–1 in this case). In general, for two outcome win-or-lose situations where the size of the wager does not influence the odds, the same edge divided by the odds formula holds. Hence with a 2 percent edge, betting on a 10–1 shot, the optimal wager is 0.2 percent of one's fortune. The growth rate of the investor's fortune is shown in figure 22.2. It is nearly symmetrical around $x^* = 0.02$. A security measure is also displayed in figure 22.2 in terms of the probability of doubling or quadrupling before halving. Since the growth rate and the security are both decreasing for $x > x^*$, it follows that it is never advisable to wager more than x^*. However, one may wish to trade off lower growth for more security using a fractional Kelly strategy. For example, a drop from $p = 0.02$ to 0.01 for a 0.5 fractional Kelly strategy decreases the growth rate by 25 percent but increases the chance of doubling before halving from 67 to 89 percent.

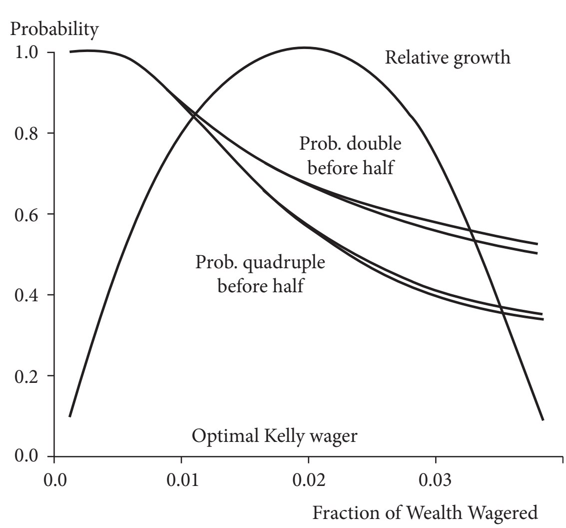


FIGURE 22.2 Relative growth versus probability of doubling before halving (blackjack).

In the blackjack example, the additional information provided by the security measure was important in reaching a final investment decision. In a flexible or adaptive decision environment competing criteria would be balanced to achieve a satisfactory path of accumulated wealth. Professional blackjack teams often use a fractional Kelly wagering strategy with the fraction = 0.2 to 0.8. See Gottlieb (1985) for further discussion including the use of adaptive strategies.

4.2 Horse Racing

In a race with n horses there are positive returns for show wagers made on the first three finishers. The set of all outcomes with probability p_{ijk} is $\Omega = (1,2,3), \ldots$ "…", $(i,j,k), \ldots$ "…", $(n-$"$-$"$2, n-$"$-$"$1, n)$. If is the probability that horse i wins, then Harville (1973) gave the probability of an (i,j,k) finish as $p_{ijk} = \frac{q_i q_j q_k}{(1-q_i)(1-q_i-q_j)} = q_i \times \frac{q_j}{1-q_i} \times \frac{q_k}{1-q_i-q_j}$. That is the probability that i wins times the probability that j wins a race without i times the probability that k wins a race that does not contain i or j. (For a bias correction to the Harville formulas see Hausch, Lo, and Ziemba 1994/2008. The bias comes from the fact that horses that do not win come in less frequently for second and third than the Harville formulas suggest. The correction replaces q_j and q_k by $\widetilde{q}_j = \frac{q_j^a}{\sum q_k^a}$, $\widetilde{q}_k = \frac{q_k^{a^2}}{\sum q_j^{a^2}}$, respectively. The selection $a = 0.81$ is recommended.) An investor wagers the fractions (x_{i1}, x_{i2}, x_{i3}) of his or her fortune w_0 on horse i to win, place, and show, respectively. One collects on a win bet when the horse is first, on a place bet when the horse is first or second, and on a show bet when the horse is first, second, or third. The order of finish does not matter for place and show bets. All bettors, wagering on a particular horse, share the net pool in proportion to the amount wagered, once the original amount of the winning bets are refunded and the winning horses share equally the resulting profits.

A particular anomaly is the place and show wager (Ziemba 1987). Let the player bets be (x_{i2}, x_{i3}) for horse i to place and show, respectively. Also let (X_{i2}, X_{i3}) be the total place and show bets of other people and the particular (i,j,k) outcome is

$$\begin{aligned}
K\left((i,j,k),x\right) = {} & \frac{Q\left(X_2 + \sum_{l=1}^{n} x_{l2}\right) - \left(x_{i2} + x_{j2} + X_{i2} + X_{j2}\right)}{2} \\
& \times \left(\frac{x_{i2}}{x_{i2} + X_{i2}} + \frac{x_{j2}}{x_{j2} + X_{j2}}\right) \\
& + \frac{Q\left(X_3 + \Sigma_{l=1}^{n} x_{l3}\right) - \left(x_{i3} + x_{j3} + x_{k3} + X_{i3} + X_{j3} + X_{k3}\right)}{3} \\
& \times \left(\frac{x_{i2}}{x_{i3} + + X_{i3}} + \frac{x_{j3}}{x_{j3} + X_{j3}} + \frac{x_{k3}}{x_{k3} + X_{k3}}\right) + w_0 \\
& - \left(\sum_{l=1, l\neq i,j,k}^{n} x_{l2} + \sum_{l=1, l\neq i,j}^{n} x_{l3}\right) + \gamma \left(\sum_{l=1}^{n} x_{l2} + \sum_{l=1}^{n} x_{l3}\right).
\end{aligned}$$

where $Q = 1-$ the track take and γ is the rebate fraction. This return function is developed in Hausch, Ziemba, and Rubinstein (1981). The return is net of transactions costs (track take) and includes the track rebate and the effect of bets on the odds.

The optimal Kelly wager is determined from the problem

$$\mathrm{Max}\, l \left\{ \sum_{i=1}^{n} \sum_{j=1, j\neq i}^{n} \sum_{k=1, k\neq i,j}^{n} \left[p_{i,j,k} log\left(K\left(i,j,k\right), x\right)\right] \right\}$$

subject to

$$\sum_{i=1}^{n} (x_{i2} + x_{i3}) \leq w_0.$$

$$x_{i2} \geq 0, x_{i3} \geq 0,\ i = 1, \ldots, n.$$

The Kelly bet test was performed in 2004 by John Swetye and William Ziemba, who searched for bets at 80 U.S. racetracks Ziemba and MacLean (2011). A \$5,000 initial wealth returned a profit of \$30,000. There was a lot of churning, with a total of more than \$1.5 million bet. The Dr Z system actually lost 7 percent, but with a rebate of 9 percent there was a 2 percent gain. Figure 22.3 presents the wealth trajectory, with a comparison to the outcome for fractional wagers of $\frac{1}{2}$ Kelly and $\frac{1}{3}$ Kelly. The smaller betting fractions provide more security but have less growth and lower final wealth.

Implicit in the above example is the ability to identify races where there is a substantial edge in the bettor's favor. There has been considerable research into that question (as surveyed by Hausch, Lo, and Ziemba 1994/2008; Hausch and Ziemba (1985, 2008)).

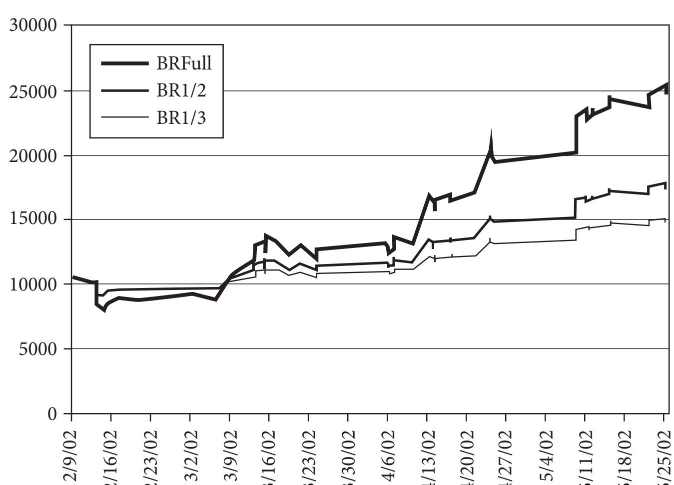


FIGURE 22.3 Wealth results for Dr Z place and show bets in 2004.

Donald Hausch et al. (1981) demonstrated the existence of anomalies in the place and show market. At thoroughbred racetracks, about 2–4 profitable wagers with an edge of 10 percent or more exist on an average day. The profitable wagers occur mainly because (1) the public has a distaste for the high probability–low payoff wagers that occur on short-priced horses to place and show and (2) the public is unable to properly evaluate the worth of place and show wagers because of their complexity—for example, in a 10-horse race there are 120 possible show finishes, each with a different payoff and chance of occurrence. In Hausch et al. (1981) and more fully in Hausch and Ziemba (1985), equations were developed that approximate the expected return and optimal Kelly wagers based on minimal amounts of data to make the edge operational in the limited time available at the track. The Kentucky Derby is discussed by Bain, Hausch and Ziemba (2005).

The bias correction to probabilities of second and third (Hausch, Lo, and Ziemba 1994/2008) approximately cancels the favorite-longshot bias, where long shots have lower expected value and favorites have higher ones for place and show bets. However, this bias is important in other wagers. (For more information on the favorite-longshot bias refer to Hausch and Ziemba 2008).

Since the Dr Z system was derived in the 1980s it has changed the character of the place and show and other racetrack betting markets. The recent changes have to do with the rebate—the return of a fraction of the amount bet, long, and short betting in Betfair in London and other locales. Track betting, some of which is done by well-funded racing syndicates, has become a tougher market to win; see Benter (1994), who discusses the world's most successful such syndicate. See Ziemba (2014) for more details on various racetrack bets.

4.3 Lotto Games

In lotto games players select a small set of numbers from a given list. The prizes are shared by those with the same numbers as those selected in the random drawing. The lottery organization bears no risk in the pari-mutuel system and takes its profits before the prizes are shared. Hausch and Ziemba (1995, 2008) surveyed these games. Ziemba et al. (1986) studied the 6/49 game played in Canada and several other countries. (The analysis was updated in Ziemba 2008 and other papers in Hausch and Ziemba 2008). Numbers ending in eight and especially nine and zero tend to be unpopular. Six tuples of unpopular numbers have an edge with expected returns exceeding their cost. The expected value approaches \$2.25 per dollar wagered when there are carryovers (that is, when the Jackpot is accumulating because it has not been won.) However, investors may still lose because of mean reversion (the unpopular numbers tend to become less unpopular over time) and gamblers' ruin (the investor has used up his resources before winning). MacLean, Ziemba, and George Blazenko (1992) investigated how an investor might do playing sets of unpopular numbers with a combined advantage using the data in table 22.6.

Table 22.6 Lotto 6/49 Data

Prizes	Prob	Value	Contribution %
Jackpot	1/13983816	$6M	42.9
Bonus	1/2330636	$0.8M	34.3
5/6	1/55492	$M	9.0
4/6	1/1032	$5,000	14.5
3/6	1/57	$150	17.6
Edge			18.1%
Kelly bet			0.00000011
Number of Tickets with 10M bankroll			11

Source: MacLean, Ziemba, and Blazenko (1992).

The optimal Kelly wagers are extremely small. The reason for this is that the bulk of the expected value is from prizes that occur with less than one in a million probability. A wealth level of $1 million is needed to justify even one $1 ticket. Figure 22.4 provides the chance that the investor will double, quadruple, or tenfold this fortune before it is halved using Kelly and fractional Kelly strategies. These chances are in the range of 40–60 percent. With fractional Kelly strategies in the range of 0.00000004 and 0.00000025 or less of the investor's initial wealth, the chance of increasing one's initial fortune tenfold before halving it is 95 percent or more. However, it takes an average of 294 billion years to achieve this goal, assuming there are 100 draws per year as there are in the Canadian lotto 6/49.

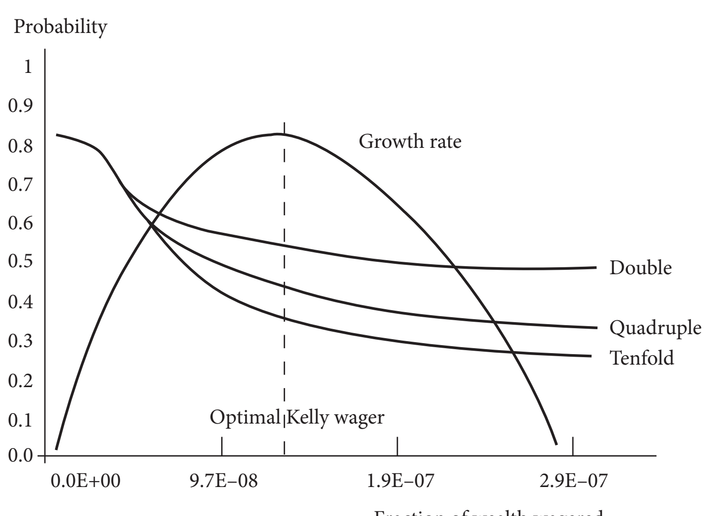


FIGURE 22.4 Lotto 6/49—Probability of multiplying before losing half of one's fortune versus bet size.

Source: Maclean, Ziemba, and Blazenko (1992).

The conclusion is that except for millionaires and pooled syndicates, it is not possible to use the unpopular numbers in a scientific way to beat the lotto and have high confidence of becoming rich; these aspiring millionaires are also most likely going to be residing in a cemetery when their distant heir finally reaches the goal.

4.4 Investing in the Turn-of-the-Year Effect

The trade-off between growth and security with the Kelly criterion was used by Ross Clark and Ziemba (1987) in their analysis of investment in the turn-of-the-year effect in the futures markets. The excess return of small cap stocks minus large cap stocks is most pronounced in January. The distribution of gains at the turn of the year from holding long positions in the Value Line futures Index of small stocks and short positions in the Standard & Poor's (S&P) futures index of large cap stocks is given in table 22.7, where each point is worth \$500. The data covers the period 1976–1977 to 1986–1987.

The Kelly strategy calculated from this distribution invests 74 percent of one's fortune in the trade. This is very aggressive considering the possible estimation errors and the market volatility.

In figure 22.5 is a graph for fractional Kelly strategies with the turn-of-the-year trade, showing the chance of reaching a wealth of \$10 million before ruin, starting from various initial wealth levels. The graph for 0.25 Kelly is much more secure. Similarly in figure 22.6, the trade-off between relative growth and probability of achieving a desired wealth level is displayed. Going from Kelly to quarter Kelly realizes a probability gain (security) of about 0.25 and an almost equivalent loss in relative growth. Ziemba used an approximate 0.25 Kelly strategy with consistent success in actual trades on this commodity in the 14 years from 1982/1983 to 1996/1997, winning each year.

Because the declining Value Line volume makes the trade risky, Ziemba did not participate in the trade from 1997 to 2008; see Ziemba (1994), Hensel and Ziemba (2000), Rendon and Ziemba (2007), and Ziemba (2011). An update on the turn-of-the-year effect is provided in Ziemba (2012), including successful trades of the turn-of-the-year effect in 2009, 2010, and 2011 using the Russell 2000 small cap futures index (long) and the S&P large cap futures index (short). Figure 22.7 shows the performance, where 100K grew to 147K in two years. See the large portfolio jumps on the far left and far right plus the middle jump for the three turn-of-the-year trades in the Anomalies Test Account. An interesting aspect of this anomaly is the shift from January to a December effect.

Table 22.7 Returns Distribution for VL/S&P Spread

Point Spread	7	6	5	4	3	2	1	0	−1
Probability	.007	.024	.07	.146	.217	.229	.171	.091	.045

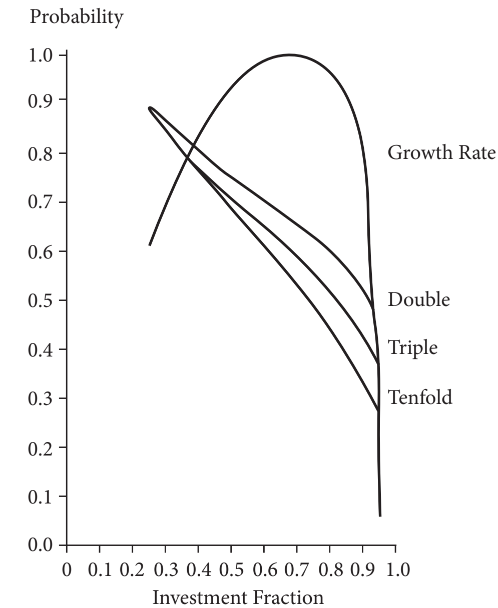


FIGURE 22.5 Turn-of-the-year effect: Probability of reaching $10 million before ruin for Kelly, half Kelly, and quarter Kelly strategies.

Source: MacLean, Ziemba, and Blazenko (1992).

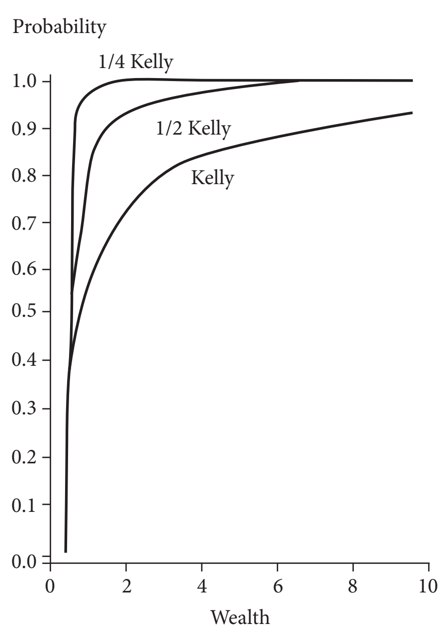


FIGURE 22.6 Relative growth versus probability of goal attainment.

Source: MacLean, Ziemba, and Blazenko (1992).

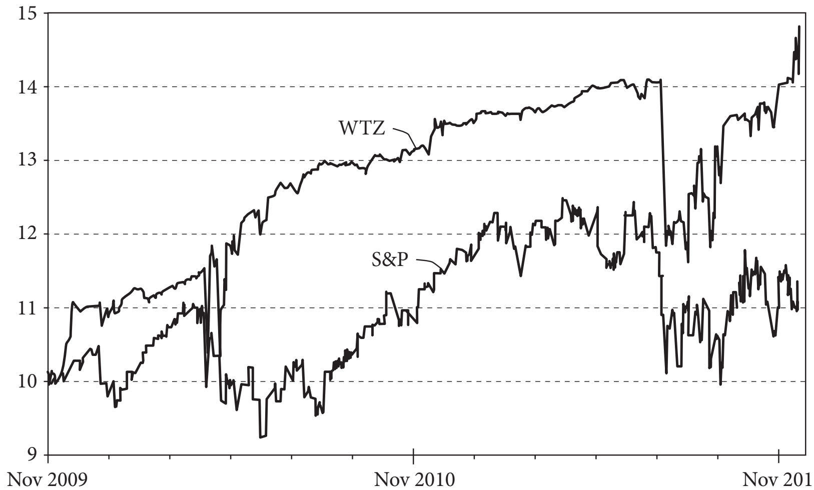


FIGURE 22.7 Private account of WTZ using anomalies, 2009–2011.

5 The Great Investors

A large game of chance is the stock market. Much has been written about the efficiency of the market, but clearly there are opportunities for excess returns. (Ziemba 2012 surveys U.S. calendar anomalies.) The legends of investing have a number of common features: excellence in selecting/evaluating opportunities, focus, deep pockets to ride out downturns, and successful investment strategies. Interestingly some great investors have followed a Kelly type investment strategy. The outperformance of some legends is documented in this final section. Details on these and other investors is given in Gergaud and Ziemba (2012).

5.1 John Maynard Keynes

Keynes ran the King's College, Cambridge University, endowment fund from 1927 until 1945. Ziemba (2003) estimated that Keynes had about an 80 percent Kelly or a negative power utility of $u(w) = -w^{-0.25}$. That is an aggressive strategy, and he lost a lot (over 50%) in the early years. The trajectory of the Cambridge Chest is shown in figure 22.8.

Over the 19 years of his management, the fund had a geometric mean return of 9.12 percent compared to the market index of −0.89 percent. So he had superior long-term performance.

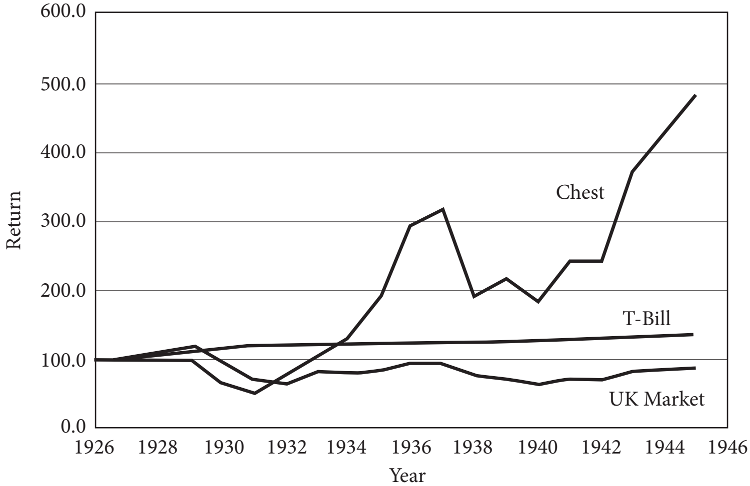


FIGURE 22.8 Portfolio performance: Keynes.

Source: Ziemba (2005).

5.2 Warren Buffett and George Soros

Two full Kelly investors who hold few and concentrated positions are the famous Warren Buffett and George Soros. The wealth paths for their funds, Berkshire Hathaway and Quantum, respectively, are shown in figure 22.9. As with Keynes, these billionaire

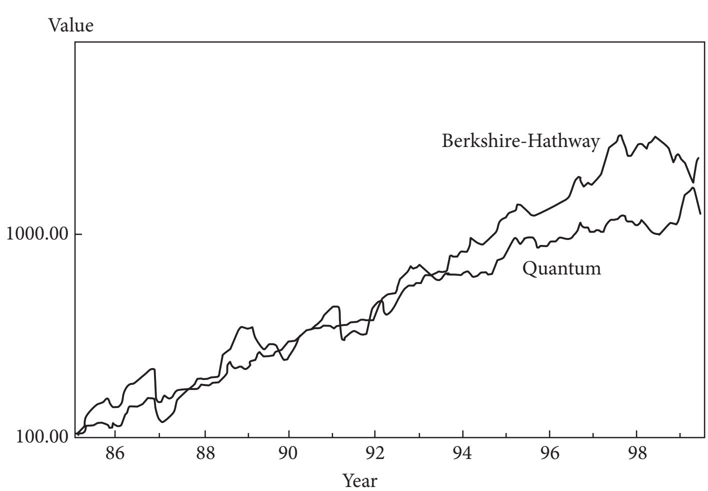


FIGURE 22.9 Portfolio performance: Buffett and Soros, 1985–2000.

Source: Ziemba 2005.

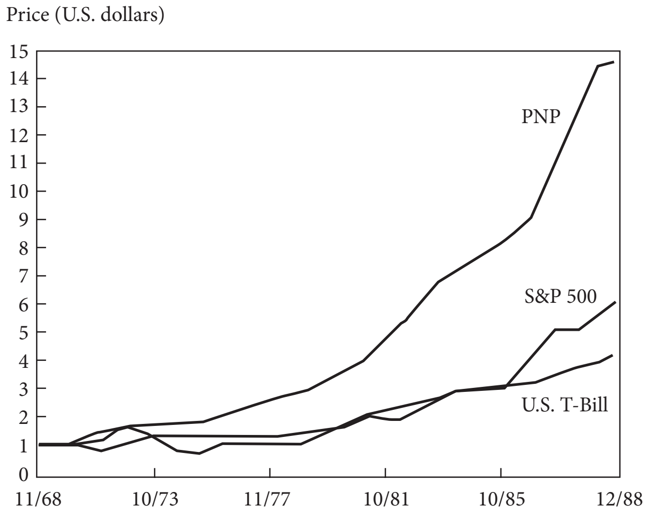


FIGURE 22.10 Portfolio performance: Thorp.

Source: Ziemba 2005).

investors had many gains but also many losses. Out of the 172 months represented in the figure, Berkshire Hathaway had 58 and Quantum 53 losing months. However, the gains were very high, as happens with the Kelly strategy. Both funds had annual growth rates in excess of 15 percent during the years 1977 to 2005.

5.3 Edward O. Thorp

Ed Thorp is famous for his work on blackjack and for applying his ideas on mispriced options and warrants and the growth optimal approach to the stock market (Rotando and Thorp 1992, Thorp and Kassouf 1967). In fact he coined the term "fortune's formula" for the Kelly strategy. Thorp ran the Princeton Newport Partners (PNP) hedge fund from 1968–1988. He had an amazing record of just three monthly losses over a 20-year period. The smooth path of PNP is presented in figure 22.10. The fund had an annual growth rate of 13.5 percent, a bit lower than that of Buffett and Soros but without dips. Thorp's later results in other hedge funds are equally impressive.

5.4 James Simon

The Renaissance Medallion fund uses ideas such as the Kelly criterion to achieve superior returns. Working under mathematician James Simons, a staff of technical researchers and traders devises edges to generate successful trades, most of which are short-term (lasting only seconds). Amazingly, the fund had only 17 monthly losses in 148 months, 3 losses in 49 quarters, and zero losses in 12+ years of trading to

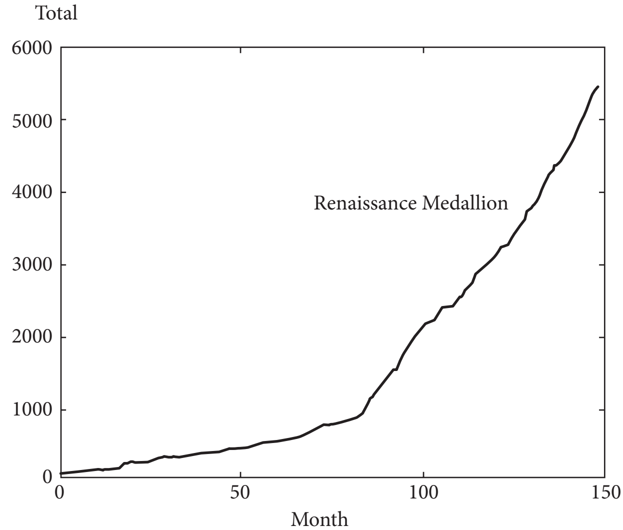


FIGURE 22.11 Portfolio performance: Simons.

Source: Ziemba and Ziemba 2007).

2005. So Renaissance Medallion had an annual growth rate of more than 30 percent, and the path slope is basically monotone increasing (figure 22.11). (See Ziemba and Ziemba 2007 for an analysis of this fund and Gergaud and Ziemba 2012 for an update.)

6 Conclusion

The Kelly strategy, where the expected logarithm of returns is maximized, subject to constraints on betting proportions, has a rich tradition in the analysis of games of chance. The strategy has many desirable properties that are mainly related to long-run asymptotic growth. Use of this strategy over many plays of favorable games usually results in unparalleled wealth. However, the strategy is aggressive since its Arrow–Pratt risk aversion index is essentially zero. This leads to violent wealth paths, and a string of bad luck can wipe out the bettor. Smoother wealth paths are obtained with fractional Kelly strategies that blend the Kelly strategy with cash. This chapter has presented the concepts in the Kelly strategy and described its good and bad properties. The strategy and its fractional Kelly modification have been applied to some familiar games of chance, such as blackjack, horse racing, commodity markets, and lotteries. The growth and security aspects of the strategy have been illustrated with tabulations and graphs, and the records of some great investors that have Kelly-like portfolios have been presented.

References

Algeot, Paul H., and Thomas M. Cover. 1988. Asymptotic optimality and asymptotic equipartition properties of log-optimum investment. *Annals of Probability* 16(2):876–898.

Aucamp, Donald C. 1993. On the extensive number of plays to achieve superior performance with the geometric mean strategy. *Management Science* 39(9):1163–1172.

Bain, Roderick S., Donald B. Hausch, and William T. Ziemba. 2006. An application of expert information to win betting on the Kentucky derby, 1981–2005. *European Journal of Finance* 12(4):283–301.

Bell, Robert M., and Thomas M. Cover. 1980. Competitive optimality of logarithmic investment. *Mathematics of Operations Research* 5(2):161–166.

Bell, Robert M., and Thomas M. Cover. 1988. Game-Theoretic Optimal Portfolios. Management Science 34(6): 724–733.

Benter, W. 1994. Computer based horse race handicapping and wagering systems: A report. In *Efficiency of Racetrack Betting Markets*, edited by Donald B. Hausch, Victor S. Y. Lo, and William T. Ziemba. San Diego, Calif.: Academic Press, 183–198.

Bernoulli, Daniel. [1738] 1954. Exposition of a new theory on the measurement of risk, translated by Louise Summer. *Econometrica* 22(1):22–36.

Breiman, Leo. 1960. Investment policies for expanding business optimal in a long-run sense. *Naval Research Logistics Quarterly* 7(4):647–651.

Breiman, Leo. 1961. Optimal Gambling Systems for Favorable Games. Proceedings of the 4th Berkeley Symposim on Mathematical Statisics and Probability 1; 63–68.

Browne, Sid. 1997. The return on investment from proportional portfolio strategies. *Advances in Applied Probability* 30(1):216–238.

Chopra, Vijay K., and William T. Ziemba. 1993. The effect of errors in means, variances, and covariances on optimal portfolio choice. *Journal of Portfolio Management* 19(2):6–11.

Clark, Ross, and William T. Ziemba. 1987. Playing the turn-of-the-year effect with index futures. *Operations Research* 35(6):799–813.

Cover, Thomas M., and Joy A. Thomas. 2006. *Elements of information theory*. 2nd ed. Hoboken, N.J.: John Wiley & Sons.

De Giorgio, Enrico. 2005. Reward-risk portfolio selection and stochastic dominance. *Journal of Banking & Finance* 29(4):895–926.

Dohi, T., H. Tanaka, N. Kaio, and S. Osaki. 1995. Alternative growth versus security in continuous dynamic trading. *European Journal of Operational Research* 84(2):265–278.

Ethier, S. N. 2004. The Kelly system maximizes median fortune. *Journal of Applied Probability* 41(4):1230–1236.

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. Whitely. 1981: Optimal Strategies in Repeated Games. Advances in Applied Probability 13:415–428.

Gergaud, Olivier, and William T. Ziemba. 2012. Great investors: Their methods, results, and evaluation. *Journal of Portfolio Management* 38(4):128–147.

Gottlieb, G. 1985. An Analytic Derivation of Blackjack Win Rates. Opeations Research 33: 971–988.

Grauer, Robert R., and Nils H. Hakansson. 1986. A half century of returns on levered and unlevered portfolios of stocks, bonds, and bills, with and without small stocks. *Journal of Business* 59(2):287–318.

——. 1987. Gains from international diversification: 1968–85 returns on portfolios of stocks and bonds. *Journal of Finance* 42(3):721–739.

Griffin, Peter A. 1984. Different measures of win rate for optimal proportional betting. *Management Science* 30(12):1540–1547.

Hakansson, Nils H. 1970. Optimal investment and consumption strategies under risk for a class of utility functions. *Econometrica* 38(5):587–607.

——. 1971. On optimal myopic portfolio policies, with and without serial correlation of yields. *Journal of Business* 44(3):324–334.

Hakansson, Nils H. and Bruce Miller. 1975. Compound-Return Mean-Variance Efficient Portfolios Never Risk Ruin. Management Science 22:391–400.

Hakansson, Nils H., and William T. Ziemba. 1995. Capital growth theory. In *Handbooks in operations research and management science*, vol. 9: *Finance*, edited by R. A. Jarrow, V. Maksimovic, and W. T. Ziemba. Amsterdam: Elsevier, 65–86.

Hanoch, G., and H. Levy. 1969. The efficiency analysis of choices involving risk. *The Review of Economic Studies* 36(3):335–346.

Harville, David. 1973. Assigning Probabilities to the Outcomes of Multi-Entry Competition. Journal of the American Statistical Association 68:312–316.

Hausch, Donald B., and William T. Ziemba. 1985. Transactions costs, extent of inefficiencies, entries and multiple wagers in a racetrack betting model. *Management Science* 31(4): 381–394.

——. 1995. Efficiency of sports and lottery betting markets. In *Handbooks in operations research and management science*, vol. 9: *Finance*, edited by R. A. Jarrow, V. Maksimovic, and W. T. Ziemba. Amsterdam: Elsevier, 545–580. (Updated in Ziemba 2008.)

——, eds. 2008. *Handbook of Sports and Lottery Markets.* Amsterdam: Elsevier.

Hausch, Donald B., Victor S. Y. Lo, and William T. Ziemba, eds. 1994. *Efficiency of racetrack betting markets.* San Diego, Calif.: Academic. (Updated in 2008 and published by World Scientific, Singapore.)

Hausch, Donald B., William T. Ziemba, and Mark Rubinstein. 1981. Efficiency of the market for racetrack betting. *Management Science* 27(12):1435–1452.

Hens, Thorsten, and Klaus Reiner Schenk-Hoppé. 2005. Evolutionary stability of portfolio rules in incomplete markets. *Journal of Mathematical Economics* 41(1–2):43–66.

Hensel, Chris R., and William T. Ziemba. 2000. Anticipation in the January effect in the US futures markets. In *Security Market Imperfections in Worldwide Equity Markets*, edited by Donald B. Kiem and William T. Ziemba. Cambridge: Cambridge University Press, 179–202.

Janacek, K. 1998. Optimal growth in gambling and investing. M.Sc. thesis, Charles University, Prague.

Kallberg, J. G., and William T. Ziemba. 1981. Remarks on optimal portfolio selection. In *Methods of operations research*, edited by Günter Bamberg and Otto Opitz. Methods of Operations Research 44. Cambridge, Mass.: Oelgeschlager, Gunn & Hain, 507–520.

——. 1983. Comparison of alternative utility functions in portfolio selection problems. *Management Science* 29(11):1257–1276.

——. 1984. Mis-specification in portfolio selection problems. In *Risk and capital: Proceedings of the second summer workshop on risk and capital held at the University of Ulm, West Germany, June 20–24, 1983*, edited by Günter Bamberg and Klaus Spreemann. Berlin: Springer-Verlag, 74–87.

Kelly, J. L., Jr. 1956. A new interpretation of information rate. *Bell System Technological Journal* 35(4):917–926.

Latané, Henry Allen. 1959. Criteria for choice among risky ventures. *Journal of Political Economy* 67(2):144–155.

Luenberger, David G. 1993. A preference foundation for log mean-variance criteria in portfolio choice problems. *Journal of Economic Dynamics and Control* 17(5–6):887–906.

——. 2009. *Investment science*. New York: Oxford University Press.

MacLean, Leonard C., and William T. Ziemba. 1991. Growth-security profiles in capital accumulation under risk. *Annals of Operations Research* 31:501–510.

——. 1999. Growth versus security tradeoffs in dynamic investment analysis. *Annals of Operations Research* 85:193–225.

——. 2000. Efficiency concepts in capital accumulation models. *Philosophical Transactions: Mathematical, Physical, and Engineering Sciences* 357(1758):2029–2037.

MacLean, Leonard C., Michael E. Foster, and William T. Ziemba. 2007. Covariance complexity and rates of return on assets. *Journal of Banking and Finance* 31(11):3503–3523.

MacLean, Leonard C., Edward O. Thorp, and William T. Ziemba. 2010. Long-term capital growth: The good and bad properties of the Kelly and fractional Kelly capital growth criteria. *Quantitative Finance* 10(7):681–687.

——. 2011. *The Kelly capital growth investment criterion*. World scientific handbook in financial economic series—vol. 3, Theory and practice. Singapore: World Scientific Publishing.

MacLean, Leonard C., Yonggan Zhao, and William T. Ziemba. 2005. Dynamic portfolio selection with process control. *Journal of Banking and Finance* 30(2):317–339.

——. 2009. *Optimal capital growth with convex loss penalties*. Working paper. Halifax, Nova Scotia: School of Business Administration, Dalhousie University.

MacLean, Leonard C., William T. Ziemba, and George Blazenko. 1992. Growth versus security in dynamic investment analysis. *Management Science* 38(11):1562–1585.

MacLean, Leonard. C., Rafael Sanegre, Yonggan Zhao, and William T. Ziemba. 2004. Capital Growth with Security. *Journal of Economic Dynamics and Control* 28(5):937–954.

MacLean, Leonard C., Edward O. Thorp, Yonggan Zhao, and William T. Ziemba. 2011. How does the *Fortune's Formula* Kelly capital growth model perform? *Journal of Portfolio Management* 37(4):96–111.

Markowitz, Harry M. 1952. Portfolio Selection. *Journal of Finance* 7(1):77–91.

——. 1987. *Mean-variance analysis in portfolio choice and capital markets*. Oxford, U.K.: Basil Blackwell.

Menger, Karl. [1934] 1967. The role of uncertainty in economics. Reprinted in *Essays in mathematical economics: In honor of Oskar Morgenstern*, edited by Oskar Shubik. Princeton, N.J.: Princeton University Press, 211–223.

Rendon, Juan, and William T. Ziemba. 2007. Is the January effect still alive in the futures market? *Financial Markets and Portfolio Management* 21(3):381–396.

Rogers, Leonard C. G. 2001. The relaxed investor and parameter uncertainty. *Finance and Stochastics* 5(2):131–154.

Rotando, Louis M., and Edward O. Thorp. 1992. The Kelly criterion and the stock market. *American Mathematical Monthly* 99(10):922–931.

Stutzer, Michael. 2003. Portfolio choice with endogenous utility: A large deviations approach. *Journal of Econometrics* 116(1–2):365–386.

Thorp, Edward O. 1962. *Beat the dealer: A winning strategy for the game of twenty-one*. New York: Blaisdell Publishing.

——. 1975. Portfolio choice and the Kelly criterion. In *Stochastic optimization models in finance*, edited by William T. Ziemba and Raymond G. Vickson. New York: Academic, 599–619. Reprinted 1971 from *1970 proceedings of the Business and Economic Statistics Section: Papers presented at the annual meeting of the American Statistical Association held in Detroit, Michigan*. Washington, D.C.: American Statistical Association, 215–224.

Thorp, Edward O., and Sheen T. Kassouf. 1967. *Beat the market: A scientific stock market system*. New York: Random House.

Ziemba, Rachel E. S., and William T. Ziemba. 2007. *Scenarios for risk management and global investment strategies*. Hoboken, N.J.: John Wiley & Sons.

Ziemba, William T. 1994. Investing in the turn-of-the-year effect in the US futures market. *Interfaces* 24(3):46–61.

——. 2003. *The stochastic programming approach to asset, liability, and wealth management*. Charlottesville, Va.: Research Foundation of AIMR.

——. 2005. The symmetric downside-risk Sharpe ratio. *Journal of Portfolio Management* 32(1):108–122.

——. 2008. Efficiency of racing, sports, and lottery betting markets. In *Handbook of sports and lottery markets*, edited by Donald B. Hausch and William T. Ziemba. Amsterdam: Elsevier, 183–222.

——. 2011. Investing in the turn-of-the-year effect. *Financial Markets and Portfolio Management* 25(4):455–472.

——. 2012. *Calendar anomalies and arbitrage*. Singapore: World Scientific.

——. 2014. *Exotic betting at the racetrack*. Singapore: World Scientific.

Ziemba, William T., and Donald B. Hausch. 1986. *Betting at the Racetrack*. San Luis Obispo, Calif.: Dr. Z Investments.

——. 1987. *Dr. Z's Beat the Racetrack*. New York: William Morrow. (Revised edition of *Beat the Racetrack* [1984])

Ziemba, William T., Shelby L. Brunelle, Antoine Gautier, and Sandra L. Schwartz. 1986. Dr. Z's 6/49 lotto guidebook. San Luis Obispo, Calif.: Dr. Z Investments, Inc.

Ziemba, William T. and L.C. MacLean. 2011. Using the Kelly Criterion for Investing. In *Stochastic Optimization Models in Finance and Energy*. Bertocchi, Consigli and Dempster, eds. Springer, London.

CHAPTER 23

EXPLOITING EXPERT ANALYSIS? EVIDENCE FROM EVENT STUDIES IN AN INFORMATION-RICH MARKET ENVIRONMENT

MICHAEL A. SMITH

Introduction

Almost ten years ago I conducted a study of horse racing tipsters, focusing primarily on Pricewise of the *Racing Post* (Smith 2003).[1] The study sought to evaluate the record of media forecasters in selecting winners, with returns judged against the benchmark of information efficiency as specified in the efficient-market hypothesis (Fama 1970).

The idea behind the Pricewise column is to identify overlays, that is, horses whose true probability of winning is understated by the market odds. The column recommends an average of between two and three win bets on a typical Saturday, sometimes more than one in a race, with additional midweek tips for prestige meetings, such as Royal Ascot. Selections are mostly restricted to races of high public interest, associated with high betting turnover, which along with the reputation of the column serve to further the goal of maintaining acceptable or high circulation of the newspaper. Pricewise has been portrayed in the media as a feature that leads to significant odds movements, and it claims to be a notably profitable column, which has been a successful feature of the *Racing Post*'s racing coverage since that paper's inception in the mid-1980s. Pricewise's function and status in U.K. racing are comparable to those of, for example, the Value Line research organization with respect to financial investments. Pricewise selections are routinely reported on Channel 4's *Morning Line* racing program as well as its afternoon racing coverage, always in the context of reported odds movements. I visited the *Racing Post*'s offices in Canada Square (Canary Wharf, London) 10 years ago and was impressed by the range of information resources available to journalists working on the paper. Subsequent technological developments can be expected to have further enhanced the arsenal of information sources and analytical capabilities of such media experts.

The prior analysis alluded to above found that newspaper tips appeared to have a significant impact on odds movements from early morning to starting price[2] (SP) and that knowledge of Pricewise selections in particular was a useful predictor of large contractions in odds. Furthermore the performance of such horses was understated by the best available early morning odds, indicated by high positive returns at those odds. The high rates of return to Pricewise horses at the best early odds suggested that the bookmakers' initial appraisal of the chances of these horses was erroneous and that the Pricewise assessment superior.

The evidence suggested that media journalists have a high degree of expertise and/or privileged information in regard to such runners relative to the early bookmaker assessment, at least concerning some runners. The returns at SP for such horses, however, though profitable, were not significantly different to zero, implying semi-strong form efficiency with respect to final odds. A number of other studies have found that pari-mutuel final odds and SP, in relation to U.S. and U.K. horse racing, respectively, are efficient in discounting published factual information and published race forecasts based on expert analysis (Figlewski 1979; Vaughan Williams 2000).

The prior Smith study did not consider evidence concerning market efficiency with respect to arbitrage opportunities as the market progresses from media publication time to race time. Thus a systematic analysis of odds over time, rather than two points in time, is desirable to chart the dynamics of the market. The more quickly early odds assimilate the information held in media forecasts, the more informationally efficient the market is.

The purpose of this chapter is to update, with a new and bigger dataset, the analysis of Pricewise in order to establish whether or not the feature's impressive record has been maintained and to exploit new market information sources that permit insights into the dynamics of horse race betting markets over time to reveal the extent of arbitrage opportunities with respect to the Pricewise selections. This will permit a fuller evaluation of the degree of information efficiency in relation to these "events."

First I will consider an abstract framework for judging the degree of information efficiency with respect to the "events" impacting the market for individual horses, in this case selection by Pricewise. The reader should bear in mind that the odds considered before commencement of the event, when SP rules, are known and fixed at the point of wager. Further, arbitrage/hedging trades of betting assets are now possible via the medium of betting exchanges in those countries where such market formats are legal; the discussion which follows assumes use of the exchanges, and the mechanism of such trading is considered in a later section.

Figure 23.1 charts some of the possible paths of odds over time from publication of media race selections at time *t1* on the x axis to race commencement at *t2*. Odds are

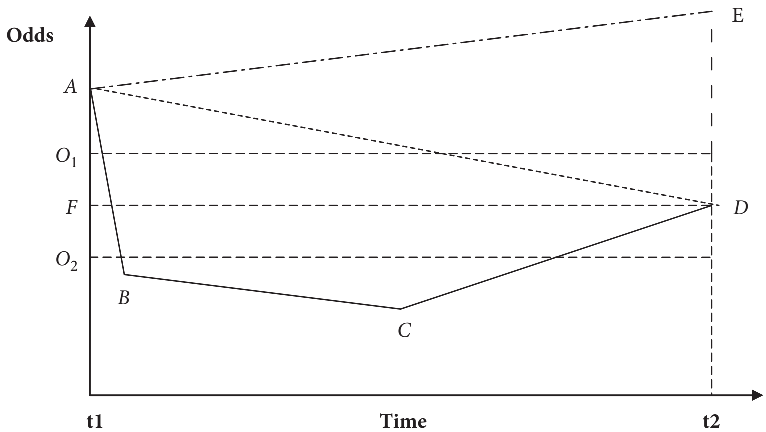


FIGURE 23.1 Alternative paths of odds to SP

represented on the y axis, with fair odds shown as F. Fair odds will vary with selections by odds category, but for any odds level an efficient SP market at *t2* will yield zero profit at odds F over a representative number of selections. Should final odds instead average O_1 long-term profits will be earned, while average odds of O_2 will lead to losses, both cases representing semi-strong information inefficiency.

If odds A at the commencement of trading are known to systematically understate the true chance of the outcome occurring, these odds will quickly contract on publication of the media selection. An informationally efficient market will approximate the path of odds to the event of *AFD*; variations on this L-shaped pattern are consistent with efficiency provided they do not contain secondary or tertiary trends that can be systematically exploited. The alternative path *AD* is not efficient, as it permits bets to be placed at high odds and subsequently laid off at lower odds. This hedging process is equivalent to buying a claim to an asset at a low price then selling at a higher price.

The path *AE* depicts a case in which the market finds the Pricewise arguments for the selection unconvincing and consequently the odds extend to race time; as odds F are fair, the market would be wrong and final odds of E would yield profits. The earlier study (Smith 2003) suggested that, in fact, this class of horse (drifters in the market) leads to profits very close to the average for all Pricewise selections at SP, suggesting that the error of judgment lies in the Pricewise column, and the market is able to correct for this error. In figure 23.1 this would be represented by a line parallel to *AE* culminating in point D.

Finally, *ABCD* illustrates a situation where the information in the selection is quickly assimilated by the market, followed by a period of overbetting and later correction, perhaps by hedging trades as early back-to-win traders lay off at lower odds from point C onward.

Clearly these paths, and other variations, may be observed in specific cases. For semi-strong information efficiency to exist, however, profitable arbitrage and hedging opportunities must not be systematic and predictable. In the results section below the actual path of price outcomes for a sample of Pricewise selections are considered.

The Betting Exchanges: An Information-Rich Market Environment

This section considers the nature and functionality of betting exchanges as a prelude to analyzing market information relating to Pricewise selections. The exchanges are information-rich in terms of odds, volumes traded at different odds, and sequence of trades over time and therefore permit clear insights into the betting period before and after publication of the Pricewise column. Similar market transparency is not evident in the corresponding bookmaker markets largely because of lack of information about trading volumes and uncertainty about the availability of odds (bookmakers frequently impose strict bet limits in relation to advertised odds pertaining to Pricewise selections).

In 2000 a small number of experienced city traders combined their expertise in sports betting to create Betfair.While not the first betting exchange, Betfair quickly established itself as the market leader—it currently has approximately 90 percent of the betting exchange market. Its turnover grew exponentially in the early years—growth generated not only from market share taken from bookmakers but, in addition, from the influx of a new cadre of traders attracted to the potential profits to be reaped from exploiting price volatility in arbitrage and hedging activities; such growth could not be maintained indefinitely, and Betfair has now consolidated its position as the leading exchange and a permanent and highly influential market mechanism in sports betting. Betfair is the eBay of betting; it permits the many-to-many double auction of state-contingent claims based on sporting events. It is Internet based and has spawned a whole new derivative software industry in the development of associated applications and trading software that enhance and customize the primary properties of the site itself. This related software permits market monitoring, order placing, direct trades, and hedging to be executed on a range of mobile electronic devices, including mobile phones and iPads in addition to desktops and laptops. Functionality will now be outlined, primarily by reference to Betfair; other exchanges have similar characteristics with minor variations.

A betting exchange does not act as market maker; rather it fulfills broking and exchange functions, displaying odds offered by layers and requested by bettors and permitting execution of bets based on these odds. Layers wager that an outcome will not happen, for example, a horse will lose, whereas backers wager to win, that is, that the specified outcome will occur. The exchange itself assumes no risk, its income generated from commission charged to clients, typically at a rate between 2–5 percent (most

Betfair clients pay 5%, whereas high-volume traders pay less) applied to net winnings on an event (horse race, cricket match, game of tennis—here, horse racing).

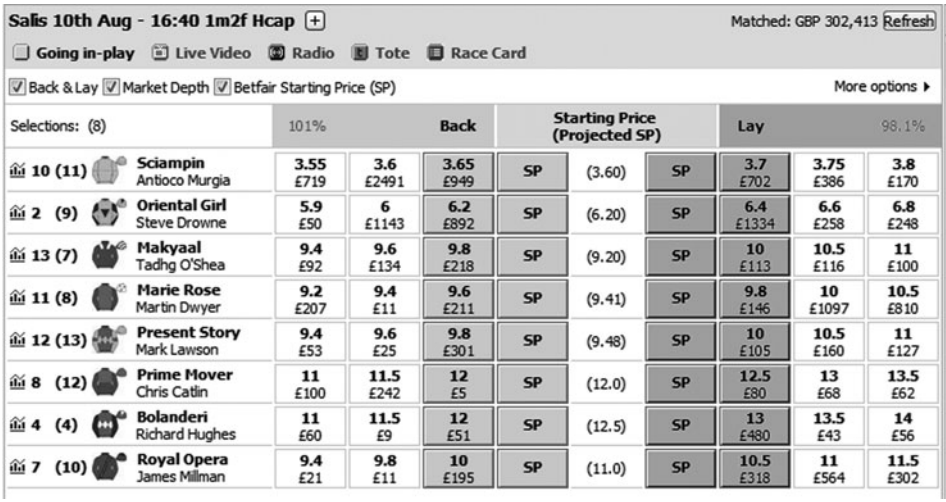
Clients can back a horse to win, lay a horse at specified odds, or request odds, all at client-specified stake limits. Stakes are pooled by odds value so that individual matched trades are not necessary, reducing the incidence of thin markets; trades and parts of trades are allocated on a first-requested and first-offered basis. Betfair securely holds monies corresponding to net potential liabilities, which must be fully funded from clients' accounts in advance of the event, thus ensuring the integrity of trades.

Once a race commences the market goes "in play." Unmatched trades can be withdrawn at this point or kept open. In play, that is, in running markets are characterized by bigger aggregate margins (higher overround on the back side and underround on the lay side), more frequent and rapid price movements, and greater price volatility than pre-event markets.

Odds changes can be monitored directly or with dedicated trading software, some of which is freely available from the Internet; much of this software downloads market information from Betfair, updating at user-specified intervals of as little as once per second. Automated trading is possible with such software, governed by decision rules specified by the trader, often triggered by associated spreadsheets linked to the software.

The ability to back and lay the same outcome also means that a range of hedging strategies, and arbitrage between the exchange and bookmaker markets, can be employed by traders. Such extensive functionality, price transparency, and low commission (relative to that implicit in bookmaker overround averaging over 1.5 percent per runner in the United Kingdom) draws in serious bettors, reducing the proportion of turnover attributable to casual bettors. As a consequence the market is more efficient, with low margins on the back and lay side. Price competition increases with volume as the market progresses, thereby reducing the amount of bias; Smith, Paton, and Vaughan Williams (2006) showed that unlike bookmaker horse race markets, Betfair odds contain very little longshot bias. This result is consistent with the literature on the economics of auctions (e.g., Klemperer 1999, 2004), which suggests that the decentralized nature of the decision-making processes in such markets can accomplish the aggregation of information in a very efficient manner.

Figure 23.2 shows a typical Betfair horse race market shortly before going in play. This race, run at Salisbury, was a modest handicap, which would attract relatively low betting volumes in the bookmaker market compared to the higher class races from which Pricewise selections are typically drawn. Nonetheless, the volume of matched trades on Betfair (top right) for this race exceeded £300,000. To put this in perspective, the tote win pool for this race amounted to only £8,428. The volume traded on the exchange is so much greater than that in the tote pool because of the large amount of trading that occurs in the exchanges, involving hedging strategies of one form or another enabled by the possibility of backing and laying the same outcome.



Salis 10th Aug - 16:40 1m2f Hcap [+] Matched: GBP 302,413 Refresh

Going in-play | Live Video | Radio | Tote | Race Card

Back & Lay | Market Depth | Betfair Starting Price (SP) | More options ▸

Selections: (8)	101%		Back	Starting Price (Projected SP)			Lay		98.1%
10 (11) Sciampin, Antioco Murgia	3.55 £719	3.6 £2491	3.65 £949	SP	(3.60)	SP	3.7 £702	3.75 £386	3.8 £170
2 (9) Oriental Girl, Steve Drowne	5.9 £50	6 £1143	6.2 £892	SP	(6.20)	SP	6.4 £1334	6.6 £258	6.8 £248
13 (7) Makyaal, Tadhg O'Shea	9.4 £92	9.6 £134	9.8 £218	SP	(9.20)	SP	10 £113	10.5 £116	11 £100
11 (8) Marie Rose, Martin Dwyer	9.2 £207	9.4 £11	9.6 £211	SP	(9.41)	SP	9.8 £146	10 £1097	10.5 £810
12 (13) Present Story, Mark Lawson	9.4 £53	9.6 £25	9.8 £301	SP	(9.48)	SP	10 £105	10.5 £160	11 £127
8 (12) Prime Mover, Chris Catlin	11 £100	11.5 £242	12 £5	SP	(12.0)	SP	12.5 £80	13 £68	13.5 £62
4 (4) Bolanderi, Richard Hughes	11 £60	11.5 £9	12 £51	SP	(12.5)	SP	13 £480	13.5 £43	14 £56
7 (10) Royal Opera, James Millman	9.4 £21	9.8 £11	10 £195	SP	(11.0)	SP	10.5 £318	11 £564	11.5 £302

FIGURE 23.2 Example of a Betfair horse race market

The side of the display headed "Back" shows the best three prices available for each runner being offered by layers to clients wishing to make win bets, with stake limits. The stake limits are the pooled amounts various clients are prepared to accept on back-to-win bets at the specified odds values. Prices on the exchange are expressed with a unit stake included, so at this point in the market the best bet-to-win odds available for Sciampin are 2.65/1 (3.65 minus 1, to a nominal unit stake) up to a maximum stake of £949 (with a minimum stake requirement of £2). Should a client wish to offer odds of 5.3 to 1 (price of 6.3) against Oriental Girl to a maximum of, say, £10, this would become the new best price. Alternatively the client could specify an offer to lay at lower odds, which would be pooled with existing values. The sum of probabilities on the back side in the illustrated market stands at 101 percent, that is, 1 percent overround, ensuring that a profit cannot be earned by backing all runners to stakes proportionate to odds probabilities. Should the book total less than one momentarily, such "dutching" strategies will quickly mop up the higher odds across the field until an overround position is restored.

Similarly, on the lay side, prices represent invitations, or requests, to lay at higher odds than are available on the back side—clients who place these orders wish to bet-to-win but are not prepared to do so at odds indicated on the back side. Other clients may choose to lay a horse at these odds should they believe that they overstate the chances of that horse or as part of a hedging strategy. For example, a client who believes that the true chance of Prime Mover winning is 20/1 may choose to lay that horse at 11.5/1 (12.5 minus 1), marginally higher than the best odds currently on offer, to a stake of up to £80. The sum of probabilities on the lay side, based on the lowest odds requested, stands at 98.1 percent, ensuring that a profit cannot be guaranteed by laying all horses to stakes proportionate to probabilities. Should this value temporarily exceed one, profit-seeking traders will quickly eliminate the excess, laying all horses in the race at the requested odds. Further details of betting exchange characteristics can be found in Jones et al. (2006), Smith and Vaughan Williams (2008) and Koning and van Velzen (2009).

Bettors may of course place conventional win bets on the exchanges as they would with a bookmaker. The advantage to them is that on average Betfair odds exceed bookmaker odds by more than the commission payable should the horse win; this is particularly so in the case of long shots (horses with a low probability of winning). In contrast, many traders attempt to lock in value arising from price movements, fully or partially hedging back or lay bets against the field.

Consider an initial back bet of stake s_B made on a horse at exchange price p_B, corresponding to odds o_B plus 1, and at a later time the back bet is partly or fully hedged by laying the same horse at the lower price, p_L equal to odds o_L plus 1.

Assume that the trader aims to carry no net exposure to liabilities and that liquidity at p_L is sufficient to achieve this aim. Let α be a hedging preference coefficient reflecting the degree of hedging across the field sought by the trader, such that $0 \leq \alpha \leq 1$.

Finally let r be the ration of best back to best lay prices, $\frac{p_B}{p_L}$.

The value of s_L wagered at p_L is a function of the initial back bet, the relative back and lay prices, and the hedging preference of the trader, as expressed in equation 23.1:

$$\begin{aligned} s_L &= s_B r - \alpha \left(s_B r - s_B \right) \\ &= s_B \left(r - \alpha \left[r - 1 \right] \right). \end{aligned} \tag{23.1}$$

If $\alpha = 0$ the resulting lay trade hedges the initial bet fully against the field and the outcome is not state-contingent. Alternatively, if α takes a positive value the paired trade closes the exposed position, but the monetary outcome remains state-contingent. In the former case ($\alpha = 0$) the trader will win an equal amount whatever the outcome of the race. If $\alpha = 1$, barring a dead heat, there are two possible outcomes: zero profit/loss if the horse loses and a profit if he horse wins. The profit π_W, should the horse win, is expressed in equations 23.2 and 23.3.

$$\pi_W = s_B \left(p_B - 1 \right) - s_L \left(p_L - 1 \right). \tag{23.2}$$

Substituting equation 23.1 in equation 23.2,

$$\begin{aligned} \pi_W &= s_B \left(p_B - 1 \right) - s_B \left(r - \alpha \left[r - 1 \right] \right) \\ &= s_B \left(\alpha \left[p_B - p_L \right] \right) - \left[r - 1 \right] \left[1 - \alpha \right]. \end{aligned} \tag{23.3}$$

Should the horse lose, the net profit on the paired trade, π_L, is expressed in equations 23.4 and 23.5.

$$\pi_L = s_L - s_B. \tag{23.4}$$

Table 23.1 Outcomes of a Matched Trade with Various Degrees of Hedging

α	p_B	p_L	s_B(£)	s_L(£)	π_W(£)	π_L (£)
0	5	4.9	490	500	10	10
0.2	5	4.9	490	498	17.8	8
0.4	5	4.9	490	496	25.6	6
0.6	5	4.9	490	494	33.4	4
0.8	5	4.9	490	492	41.2	2
1	5	4.9	490	490	49	0

Substituting equation 23.1 in equation 23.4 to yield an expression containing α,

$$\begin{aligned}\pi_L &= s_B\,(r-\alpha\,[r-1]) - s_B\\ &= s_B\,(r-\alpha\,[r-1]) - 1\\ &= s_B\,(r-1)\,(1-\alpha)\,.\end{aligned} \tag{23.5}$$

Table 23.1 shows a simple application of the above formulas, assuming bet and lay prices for a horse of 5 and 4.9, respectively, with an initial back stake of £490. The impact on profits of different values of α is shown for the outcomes of the horse winning or losing. The anticipated direction of price movement can of course be reversed, with a lay bet being executed first followed by a back bet to complete the paired trade. Rearrangement of the above expressions permits calculation of back stake levels and their consequences for profits in such cases.

Variations of the above formal expressions are built into algorithms contained in commercial derivative software that interfaces with the exchange markets, permitting the various trader preferences and bet instructions to be executed.

The success of the trader primarily depends on the ability to predict a price movement and its direction, when it will occur, and how long to leave exposed trades before closing a position. Should a trader predict and bet on a price movement wrongly, he or she can exit the trade as per the staking and profit expressions above, but losses will be incurred for one or other of the win/lose race outcomes. These factors will be considered below in relation to the charts showing the dynamics of prices over time for a number of Pricewise selections.

Results and Discussion

The original sample from the Smith (2003) study, covering the period 1998–2000, is referred to as sample 1; for the new database covering 2002–2011, sample 2.

Table 23.2 Statistics for Pricewise Selections, 1998–2000 and 2002–2011

	Sample 1 (1998–2000)	Sample 2 (2002–2011)
N	344	591
Average odds (early max)	14.3/1	12.9/1
Average odds (SP)	9.7/1	8.4/1
Average winning odds (early max)	10/4/1	8.9/1
Average winning odds (SP)	6.8/1	5.8/1
Winning strike rate (winners/runners)	0.1424	0.1506
Price movement from early max odds to SP	−0.043*** (0.0025)	−0.054*** (0.0022)

	Sample 1 Returns		Sample 2 Returns	
	Early Max	Starting Price	Early Max	Starting Price
Unit stake	0.617	0.118	0.497	0.023
Unit stake after deductions	0.520	0.051	0.497	0.023
Unit stake (weighted by odds probability)	0.559*** (0.2058)	0.113 (0.1467)	0.421*** (0.1441)	0.115 (0.1046)
Unit stake (weighted), after deductions	0.466** (0.1934)	0.046 (0.1379)	0.421*** (0.1441)	0.115 (0.1046)

Note: Values in parentheses are robust standard errors.
** denotes $p = 0.05$.
*** denotes $p = 0.01$.

Comparisons between sample 1 and sample 2 are initially facilitated by reference to bookmaker odds, as exchanges did not exist during the sample 1 period.

The descriptive statistics of odds and returns are summarized in the first seven rows of table 23.2. Average odds of Pricewise horses at both best early odds (early max) and at SP are lower in sample 2 than in sample 1, possibly an indirect consequence of the reduction in average field sizes in U.K. racing between the two periods (Smith and Vaughan Williams 2010). This tendency is also reflected in the lower average SP of winning Pricewise selections and increased strike rate in the latter period. All these aspects can be explained by declining runners per race as opposed to strategic changes in the approach of the column, which are otherwise not evident.

Odds movements from early max to SP for each Pricewise horse were recorded in both datasets by means of a measure, p_m, adapted from Law and Peel (2002), shown in equation 23.6.

$$p_m = \log\left(\frac{1}{1-p_1}\right) - \log\left(\frac{1}{1-p_2}\right), \tag{23.6}$$

where p_1 is the odds probability corresponding to the early maximum odds available for a Pricewise selection and p_2 is the odds probability at SP. The virtue of this method is that it places greater emphasis on movements from shorter odds and therefore more closely approximates a trade-weighted measurement. A negative value of p_m indicates that the Pricewise selection contracted from early odds to SP; the mean values for samples 1 and 2 are −0.043 and −0.054, respectively. To add perspective to these numbers, a value of −0.05 corresponds to a contraction from 20/1 to 10/1 against or from 6/1 to 9/2 against. The odds against Pricewise selections therefore typically contract substantially from early morning to SP.

Table 23.2 shows that Pricewise selections during both periods earned positive profits, to a unit stake, of around 50 percent at early max odds, even after deductions in the case of sample 1.[3] The corresponding returns at SP are much lower, as a consequence of strong odds contraction, but remain positive, even after deductions in sample 1. In order to generate measures of statistical significance the unit stake returns were weighted by the reciprocal of the odds plus one to reduce the impact of heteroskedasticity. All returns at early max odds are significant at $p = 0.05$ or better, though the SP positive profits are insignificant. The before deductions SP returns are similar between the two samples, however, and are suggestive of a semi-strong form inefficiency.

Table 23.3 shows the results of tests of independence between the two sets of sample returns at both sets of odds. There is no evidence of significant differences in mean returns between the samples, although the contraction of odds from early max to SP in sample 2 appears to have become more marked than in sample 1.

Overall the evidence from this comparative study of returns strongly suggests that it is possible for an expert to identify wrongly priced horses yielding high positive profits prior to public dissemination of the selection decisions. The possibility of subsequent undervaluation of the information contained therein at final market odds also remains open, given the continued positive SP returns across the two samples, their lack of statistical significance notwithstanding.

Exchange prices and traded volumes were acquired for 193 Pricewise horses that ran during the period September 2010–July 2011, a subset of sample 2 outlined above, with

Table 23.3 Difference of Sample 1 and Sample 2 Means: Independent Sample *T*-test Results

	Early Max (t-values)	Starting Price (t-values)
Unit stake (weighted by odds probability)	0.27	−0.047
Unit stake (weighted), after deductions	−0.092	1.10
Price movement from early max odds to SP	−24.03***	

the paths of exchange prices charted from the commencement of each market through time of publication of the Pricewise selections and the remainder of the pre-event trading period. Detailed market data were sourced from Fracsoft, an organization which employs specialist software to record the exchange markets for most of their duration. Recordings begin typically between 8.30 A.M. and 10 A.M. on the day of the race, tracking the remainder of the pre-event period and continuing into the in play segment through to the conclusion of the race. Only the pre-event data were required for the current study.

The problem with this data source is that the column is published, both in the hard copy *Racing Post* and on the corresponding website, some time prior to the commencement of Fracsoft recordings. Newspaper outlets nationwide typically will hold stocks of the paper from 5 or 6 A.M., early editions being available in London and its region much earlier. A market recording that commences at 8.30 A.M. will therefore not trace impacts at the point of publication. To overcome this problem, additional price data from early in the pre-event market period were acquired direct from Betfair's historical data electronic archive for the analyzed races. These additional data have their own limitations in that for each horse only first and last times traded for each price are given along with aggregate amounts traded by price per horse across the whole of the market period for these prices. Prior to the time of the detailed Fracsoft recordings, therefore, the corresponding segments of the charts derived from the data exhibit node points in an envelope encompassing first- or last-price values traded over time, with no details of price fluctuations over time at values already traded. Nonetheless it is possible to reconstruct primary price movements during the early market period sufficient to observe the ostensible impact of Pricewise selections before and after their publication.

Figures 23.3 through 23.7 depict exchange price movements for five Pricewise selections chosen from the 193 charts constructed as exemplars of frequently occurring patterns. There are some clear characteristics evident in the pattern of exchange price movements that apply to the vast majority of Pricewise selections. First, prior to publication of the column's tips they are often subject to initial speculative offers of low odds when the market opens, usually on the day before the races are scheduled to be run. Of course at this point there is no public knowledge of the Pricewise tips, which are published the following day. The offer of poor odds at the start of the market for the majority of a field of runners, with resulting high levels of overround, is commonplace on the exchanges. This tendency can be seen in four of the five exemplar charts for Pricewise horses. As competition and liquidity increase, more realistic prices for most runners in the typical field will be offered quite quickly following this initial period, again evident in the charts.

Following this early period of speculative pricing, inspection of the path of prices for Pricewise horses reveals a second phase during which there is, in the majority of cases, a sustained contraction in exchange odds. Of the 193 horses for which charts were constructed, 85 percent revealed a marked reduction in odds at this stage of the market. The exact timing of such market support varies considerably, as shown in table 23.4, which details the distribution of days/times when these sustained odds reductions

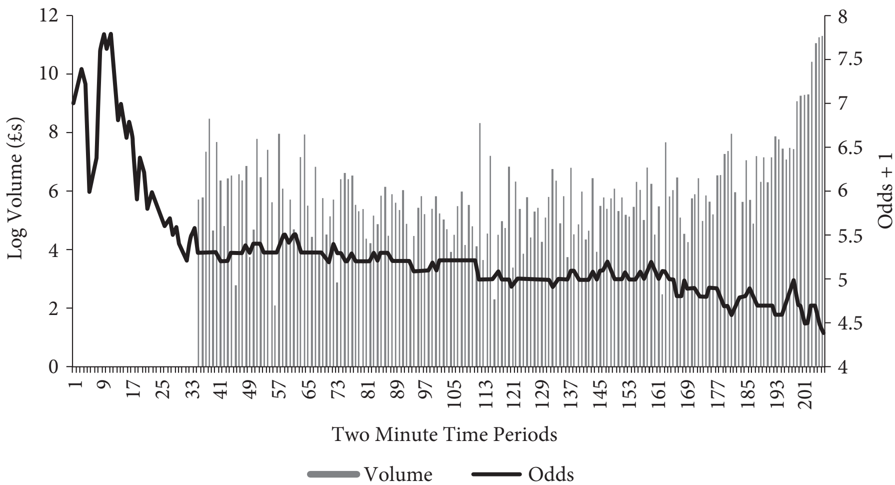


FIGURE 23.3 Steamer
Horse: Desert Law

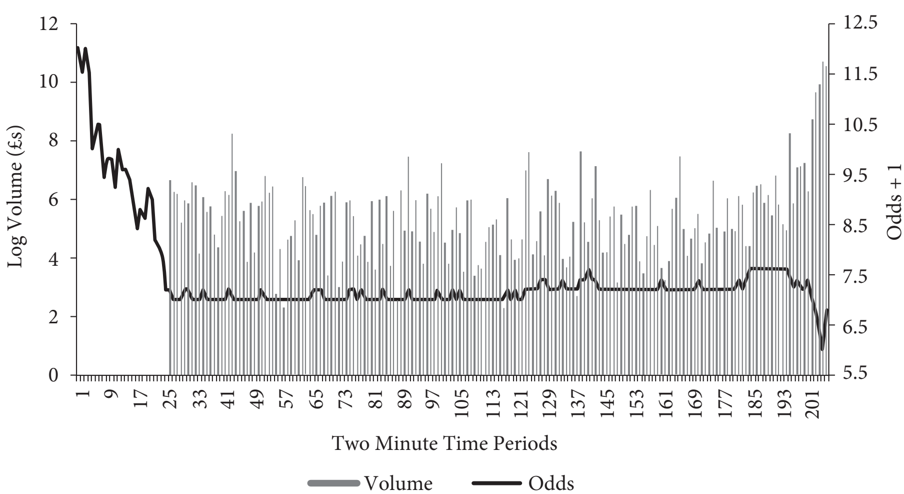


FIGURE 23.4 S-Shaped
Horse: Minella Four Star

commence; the first and last time traded data accessed directly from Betfair enable easy identification of these changes in primary trend. The five exemplar charts exhibit marked odds contractions of this type, generally following the initial speculative pricing period and prior to the period of Fracsoft recording.

Only 15 percent of observed charts exhibited no significant contraction; this category also includes observed contractions which were judged to be consistent with subsequent

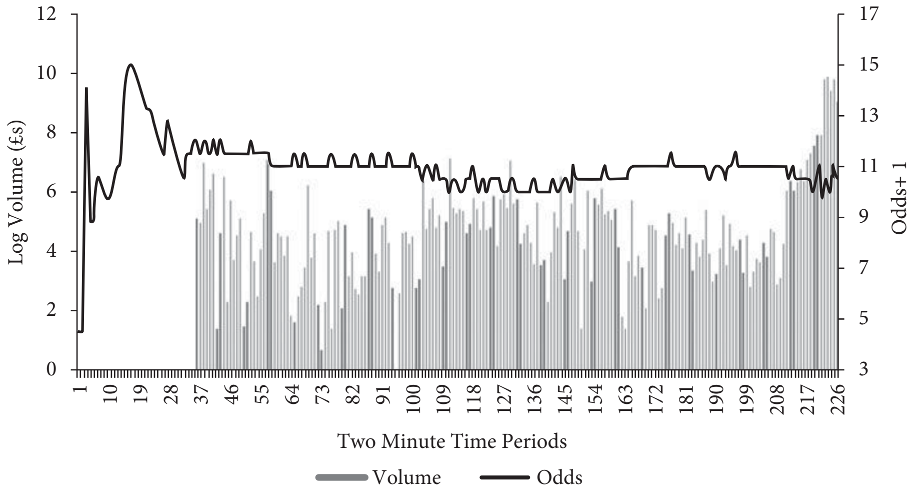


FIGURE 23.5 L-Shaped
Horse: Hawkeyethenoo

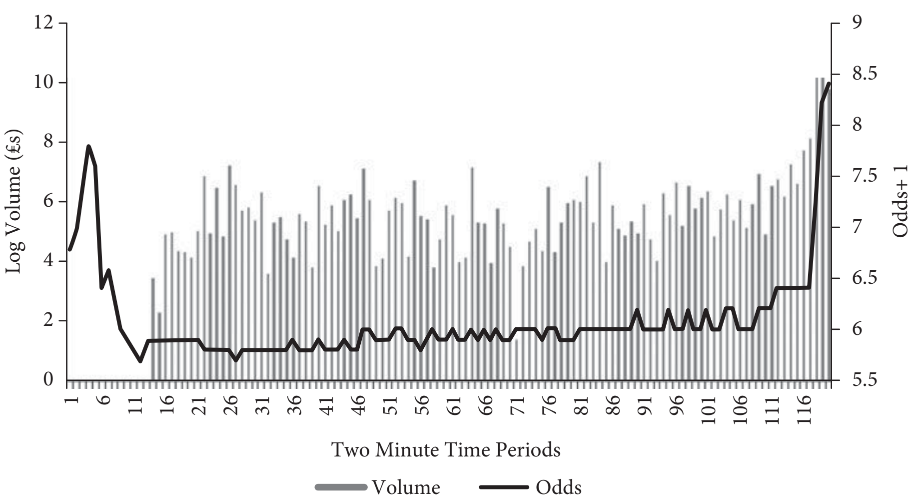


FIGURE 23.6 Cup-Shaped
Horse: Pastoral Player

price volatility and trends in the odds over time for the horses concerned and for that reason were not counted as significant.

To further consider the remaining 85 percent of observations that did contract, a word on distribution arrangements for the *Racing Post* is in order. A consequence of the overnight declaration system whereby runners for the following day's races must be declared by trainers by midday on the day prior to the races is that full and focused

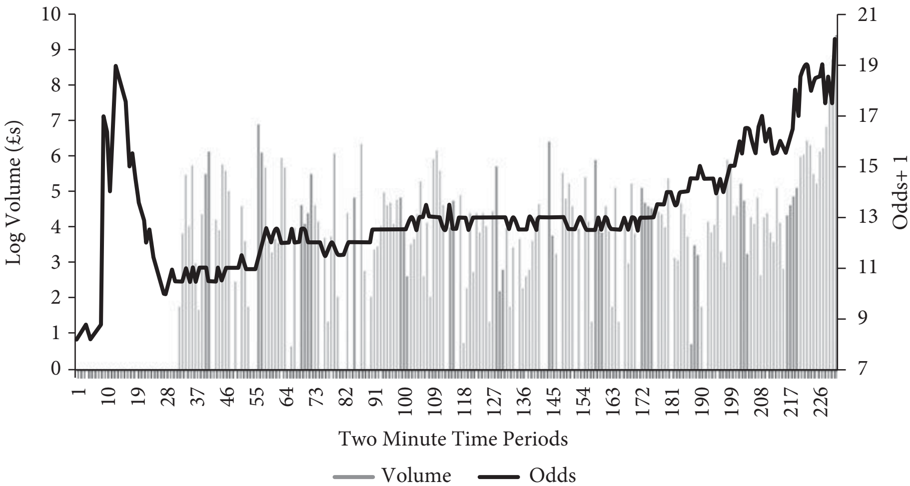


FIGURE 23.7 Drifter
Horse: Empirico

coverage of a day's race meetings by *Racing Post* cannot feasibly be published prior to the day of the races concerned. *Racing Post* hard copy is distributed early on the morning of the races to be run that day, typically as early as 5 A.M. via newspaper outlets nationwide. Distribution in London may begin earlier than this, and therefore midnight to 5 A.M. is assumed here as the period of initial distribution. *Racing Post* tips are also made available online but not until well after hard copy distribution commences so as not to compromise circulation figures of the newspaper.

Table 23.4 shows that only 18.6 percent of charts surveyed exhibited a marked contraction in odds commencing in the post-distribution period (from midnight of the day of the races onward). As many as 11.4 percent of the observations exhibited sharp

Table 23.4 Starting Time of Odds Contractions (Primary Trend)

Time	Frequency	Percentage of Total Charts Surveyed
Two days before the race	22	11.4
Morning of day before race	9	4.7
1200<1600, day before race	34	17.6
1600<2000, day before race	74	38.3
2000<midnight, day before race	16	8.3
Midnight<0500, day of race	3	15.5
0500 or later, day of race	6	3.1
No evidence of significant contraction	29	15.0
Total	193	

contractions in price commencing two days before the races concerned were run, typically being races attracting substantial ante post betting interest. Whereas Pricewise does give ante post recommended bets for some races, the present study excluded these, being careful to include only tips published on the relevant race day, the better to identify potential cause and effect in an event study methodology. Most primary trend price contractions commenced in the period from noon to 8 P.M. (1200 < 2000) of the day prior to publication, accounting for 56 percent of the charts surveyed. Most of the subsequent price contraction in such cases appears to have occurred by the time Fracsoft recording commences, typically between 8.30 and 10 A.M. on race day, at which point a new lower price floor characterized by less volatility compared to trading prior to publication is invariably evident, at least until the final 10 to 20 minutes of the market. No explicit control against corresponding non-tipped horses was carried out, but the fact that overround on odds offered decreases over time in these markets (levels of 50% or more are commonplace early in the market) to approaching zero at race time implies that price extensions over time, not contractions, are the norm for runners in general. More significantly, actual trades as opposed to offers on average exhibit little trend across the market as a whole.

There are several possible explanations for these strong price contractions prior to publication.

1. One of the key criteria in the selection process for the Pricewise column may be existing evidence of strong market support; that is, odds contraction precedes selection, in which case the assumed cause and effect (that Pricewise selection causes the relevant odds to contract) is reversed.
2. Pricewise selection and odds contraction are strongly associated but independent, implying that market participants employ information search and analysis techniques equivalent to those applied by the Pricewise column.
3. The exhibited price contractions arise from betting by insiders, having privileged knowledge of Pricewise horses prior to publication, either to retain as win bets or to subsequently hedge following the anticipated reduction in odds.
4. The explanation may lay in a combination of the above, as they are not mutually exclusive.

Whatever the explanation, it would appear that watching the market in its early stages for primary trend contracting odds, against the tide of general price extension through competition between layers, may give signals as to which horses may turn out to be Pricewise selections.

The phases of the market considered thus far have been confined to the price envelope containing first and last odds value trades[4] from Betfair's historical data archive. The discussion now moves to an examination of the remaining period of pre-event trading, for which the charts depict exchange price values (odds plus one), traded at two-minute intervals, and volumes wagered during those intervals. On the exemplar charts, figures

Table 23.5 Advised Odds, SP, and Amounts Traded on Pricewise Horses in the Exemplar Charts

Pricewise Selection (chart-depicted horses only)	Bookmaker Odds Advised and (SP)	Total Traded on Horse Prior to Commencement of Fracsoft Recording (£000)	Total Traded on Horse at End of Pre-Event Period (£000)	Ratio of Totals Traded (end-period/ Fracsoft-start)
Desert Law	6/1 (3/1)	76.05	458.14	6.02
Minella Four Star	10/1 (5/1)	45.37	231.50	5.10
Hawkeyethenoo	16/1 (8/1)	12.53	129.51	10.34
Pastoral Player	6/1 (13/2)	31.13	146.27	4.70
Empirico	14/1 (14/1)	18.08	55.91	3.09

23.3 through 23.7, the primary vertical axis measures volume traded expressed as a natural logarithm (to make the scale manageable), and exchange prices (odds plus one) are expressed on the secondary vertical axis. Table 23.5 shows the amounts traded on the exemplar horses to give perspective to the log scale and trading volumes over time.

It is clear from the charts that in each case volume traded accelerates substantially during the final stages of the pre-event market; these characteristics are typical of the 193 horses for which charts were constructed. Inspection of the charts in total lends itself to identification of a typology of market dynamics; five generic categories were created to represent market types commonly observed. Each Pricewise horse was allocated to one of the categories (or more than one if the market appeared to be a hybrid, which was the case for about 5% of horses). The categories were named as follows: Steamer, S-Shaped, L-Shaped, Cup-Shaped, and Drifter. These categories are now described.

Steamer: Primary trend of exchange price (odds plus one) throughout the pre-event market period is downward. This trend is often, but not always, steeper in the pre-Fracsoft recording period. Steamer is exemplified by figure 23.3.

S-Shaped: Following a significant decline in exchange price in the pre-recording period, a period of price stability and low volatility extends throughout the day. As volume increases toward the end of the pre-event period, a further primary trend decline in odds occurs. S-Shape is exemplified by figure 23.4.

L-Shaped: In most cases a significant decline in exchange price occurs in the pre-recording period, though in a small number of cases the exchange price extends significantly instead. A long period of price stability and low volatility continues throughout the day, to the end of the pre-event period. L-Shaped is exemplified by figure 23.5.

Cup-Shaped: Following a significant decline in exchange price during the pre-recording period, a long period of price stability and low volatility follows throughout the day. As volume increases toward the end of the pre-event period, the early primary trend decline in odds is reversed. A small number of cases approximates the stereotypical quadratic cup shape. Cup-Shaped is exemplified by figure 23.6.

Drifter: There may or may not be an initial primary trend reduction in odds prior to the Fracsoft recording period. The recording period is characterized by a primary trend extension of exchange price throughout the pre-event market period; that is, odds increase over time to the conclusion of the market. Drifter is exemplified by figure 23.7.

The frequency distribution of this allocation of horses to categories is shown in table 23.6. Traders who back a Pricewise selection that subsequently contracts in odds early in the market, which occurs in the majority of observed selections, can exploit their position by laying, either fully or partially or in a staged sequence of hedging bets later in the pre-event period. However, as 85 percent of these horses exhibit the bulk of primary trend contractions prior to publication, this strategy is easier to identify than to execute.

In terms of the market shapes suggested above these are primarily distinguished by the path of prices during the Fracsoft recording period. The L-Shaped category offers the best fit with an efficient market in terms of arbitrage opportunities. Table 23.6 shows this to be empirically the mean outcome, with the Steamer and S-Shaped categories (exhibiting odds contractions during or late in the market) and the Cup-Shaped and Drifter categories (exhibiting odds extensions during or late in the market) distributed approximately normally around the L-Shape type. This distribution is consistent with market efficiency, from an arbitrage over time perspective, suggesting the path of prices for the sampled horses to be a random variable normally distributed around the L-Shaped pattern.

Traders not fortunate enough to place win bets prior to publication are therefore presented with the usual dilemmas when attempting to predict price movements and associated trading strategies. Should the trader back Pricewise selections shortly after publication, hoping to lay at lower odds later on? If so they are betting on the emergence of the S-Shaped or Steamer patterns or are relying on favorable random movements sufficient to offset the margin between back and lay prices. Or should the trader lay Pricewise selections shortly after publication, hoping to back at higher odds later on?

Table 23.6 Pricewise Horses Distributed according to Market Shape Typology

Category	Frequency	Percentage of Total Charts
Steamer	36	18.7
S-Shaped	30	15.5
L-Shaped	49	25.4
Cup-Shaped	48	24.9
Drifter	30	15.5
Total	193	

This strategy relies on the emergence of the Cup-Shaped or Drifter patterns or, again, favorable random price changes. Implicit in this strategy is the assumption that the market has overreacted to the Pricewise information and that a market correction will occur in the form of drifting odds. The empirical evidence suggests, however, that the exchange prices at the end of the pre-event period accurately reflect the value of Pricewise information: for the subsample from which charts were derived the Betfair SP return was −0.05 percent, factoring in 5 percent commission (compared to a loss of 9.9% at bookmaker SP). The distribution of charts around the L-Shaped market type, and a zero exchange return at final odds, imply no systematic overreaction or underreaction earlier in the recording period of the pre-event market.

Another common strategy employed by exchange traders, known as scalping, seeks to exploit volatility in price movements. Traders can amass frequent low-margin positive returns from such trading during periods when the price exhibits no trend. The problem with this trading approach is that gains can quickly be wiped out by the emergence of a trend in the prices; in addition, traders have to be skillful in timing decisions about the optimum point in time to exit trades and at what levels of profit and loss open positions should be closed. The evidence is that many traders find this balance difficult to achieve; empirical studies of financial asset trading suggest that traders hold on to losing positions too long and not long enough to winning positions—the so-called disposition effect (Shefrin and Statman 1994; Weber and Camerer 1998). This tendency is explained by prospect theory, which posits that the utility function for gains is less steep than that for losses (Kahneman and Tversky 1979). Scalping can be seen simply as noise trading; it is not central to the pursuit of value from Pricewise selections but does explain much of the market activity between step changes in price observed in the middle segments of the recording periods in many of the charts.

Traders employ analyses of volumes offered, requested, and traded to predict the direction and extent of exchange price changes. Weight of money indicators are routinely calculated and updated in the derivative trading software available on the market. A common technique is to monitor the relative percentages of money associated with price offers (back side) and price requests (lay side), including the best three prices on each side. For example, considering Sciampin in the Salisbury race illustrated in figure 23.2, the total amount of unmatched orders for the best three prices on either side is £5417, of which £4159, or 76.8 percent, is on the back side, with the remaining 23.2 percent on the lay side. Prima facie, many traders would interpret these specific values as a predictive indicator of imminent odds extension.

Significant primary trend price contractions (S-Shaped) or extensions (Cup-Shaped) are likely to occur in the final stages of the pre-event market; consistent success in predicting the direction of these changes would clearly be of considerable trading value. Bookmakers are permitted to run betting exchange accounts and frequently use these to lay off unwanted liabilities that may unduly balance their books. In the final stages of the pre-event market bookmaker odds movements (though not the associated volumes) are well publicized; evidence of marked price contractions in the bookmaker market for specific horses may signal corresponding falls in exchange price. If such support in the bookmaker market is for the Pricewise horse(s) in a race, then volumes finding their way into the exchanges to reduce bookmaker liabilities may considerably outweigh volumes arising from traders hedging their early back to win positions on Pricewise. Such a scenario will lead to downward pressure on prices and a subsequent S-Shaped market type. Conversely, if late support in the bookmaker and exchange markets has been for other horses in the field, price extension of the Pricewise horse(s) may occur, especially if higher than normal exchange trading volumes associated with the early odds contraction of Pricewise selections leads to high levels of hedging.

Conclusions

This chapter has reported the results of a study of returns and odds movements associated with an influential media horse racing pundit. In the comparative study of odds at two points in time, early morning and at race time, both samples considered suggest that expert analysis can yield highly profitable, statistically significant returns prior to public dissemination of the results of such analysis. These "abnormal returns" arise from bookmaker mispricing and the fact that the market drives down returns at SP as a consequence of strong odds contraction. Even at SP, however, profits of over 10 percent remain. While not statistically significant (the associated p value is 0.18 for the SP returns pooled over both samples), these positive returns persist over both samples and are very similar, a result at variance with most large-sample studies of betting markets selection strategies, which invariably yield losses. The persistence of positive returns over time is indicative also of a lack of learning from past experience. On these grounds the results at both points in time arguably permit a conclusion of semi-strong form inefficiency.

The second part of the chapter considered the dynamics of the market from early odds to SP. This was based on charts derived from betting exchange data, and there appears to be prima facie evidence that exchange traders are collectively skilled in accurately incorporating the information contained in the Pricewise column into odds; given the broadly normal distribution of path-of-price types around that typifying market efficiency, there do not appear to be systematic opportunities for profitable hedging strategies. It may be, however, that skillful traders employ more sophisticated methods based on, for example, weight of money, to profitably exploit the expert analysis of tipsters.

More robust statistical measures would be possible with a full record of Pricewise selections going back over 25 years. These are available via the archives of the *Racing Post*, and perhaps one day I will indulge the wish to complete the set. In the meantime, the remarkable reputation of the Pricewise column continues to exert a powerful influence on U.K. horse race betting markets.

Notes

1. The *Racing Post* is the leading racing newspaper by circulation in the United Kingdom. The study in question employed the nom de plume "Winsome" for Pricewise; in retrospect the reasons for being so guarded in identifying the real name of the feature are unclear.
2. Starting price is the odds value at which bets are settled in the absence of a fixed-odds agreement between bookmaker and bettor. The SP for each horse is determined by inspectors who monitor the range of on-course odds available from competing bookmakers close to the time of the race.
3. During the period of sample 1 bookmakers made deductions to cover the betting tax of 6 percent of winnings and an additional 3 percent to cover operating costs; this deduction ceased when betting tax was abolished in October 2001.
4. Many of the last observed odds values do not appear in the early stages of the charts as they are captured in the recorded Fracsoft phase.

References

Fama, Eugene F. 1970. Efficient capital markets: A review of theory and empirical work. *Journal of Finance* 25(2):383–417.

Figlewski, Stephen. 1979. Subjective information and market efficiency in a betting market. *Journal of Political Economy* 87(1):75–88.

Jones, Peter, David Turner, David Hillier, and Daphne Comfort. 2006. New business models and the regulatory state: A retail case study of betting exchanges. *Innovative Marketing* 2(3):112–119.

Kahneman, Daniel, and Amos Tversky. 1979. Prospect theory: An analysis of decisions under risk. *Econometrica* 47(2):263–292.

Klemperer, Paul. 1999. Auction theory: A guide to the Literature. *Journal of Economic Surveys* 13(3):227–286.

——. 2004. *Auctions: Theory and practice.* Economics Papers 2004-W09. Economics Group, Nuffield College, University of Oxford.

Koning, Ruud H., and Bart van Velzen. 2009. Betting exchanges: The future of sports betting? *International Journal of Sport Finance* 4(1):42–62.

Law, David, and David A. Peel. 2002. Insider trading, herding behaviour and market plungers in the British horse race betting market. *Economica* 69(274):327–338.

Shefrin, Hersh, and Meir Statman. 1994. Behavioral capital asset pricing theory. *Journal of Financial and Quantitative Analysis* 29(3):323–349.

Smith, Michael A. 2003. The impact of tipster information on bookmakers' prices in UK Horse-Race Markets. In *The economics of gambling*, edited by Leighton Vaughan Williams. London: Routledge, 67–79.

Smith, Michael A., and Leighton Vaughan Williams. 2008. Betting exchanges: A technological revolution in sports betting. In *Handbook of sports and lottery markets*, edited by William T. Ziemba. Amsterdam: Elsevier, 403–418.

——. 2010. Forecasting horse race outcomes: New evidence on odds bias in UK betting markets. *International Journal of Forecasting* 26(3):543–550.

Smith, Michael A., David Paton, and Leighton Vaughan Williams. 2006. Market efficiency in person-to-person betting. *Economica* 73(292):673–689.

Vaughan Williams, Leighton. 2000. Can forecasters forecast successfully? Evidence from UK betting markets. *Journal of Forecasting* 19(6):505–513.

Weber, Martin, and Colin F. Camerer. 1998. The disposition effect in securities trading: An experimental analysis. *Journal of Economic Behavior & Organization* 33(2): 167–184.

SECTION V

MOTIVATION, BEHAVIOR, AND DECISION-MAKING IN BETTING MARKETS

This page intentionally left blank

CHAPTER 24

BETTING MOTIVATION AND BEHAVIOR

ALISTAIR BRUCE

1 Introduction

This chapter is concerned with exploring individuals' motivations for betting and other aspects of their behavior as bettors. These topics have generated a substantial literature over the past four decades. While the study of betting behavior has been dominated by the lens of economic analysis, it has also occupied the attention of psychologists, sociologists, and those interested in designing the legal and regulatory regimes within which betting takes place or those tasked with addressing and treating the negative effects of excessive exposure to betting from a medical or social care perspective.

A major contemporary factor behind the interest in and motivation for academic research on betting is the large increase in gambling activity in many countries (see, for example, Gambling Commission 2011). This finds expression via an increasingly wide diversity of betting media, forms of betting markets and forms of engagement with betting markets. This has in part been fueled by, and in part has fueled, changes in the regulatory and legal frameworks within which betting takes place. In the United Kingdom, for example, the deregulation of betting and other forms of gambling has been as much a feature of the past 30 years as the rise in betting as a participative leisure pursuit, though which drives which is neither clear nor, probably, straightforward. For example, 30 years ago, the typical U.K. betting office had sound-only access to betting shows and race commentary, an overwhelmingly male-dominated and working-class clientele, a highly limited range of betting media focused on horse, and to a lesser degree greyhound, racing, no seating or refreshments. The interiors of betting offices were, by regulatory design, invisible from the street. Betting was regarded as an activity at the margins of social acceptability—indeed betting offices had only been legalized in 1960. The contrast with the contemporary betting office is striking. Premises today enjoy sophisticated technological support, typically with multiple TV screens relaying a diversity of information, live broadcasts of betting-related events, expert analysis, and guidance. The modern clientele reflects, to a far greater degree, a cross section of society in general. We can see into betting shops, which offer a bright, clean, and welcoming image, inviting clients to participate in what is now seen as a mainstream leisure experience, with the opportunity to sit down and enjoy refreshments as they bet. The menu of betting opportunities has increased significantly to embrace most mainstream sports as well as much wider media, such as outcomes in political elections, game shows, and financial markets. More strikingly, perhaps, the mode of engagement of bettors with betting markets is in a state of flux, with the rapid emergence of betting via Internet and mobile devices, while new forms of betting market, most notably person-to-person betting exchanges, are also a feature of a rapidly changing landscape.

The above is just one illustration of the transformative change, in the United Kingdom and elsewhere, that forms the context for this survey. The approach adopted here in addressing motivational and behavioral issues relating to betting is naturally shaped by the author's disciplinary background and experience as a researcher in the field over the past 25 years and, as such, reflects a personal perspective on the issues under scrutiny. Accordingly, the emphasis in this chapter is on examining betting motivation and behavior principally, though not exclusively, through

(i) the application of *economic analysis* to betting;
(ii) insights based on *naturalistic* rather than *experimental* research;
(iii) the lens of *horse race* betting;
(iv) the author's (and his collaborators') own work and insights.

Clearly the above basis for addressing motivational and behavioral issues cannot claim to offer a comprehensive coverage of available insights; it does, though, attempt to provide a consistent and manageable approach to a complex and diverse field.

The chapter is organized as follows. Section 2 explains the approach outlined above by examining economists' particular interest in betting as a research medium and the susceptibility of betting to economic analysis. This is followed by a discussion of the merits of naturalistic vis-à-vis other methodological traditions in the study of betting behavior. Finally, the focus on horse race betting as the principal medium of interest addressed in the chapter is explained.

This provides the basis in Section 3 for considering how individuals might be motivated to engage in betting as an activity and to become active participants in betting markets, as consumers. A variety of core motivations are considered, including some that have been the subject of empirical inquiry. An important consideration here is the belief that we can often infer something of the motivation of the individual bettor by examining the particular circumstances of that bettor's betting activity, such as how, when, or where the bet was made.

Section 4 moves beyond betting motivation per se to explore some of the influences on and relationships between further aspects of betting behavior. This discussion is centered on the concepts of risk-taking and betting performance and the factors that appear to influence each of these aspects. It will become clear that many of the insights into behavior that are developed and discussed herein are the by-products of studies, the principal objectives of which lie elsewhere. So, for example, while a large proportion of the empirical analysis of betting has occupied itself with issues of betting market efficiency, variously defined, the roots of inefficiency at the level of the market lie in individual bettors' and/or bookmakers' behavior. As such, studies with these more macro-level objectives have the capacity to enrich greatly our understanding of micro-level behavior. Finally, in section 4, some observations are offered on that most widespread and researched behavioral phenomenon in betting, the favorite-longshot bias. Brief concluding remarks follow.

2 Studying Betting Motivation and Behavior: Economic Analysis, Naturalistic Inquiry and Horse Race Betting

As noted above, the approach adopted in this chapter reflects the author's particular "take" on the body of theoretical and empirical work in the area of betting motivation and behavior, in terms of its disciplinary orientation, methodological preference, and the dominance of a particular betting medium within the literature.

2.1 The Economic Analysis of Betting

The justification for adopting a largely economic approach to the study of betting reflects both the willingness of economists to engage with the phenomenon and the preeminence of economics as a disciplinary basis for investigation within the literature.

Economists' interest in betting has been motivated by a variety of factors. First there is the apparent violation of the received view of economic rationality that is embodied in betting as an activity and which explains, in particular, the interest of behavioral economists. There is also interest in the study of decision-making in an economic context and under uncertainty, how decisions are framed and made and how such issues as the complexity of decision problem are managed by individuals. As a form of decision context, betting is of particular interest to economists as bets generate unequivocal outcomes with clear financial implications, which facilitate measurement of decision performance under a range of conditions. Third, economists have an enduring interest in the functioning and efficiency characteristics of markets and how these may be affected by biases in market participants' behavior or by information asymmetries.

In this section the emphasis is on exploring how economists seek to make sense of the apparent irrationality of betting. In a sense, developing an understanding of this issue clarifies the origins of economists' interests both in the details and processes associated with decision-making and the consequences for market efficiency which reflect decision-making in aggregate.

2.1.1 *Betting and Economic Rationality*

The fact that betting is an activity in which individuals voluntarily engage, using their own money, and which in the long run will almost inevitably lead to financial loss, poses difficulties for economists who feel the need to explain all economic behavior as motivated by narrowly defined, seemingly rational economic considerations. To such "purist" economists, therefore, betting is regarded as constituting eccentric behavior that defies rational explanation. The irrefutable fact that betting is a mass-participation and increasingly popular pursuit with material financial consequences, however, means that it over the past twenty years in particular, the behavioral economics tradition has become an increasingly important and vibrant sector of the discipline. In many areas of economic life an inconvenient truth is that human behavior appears to accommodate motivations other than the maximization of a narrowly defined welfare function. In the particular context of the analysis of decision-making, Richard Thaler (1991, xii) captured the limitations of the "purist" economic perspective.

> What I kept noticing was that people did not seem to behave the way they were supposed to .... I was constantly confronted with the contrast between the models my colleagues were constructing and the behaviour I was so frequently observing.

Thaler went on to describe his discovery of the work of Daniel Kahneman, Amos Tversky, and others (see, for example, Kahnemann and Tversky 1979; Gilovich, Griffin, and Kahnemann 2002, xxi) on judgment, heuristics, and biases, which he saw as key to developing a more inclusive perspective on decision-making, one that embraced concepts from economics and psychology. His conclusions on what he has termed "quasi rational behavior" are worth repeating.

> Quasi rational behavior exists, and it matters. In some well-defined situations, people make decisions that are systematically and substantively different from those predicted by the standard economic model. Quasi rational behavior can be observed under careful laboratory controls and in natural economic settings such as the stock market. Market economies and their institutions are different from the way they would be if everyone were completely rational.

Even in the relatively short period since Thaler made these observations, the dialogue and development of trust and mutual respect between the disciplines of psychology and economics has been noticeable, encouraged by a growing acceptance that each can enrich the understanding of a phenomenon of common interest by accommodating perspectives from the other. This increased readiness to remove the disciplinary blinkers is epitomized by the popularity of journals that span the boundaries of psychology and economics, most notably the *Journal of Economic Behavior and Organization.* Also, in the context of betting in particular, even the most casual engagement by an academic economist with the behavior of a betting office clientele or an on-course betting market would likely encourage the academic to seek to rationalize betting activity in terms of far more complex and shifting motivations than merely the desire to make money. This translates into an interest in motivation, as characterized by the bettor's utility function and, in particular, elements of utility that work against financial returns maximization in financial decision-making. Economists have shown considerably greater tolerance than has the academic finance community for the possibility of broader utility functions, and consequently there is greater acceptance by economists that deviations from market efficiency can be rationalized in terms of noneconomic or nonfinancial drivers of behavior.

The judgmental heuristics and resultant biases to which Thaler refers form an important element in the plethora of empirical studies that focus on the efficiency characteristics of betting markets. As noted above, a pervasive, and some might say obsessive, feature of work in this area has been the investigation of the phenomenon of the favorite-longshot bias as a contributory factor in market inefficiency. Favorite-longshot bias describes the tendency of favorites (shorter odds horses) to be underbet and long shots (longer odds horses) to be overbet relative to their objective probability of success. The causes of this particular bias are an issue of some debate within the literature, and while psychologically rooted factors are held to be important in some quarters, other contributors view the bias more as simply the function of a rational economic response to the threat of adverse selection under conditions of asymmetric information (see Shin 1991, 1992, 1993). The issue of the favorite-longshot bias is discussed in the concluding section.

2.2 Naturalistic and Laboratory-Based Inquiry in Betting Research

Thaler points to the incidence of, and opportunity to observe, quasi rational behavior in both controlled laboratory and natural contexts, and at this point it is appropriate to explore some of the methodological issues relating to the economic analysis of betting. One view is that, in essence, laboratory-based and naturalistic insights are complementary and mutually corroborative methods. As Bernard Baars (1980) has observed, "without naturalistic facts, experimental work may become narrow and blind . . . without experimental research, the naturalistic approach runs the danger of being shallow and uncertain." Although the literature on betting behavior is indeed well represented by both laboratory-based and naturalistic analysis, it seems fair to suggest that, in contrast to the inclusive tone of the view expressed above, there remains some implicit skepticism from each tradition regarding the contribution of the other.

From a personal perspective, implicit in a preference for naturalistic investigation is a failure to be entirely convinced by the ability of the laboratory setting to capture the full richness of the natural decision environment or to replicate the incentive-penalty conditions experienced by the bettor in a real setting. The susceptibility of individual decisions to small changes in aspects of the decision environment has been observed (see, for example, Payne 1982). Equally there is some skepticism as to whether observed behavior, in the laboratory, of relatively small groups of incentivized subjects who are unfamiliar with the decision task assigned sheds much light on our understanding of real bettors committing their own resources in a decision context with which they are more or less familiar.

A preferred method of investigation with colleagues has been the interrogation of large datasets of actual betting decisions, generally made available via data-access agreements with large bookmaking organizations. This preference may be a function of an economics disciplinary base to some degree, but from a methodological perspective there is a strong appeal in being able to observe and analyze decisions that are made

- in a natural setting
- by decision-makers who are to a greater or lesser degree task-familiar,
- by decision-makers oblivious to scrutiny
- committing their own money
- with attendant real financial gains/losses
- in a large and diverse set of individual betting markets
- with attendant opportunity for multifaceted cross-market comparison
- with the opportunity to input complementary market information.

Working with datasets that embody the above characteristics permits an unparalleled range of avenues of inquiry. Typically, for each bet made, we have self-documented details of the selection made as well as the stake and the type of wager (e.g., win or each-way) supplemented by data on the evolution of odds in the market, location (e.g., on-course, off-course, telephone), and precise date and time of bet placement; race-related data, such as racecourse location, number of runners, race type (handicap/non-handicap, etc.), race class (grade of race), position of race in the race card, ground conditions ("going"), race result with associated starting prices (SPs), and returns to each betting decision. These data offer rich potential for developing insights into, inter alia, betting motivation, levels of participation, risk appetite (in terms of size of investment, inherent riskiness of bet type or selection made), the effect of complexity (in terms of race type, number of runners), and relative performance, variously measured.

While the appeal of working with this type of resource is compelling, there are associated limitations. One is the anonymity of subjects in a natural setting. This has the advantage of ensuring the absence of distortive observation effects on behavior, but it denies the opportunity to interrogate subjects directly as to their motivations for gambling. Here, by contrast, laboratory studies can shed light on the factors and processes underpinning decisions made by interrogating subjects ex post. Equally, laboratory simulations can sharpen the focus on the relationship between particular variables in a controlled environment without the noise and turbulence which attend the natural setting. To the extent that the noise and turbulence materially influence the processes involved in decision-making, however, isolation of particular variables might be regarded as an abstraction from reality.

Ultimately there are advantages and drawbacks associated with each method, and a consensus around the complementarity of each is probably fair.

The above discussion, with its focus on the contrast between laboratory-based and naturalistic inquiry, reflects a further methodological bias, that is, a preference for quantitative versus qualitative analysis of betting behavior. This requires some comment. In large part the preference for naturalistic over laboratory-based inquiry turns on the appeal of data elicited anonymously in a natural decision setting versus data generated in a synthetic decision environment as a basis for seeking to explain naturally occurring aspects of decision behavior. Questions regarding the reliability of the quantitative interrogation of laboratory-generated data are amplified where qualitative analysis is concerned. Necessarily, whether via instruments such as questionnaire-based inquiry or interviews, eliciting qualitative information from bettors on aspects of their betting behavior requires conscious self-reflection by subjects. This may, particularly for an activity which has only recently moved from the margins of social acceptability, invite responses that reflect subjects' perceptions of socially acceptable norms in relation to betting activity or unwarranted and unrealistic self-perceptions rather than frank and honest reporting of motivation or other behavioral aspects. Equally, as with laboratory-based quantitative inquiry, there are potentially distortive effects associated with the investigator's framing of the qualitative instrument. These types of reservations may explain, in part, the dearth of qualitative studies of bettor behavior in the empirical literature.

2.3 Horse Race Betting as a Focus

The justification for using, in this chapter, horse race betting as the dominant form of betting medium through which to offer more general insights rests on a number of factors.

First, in many countries where betting is a significant pastime, such as the United Kingdom, horse race betting remains the single most popular betting medium. Equally, for most of the largest bookmaking organizations, horse race betting remains the single most important element in their core business. That said, a striking feature has been the rapid innovation in both new betting media (i.e., types of events on which bets can be placed) and new forms of betting market (such as betting exchanges and digital market media). So, in the context of these changes, can the focus on horse race betting markets be justified? It is argued here that it can be justified in the sense that a significant element of the betting clientele that engages with new betting media is also engaged with (and likely began betting on) horse race betting markets. As such the populations are far from discrete, and we might therefore expect their underlying motivations and behaviors to be similar.

Second, although its origins were in traditional bookmaker-based markets, horse racing is a dominant betting medium in new market forms. For example, Betfair, the leading person-to-person (betting exchange) operator in the United Kingdom and an innovator in sports betting, returned over 43 percent of its core revenue from horse race betting in financial year 2010/2011.

A third reason for adopting this focus reflects the fact that the majority of academic studies of betting have, over the years, used horse race betting data in their analysis. This reflects the richness and accessibility of continually expanding data on individual markets complemented by the availability of comprehensive and detailed market-relevant information. As such, in reflecting on and drawing insights from this literature, horse race betting is the naturally dominant milieu.

3 Economic Analysis of the Motivations for Betting

This section addresses the issue of betting motivation based on economic analysis. It is important at the outset to acknowledge that betting motivation is itself a complex and multifaceted issue. For example, bettors are characterized both by those who are solely interested in a recreational experience for a trivial outlay/return and those who regard themselves as professionals, placing high stakes. There are highly experienced and sophisticated participants as well as complete novices; extremely well-informed bettors and wholly uninformed and whimsical decision-makers. There are on-course, off-course, online, and telephone bettors; bettors who prefer bookmaker markets, pari-mutuel markets, or betting exchanges; male and female bettors and many more bases for disaggregating bettors as a population. These spectra of diversity are matched by a breadth of motivational drives for participation, and no doubt for many bettors the motivation comprises a blend of factors. It seems reasonable to suggest that the aggregate population of participants in betting markets is considerably more diverse and motivationally complex than the populations associated with most financial markets, where a more uniform and financially focused motivation would generally appear appropriate.

As a route into this complex world, we begin this section by discussing a paper by Alistair Bruce and Johnnie Johnson (1992) that represented an early attempt to validate a four-way classification of off-course bettors' motivations. Bruce and Johnson examined the proposition that different dominant motivations to bet might be associated with betting in distinct time windows of the betting day and with whether the bettor elected to "take" the price (odds) on offer at the time of bet placement rather than have a winning bet settled at starting price (SP) (the independently adjudicated market price at the close of the market). Four motivations were identified: financial gain (drawing, for example, on Cornish 1978), intellectual challenge (Downes et al. 1976; Langer 1983; Letarte, Ladouceur, and Mayrand 1986), social interaction (Filby and Harvey 1989; Saunders and Turner 1987), and excitement (Gilovich and Douglas 1986, Anderson and Brown 1984). It was argued that each motivation could be associated with a particular time frame within the betting day or a combination of time frame and price-taking behavior. So, for example, it was suggested that those bettors principally motivated by social interaction would be more likely to bet at times when the betting office was relatively heavily populated, whereas those for whom financial return was most important might be expected to operate immediately prior to a betting event, when the fullest exposure of market-relevant information is available. Comparisons were then made between the motivationally distinct subsets, and these generated results that appeared to support the basis for subgroup definition. Accordingly, the financial gain subgroup was associated with both higher levels of staking and higher levels of profitability. There was also a tendency for this group to bias betting more toward favorites, suggesting their more sophisticated understanding of the relationship between subjective (implicit in odds) and objective (revealed by outcomes) probabilities or, in other words, a sharper appreciation of "value" opportunities.

The identification of intellectual challenge as a motivation for betting prompted further investigation of the effect of different levels of complexity in betting event on participation. In other words, to what extent are bettors either incentivized or deterred by the challenge of more complex events? Most of the literature relating to this issue would suggest that complexity should inhibit participation. Decision problems involving more alternatives and/or attributes increase the cognitive and computational stress on the decision-maker. The power to discriminate between alternatives is compromised. Johnson and Bruce (1997a) later found, however, that alternative-defined complexity is positively correlated with participation in terms both of bets placed and stakes wagered. That is to say, bettors appear to have an appetite for more complex events featuring larger numbers of runners, and this would tend to support the intellectual challenge motivation. The results relating to attribute-defined complexity are more in line with the consensus in the literature, suggesting that complexity in the attributes set deters participation. The effects of complexity on other aspects of betting, such as risk-taking and performance, are considered in a later section.

To the extent that complexity may be viewed, alternatively, as a negative and unattractive or a positive and attractive feature, that is, as a deterrent or incentive to engage with a complex betting challenge, risk might be viewed similarly. Depending on an individual's attitude to and appetite for risk, betting generally may be viewed as an unattractive or attractive pursuit. Given that risk is inherent in betting, it seems reasonable to argue that it might form part of the appeal, alongside other factors, to individuals who do bet. Section 4 illustrates how the acceptance and management of risk, or risk behavior, in betting might be influenced by a range of factors. In the meantime, it is worth making two points about the nature of risk as an incentive to bet, which are materially distinct from the satisfaction of an appetite for risk per se. First, betting offers an opportunity to experience risk in a controlled setting, with self-determined limits on downside consequences. This may be attractive to individuals who are uncomfortable with risk in a general sense. Second, exposing oneself to financial risk in relation to the outcome of an event, a horse race for example, may amplify the enjoyment of the event. So, by paralleling the experience of simply viewing the event with a financial stake in the event's outcome, and associating one's interests with a specific horse, the bettor adds material to emotional engagement. This so-called mimetic dimension to participation has been observed as important in a variety of contexts. Patrick Murphy, John Williams, and Eric Dunning (1990), for example, observed this type of effect in relation to the identification with a particular team for soccer spectators.

The diversity of the aggregate population of bettors in terms of a range of behavioral aspects has already been stressed and is a recurring theme in this chapter. For many bettors themselves, of course, there will be an awareness of this diversity and, in particular, an awareness of who their fellow bettors are. This in itself may form the basis for betting motivation. For example, for more expert, informed, analytical, and serious bettors, part of the motivation for betting may relate to their awareness that they are playing in a market alongside novice, uninformed, irrational, purely recreational bettors. Serious bettors will believe, and possibly with some justification, that they are likely to be better decision-makers than this group and that the presence of novices may actually improve the odds available to the serious bettor. The presence of so-called mug punters is, naturally, well understood by bookmakers and an important element in guaranteeing their return. Bookmakers may even deliberately manipulate odds to tempt "mugs" toward invariably fruitless investments. For example, a horse whose probability of success in the view of the bookmaker is 1 in 300, may be priced at 50–1 to give the novice punter the idea that its chances of success are not so remote as they actually are. At the same time, bookmakers may lengthen the odds of short-priced horses to induce betting revenue from novice bettors who tend not to be attracted by very short odds and are unlikely to be aware of the relative "value" in the short odds range. More serious, higher stakes bettors, with a more sophisticated understanding of the objective probability of short odds horses, may thereby be able to benefit from better prices in the short odds range. In passing, it is worth noting that this sort of activity, of course, contributes to the favorite-longshot bias. And even without manipulative bookmaker behavior, the tendency toward a random distribution of betting across opportunities by less informed or less sophisticated bettors would be inclined to contribute to the favorite-longshot effect.

If one basis for identifying motivationally distinct subgroups relates to the novice/expert axis, another may be the form of betting market in which they choose to participate. In a U.K. context, the traditional distinction between bookmaker-based and pari-mutuel markets has in recent years been augmented by the development of new market media, most notably betting exchanges. It could be argued that each form of market satisfies different motivations.

Bookmaker-based markets may appeal more to those more traditional bettors who appreciate the opportunity to know with certainty the return to a successful bet at the time of bet placement. Pari-mutuel markets, by contrast, may tend to be more attractive to more risk-loving bettors, who are content to accept the price dictated by the market, or by those who prefer betting on outsiders, where the return to a winning pari-mutuel bet is invariably higher, and sometimes significantly so, than in the equivalent bookmaker market. Exchange betting arguably taps into the motivations of a wholly different set of bettors, those who enjoy the buzz of a direct screen-based interface with a live market. In many senses the interactive opportunities offered by exchange betting may be regarded as more akin to online poker in their more dynamic and interactive nature. They are also likely to appeal to those who appreciate rapid access to the range of information relevant to the market, given the linkages and real-time updates which are a feature of most betting exchanges.

Another basis for discriminating between groups is the on-course/off-course distinction. In general terms it is tempting to suggest that those who incur the time and financial costs of attending a racecourse and who bet during their time there are likely to be materially different from those whose betting is conducted via an off-course venue, such as a betting office, telephone account, or online. A reasonable assumption is that on-course bettors may be attracted to the course principally by their enthusiasm for the sport of racing, whereas betting for them may be an essentially secondary activity, a function of their access to betting opportunities on-course rather than an innate interest in betting. By contrast, it seems fair to suggest that those who choose to bet off-course are more likely to be principally motivated by betting itself rather than the underlying sport. Bruce et al. (2009) explored behavioral differences between on- and off-course betting populations, and the results reveal evidence for stronger favorite-longshot bias in the on-course population. This is seen as symptomatic of a population that has a less sophisticated understanding of the relationship between objective probability and odds-implicit probability, which in turn suggests material distinctions in betting motivation between on- and off-course bettors.

In relation to betting motivation in general, it is clear from the above that there is a wide diversity of factors driving betting participation and, therefore, a high level of motivational heterogeneity in the aggregate betting population. It seems reasonable to suggest that a major influence underpinning this diversity has been the deregulation of betting in many jurisdictions. As more liberal regulatory attitudes toward betting and gambling more generally have developed, this has encouraged betting and gambling operators to invest in the development of new forms of betting markets and new media for betting beyond the traditional horse and greyhound racing focus. It is tempting to argue that these innovations have encouraged previously untapped sections of the population into betting and gambling. In addition, in a U.K. context, though doubtless this also applies elsewhere, the introduction of the National Lottery as a gambling opportunity made available to all on a regular and highly publicized basis appears to have been an important factor in legitimizing gambling and betting as mainstream leisure pursuits and challenging an entrenched stigma in British society. In other words, many whose urge to gamble has been suppressed in the past have felt increasingly able to express their desire through active participation.

4 Betting Behavior

The range of observable and measurable behaviors associated with betting is considerable. Apart from shedding light on the motivation to engage in betting, as discussed above, aspects of the betting decision can offer insights into perceptions of and responses to risk, variously defined, attitudes to insurance, approaches to more and less complex decision scenarios, subjective perceptions of probability, confidence, and decision performance. In this section, initially, attention is focused on two areas of behavior—risk strategies adopted in the face of complex and uncertain decision problems, and decision performance. This focus is justified in terms of the fundamental role of risk in the activity of betting and the key role of performance in evaluating decision outcomes. In each case alternative measures of the dependent variables, risk and performance, are considered and how each is influenced by a variety of independent variables is reported.

It will become apparent during this section that there are often significant correlations between particular betting motivations, as discussed above, and attitudes to risk and performance. These associations are clearest where motivationally distinct subgroups, whether defined in terms of time of bet placement, location of bet placement, or some other factor, are characterized by distinct attitudes toward/response to risk or patterns of performance.

4.1 Risk

4.1.1 *Measuring Risk in Betting*

In considering how bettors respond to the presence of risk, it must first be acknowledged that the voluntary participation of individuals in betting markets suggests some degree of inherent risk appetite. The variety of betting options available spans a very wide range of risk exposures. It is this vast range which affords the opportunity to explore attitudes toward and responses to different degrees of risk.

There are a number of perspectives on risk that translate into operational measures of riskiness or risk-taking. We can focus, for example on the *type of bet* selected and differentiate between higher- and lower-risk bets. Thus a "multiple" win bet, where a return is dependent on predicting simultaneously and in combination the unique outcome of a number of discrete events, is at one end of the risk spectrum. By contrast, an "each-way" bet is a relatively low-risk investment, where the stake is split between the probability of a horse winning and it being placed (second; second or third; second, third, or fourth or occasionally, additionally, fifth place). This embeds an element of insurance into the bet to mitigate risk. A whole range of so-called exotic bets offers a wide menu of risk characteristics.

A second basis for measuring the riskiness of a bet is associated with the subjective probability inherent in the odds at which the bet is struck, where clearly the implied riskiness of a 16/1 selection (implied probability of success = <6%) is materially different from a 5/4 selection (implied probability = c. 44%). A more sophisticated version of this type of measure uses a version of odds adjusted by market context, that is, raw odds/mean market odds as a measure of relative risk. Related to this is the position of a selected horse in the betting market; that is, the market's view of the horse's ranked probability of success within the field, whereby comparative favorites are inherently less risky propositions.

A third basis for judging risk relates not to the selection per se but to the amount of money risked in the investment, the stake.

Although they are presented here as discrete insights into risk, it will be seen that there are often important interactions between these risk variables whereby, for example, increased risk exposure in one sense may appear to be offset by reduced exposure in terms of an alternative risk measure.

In investigating the influences on the risk behavior of bettors, a number of factors have come under consideration. Here the focus is on four main potential influences: time of bet placement (variously defined), location of bet placement, complexity of the betting decision, and gender of the bettor.

4.1.1.1 Risk and Bet Timing

Johnson and Bruce's (1992) empirical study of patterns of betting behavior investigated the relationship between the time of bet placement and various bet characteristics. For each of a sample of 1,200 bets placed in U.K. betting offices, the exact time of bet placement was recorded. This formed the basis for splitting the population into three subgroups. The first subgroup comprised all bets placed prior to the first report of the active betting market in the betting office (the "first show"). The second comprised all wagers placed during the active market period (typically around 15 minutes) minus all bets placed in the last 30 seconds before the start of the relevant race. The final subgroup comprised all bets placed later than 30 seconds before the "off" time, including any bets placed after the "off." The results revealed a number of notable distinctions between the periods. In terms of staking behavior, there was more evidence for significantly higher staking in the last time period than in either of the earlier periods, suggesting a greater willingness to incur risk in terms of the level of investment at risk by bettors in this group. This applied to bets placed at both starting price (SP) and board price (BP), the prevailing market price at the time of bet placement. Bets placed in the last period carried average stakes of £12.54 (BP) and £4.75 (SP) compared with equivalent values, for the earliest period, of £4.76 (BP) and £1.83 (SP) and, for the middle period, of £7.50 (BP) and £3.51 (SP), respectively. The results were interpreted as identifying a group of late bettors who were prepared to incur increased risk in terms of stake because they were confident that they had absorbed all of the price-embedded information generated throughout the life of the markets as a basis for their decision. The performance associated with this time-defined subgroup, reported below, supports the notion of an informationally privileged set of bettors.

It was observed earlier that time-defined subgroups of bettors could allow insights into subgroups that were also distinctive in terms of their dominant motivation. As part of this study, further aspects of risk exposure across time also were considered. These included the comparative propensity, across four time-defined subgroups this time, to bet on relative favorites and relative long shots. The results indicated a significant and positive difference between the proportion of bets placed on first and second favorites by bettors in the latest time period (post-30 seconds before the "off-time") at 60 percent compared with bets placed in three earlier periods. Equally the percentages of bets placed on horses outside the first three positions of favoritism and on horses with odds of 20/1 or more were significantly lower for bettors in the latest time period. The results of these two studies, taken together, suggest that risk exposure in terms of staking levels is inversely related to the market's assessment of risk.

Another insight into the relationship between risk-taking and time considers how the propensity to take risk may vary through the course of a betting day rather than in relation to a particular betting market. There is a strand of literature which suggests that risk attitudes may change as a result of the (normal) experience of loss during a betting day. One example of this is Gluck"s Second Law, which states that the best time to bet the favorite is in the last race of the day, the reasoning being that most impoverished bettors will incur greater risk by placing their money on relative "outsiders" in later races in order to recoup losses incurred earlier. As such the odds of the favorites in later races will be more favorable than would normally be the case. Johnson and Bruce (1993) investigated this phenomenon and explored various comparisons of subsets of earlier versus later races within a betting day. The results highlighted striking differences between early and late behavior. It was clear that betting on favorites, at 30.9 percent of bets in the last three races, was significantly higher than in the first three races (20.1%). This suggests greater risk aversion as later bettors default to the market view. Mean staking levels, as an alternative insight into risk-taking, however, tell a rather different story, with significantly higher mean staking in the last three races (£6.51) compared with the first three (£4.34). Again, taken together the results suggest offsetting behaviors in terms of the two perspectives on risk-taking, which endorses the importance of adopting a range of insights into, and measures of, risk behavior.

4.1.1.2 Risk and Bet Location

A further potential influence on the propensity of bettors to incur risk relates to the location from which they place their bets, where the distinction between the on-course and off-course betting populations appears to be significant. Bruce et al. (2009) explored differential incidence of favorite-longshot bias between location-defined subgroups of bettors, finding evidence for significantly greater concentration of betting toward relatively long-odds horses in the on-course market, compared with both the aggregate off-course market and each of its component submarkets (betting office, telephone, "away" track). Similarly, off-course bettors displayed a significantly greater propensity to bet on relative favorites than their on-course counterparts. While these results may, in some degree, reflect a greater risk appetite by on-course bettors, perhaps because of the dominance of recreational and less financially focused bettors within the on-course group, it may equally reflect the superior awareness by off-course, more serious bettors of the relationship between odds-implied and objective probabilities across the odds range.

4.1.1.3 Risk and Complexity

Finally, it might be argued that the type of betting event, which affects the complexity of the decision problem, may influence bettors' attitudes toward/appetite for risk. In the context of horse race betting markets, key factors in determining complexity are the number of runners in the race and the race's status as a handicap or non-handicap event. For all races, and their associated betting markets, complexity is seen as increasing with the number of runners (alternative-related complexity). Attribute-related complexity is greater in markets relating to handicap races, where the value of a key attribute (weight carried by the horse) is designed to compensate for other horse-specific, ability-related attributes in order to contrive a competitive race. Comparison of risk behavior across events of differing complexity suggests a range of effects. Using a probit model, Johnson and Bruce (1997a) noted an increased tendency to limit risk by betting on favorites as both alternative- and attribute-based complexity increase but with a more marked effect evident as alternatives increase. Johnson and Bruce (1997b) employed three further measures of risk-taking in investigating the determinants of staking, absolute odds, and relative odds (odds relative to mean market odds) as dependent variables, using both multivariate analysis of variance and separate univariate analyses. Distinctions between win and each-way betting were explored also. The results suggest that staking levels are not influenced by either alternative- or attribute-based complexity. In terms of raw odds as a measure of risk, there is no significant difference between handicaps and non-handicaps, but there is a propensity for odds to increase (i.e., greater risk) as the number of runners increases. Using relative odds, there appears to be greater risk propensity in handicaps than non-handicaps but little discernible effect on relative risk as number of runners varies. It is clear from this set of results that effects are highly sensitive to the particular measure of risk employed. There is a clear indication that the win/each-way decision forms an important element in the management of risk, particularly in relation to bettors who appear to be risk averse in terms of their choice of non-handicap (low attribute-based complexity) races. For this group, the each-way option appears to facilitate betting on longer odds horses. Perhaps the most important contribution of this paper, however, is to demonstrate the importance of interactive effects between alternative and attribute-defined complexity.

Most strikingly, the combination of high-attribute and high alternative–based complexity, handicap races with very large numbers of runners, appears to elicit very high levels of risk exposure, suggestive of a multiplicative relationship between complexity forms.

4.1.1.4 Risk and Gender

The potential impact of a bettor's gender on attitudes to risk and risk behavior is an issue of increasing importance, as more women gamble (any numbers on this trend?).

The contemporary literature on gender differences in risk appetite among decision-makers, suggesting a greater risk propensity among males (Keinan, Meir, and Gome-Nevirovsky 1984; Levin, Snyder, and Chapman (1988), motivated an investigation of gender-based differences in risk-taking in betting by Bruce and Johnson (1994). The results of this study were mixed. Males were significantly more likely than females to select "win" bets, rather than "each-way" bets, which embody an element of insurance. While there was no significant difference between genders in preference for more complex (and inherently riskier) accumulator bets, there was evidence that women preferred "any to come" accumulators, again suggesting a female preference for some element of risk management as compared with men.

A follow-up study (Bruce and Johnson 1996) focusing simply on "single" bets demonstrated a significant tendency for women to select inherently riskier longer odds selections as compared with men; 42 percent of men's bets were placed on horses with forecast odds of 5/1 or less compared with just 26 percent of women's bets. On the other hand, it appeared that women controlled their risk exposure in terms of odds by demonstrating a preference for each-way betting, with the proportions of female and male bets placed "each-way" being 39 and 25 percent, respectively. Perhaps the strongest message to emerge from these investigations into gender influences on risk is that perceptions and management of risk differ between men and women.

4.2 Performance

Just as the riskiness of a bet can be measured in various ways, so betting performance can be viewed from a number of perspectives. Equally there are a number of factors that have been identified by empirical inquiry as materially influential in determining performance. This section begins with a consideration of performance measure before addressing those factors which have been identified as significant in determining or affecting performance.

Very simply, in the context of a betting market, bets represent discrete decision events so that one calculation of decision performance relates straightforwardly to the rates of return realized by specific bets, individually or in aggregate. A related measure measures profitability per bet. Further measures include returns and profits per unit of stake wagered, which give a greater sense of the performance of bettors in aggregate or subgroups of bettors. There are then measures of performance that focus on the proportion of bets or stakes which are successful in generating a return or profit.

In terms of the factors that might potentially influence performance, location and timing of bet placement as well as the influence of complexity and excitement on performance and the issue of gender-based performance differentials are considered.

4.2.1 *Performance and Location of Bet Placement*

One of the benefits of working with betting data that segments bets according to where they are placed is the insight this can generate into comparative performance across locations. A recent study (Bruce, Johnson, and Peirson 2012) compared the performance of "attendee" bets (bets placed at the racecourse on horse races run at the venue) and bets placed remotely from the event venue (either off-course bets or bets placed in an on-course setting on-course but on events at another course). The results demonstrate a significant performance advantage for remote bets. The interpretation of these results centers on the likely distinctions between so-called professional and recreational bettors, which are held to be dominant influences on the remote and attendee populations, respectively. An interesting feature of this contribution's results is that they run contrary to an established empirical phenomenon observed in other financial market contexts, the home asset bias effect, which would tend to suggest performance advantages where investors and assets are closely co-located (see, for example, Coval and Moskowitz 2001). A potentially important element in the identified performance differentials may relate to the highly charged atmosphere of the on-course setting, which may militate against calm and rational decision-making, especially for recreational bettors.

4.2.2 *Performance and Timing of Bet Placement*

A range of time-related effects on performance has been identified in the literature. These include day of the week effects (e.g., Sung, Johnson and Peirson 2008), timing of bet within the structure of the betting/racing day, and timing of bet within the duration of individual betting markets.

In terms of performance effects associated with the timing of bet placement within the betting day, as noted above, there is a literature relating specifically to last-race effects (see, for example, Kopelman and Minkin 1991), where one contention is that there are material changes in bettor behavior in the last race of the day, when bettors will typically seek to recoup earlier losses. The comparative performance of bets in early and late races as evidenced by Johnson and Bruce (1993) is generally suggestive of a mild performance advantage for last-race bettors. Comparison of the proportion of winning bets in the last race of the day (18.5%) with that in the first three races (10.2%) yields a striking and statistically significant difference. These results may, in part, reflect the relative risk aversion associated with later race betting.

Within the more strictly time-defined confines of individual betting markets, an interesting question relates to the degree to which markets yield performance-relevant information through their evolution, as reflected in the changing patterns of available prices. To the extent that this occurs, one might anticipate superior performance by bets placed very late within the life of the market, where the exposure of event-relevant information is most complete. In terms of the proportions of stakes producing both a return and a profit, Johnson and Bruce (1992) found highly significant performance advantages for end of market bettors as compared with earlier market bettors. Most strikingly, more than half of the betting office stakes wagered at "board price" (odds guaranteed at time of bet placement) generated a net profit compared with percentage rates of profit in the teens earlier in the market.

4.2.3 *Performance and Complexity*

It has been observed that betting decision contexts which are characterized by high levels of complexity may influence both levels of participation and the particular strategies adopted in dealing with high levels of complexity. Unsurprisingly, perhaps, there is also some evidence to suggest that complexity affects levels of performance in betting. Bruce and Johnson (1996) employed two dimensions of complexity (alternative and attribute-defined) to discriminate between more and less complex betting events. The results demonstrate that performance in complex events defined in terms of alternatives (here, the number of runners in a horse race) is significantly compromised compared with that in less complex settings. For races featuring 12 or fewer runners, performance is significantly better than in races featuring more than 12 runners. This is a consistent result across both returns and profits to bets and stakes, and in terms of comparative average returns to stake, where the less and more complex sets of events yield figures of 0.92 and 0.52, respectively. Interestingly, using an attribute-based complexity measure, where complex (by design) handicap races are compared with non-handicap events, the distinction in performance effectively disappears, suggesting that bettors' decision performance is relatively invulnerable to attribute as compared with alternative-defined complexity.

4.2.4 *Performance and Excitement*

For many bettors it is fair to suggest that an important driver of participation is the excitement associated with holding a financial stake in an event of uncertain outcome with the prospect of significant financial gain. This raises the question of the degree to which bettor excitement may inhibit decision performance, in the sense that high levels of arousal may be expected to compromise the capacity of for rational and analytical consideration of the decision problem. One insight into this was offered by Bruce and Johnson (1995), who sought to measure the cost (or value to the bettor) of excitement. That paper reviewed the literature on the importance of excitement to leisure experiences generally and noted the role of betting in amplifying the core enjoyment of the leisure experience per se. A set of time intervals within the betting day, defined in terms of the intensity of associated excitement levels, was developed. The results suggested quite strong distinctions between bets placed in the relatively tranquil morning period, remote from the buildup to and running of the race, and bets placed in the turbulent and highly charged atmosphere associated with the second half of the live on-course market, immediately prior to the race. These distinctions were evident in terms of performance, with the high-excitement zone characterized by significantly poorer financial outcomes. As noted above, levels of excitement might be expected to vary with place as well as across time, and the results reported by Bruce, Johnson and Peirson (2012) in relation to location of bet placement are suggestive of performance penalties associated with high-excitement venues.

4.2.5 *Performance and Gender*

Gender differences in the form of betting, including preferences for particular types of bet, were explored above. Comparisons of the performance of male and female bettors also yield interesting results. The more recent general literature relating to relative decision performance of men and women suggests negligible differences as compared with earlier contributions (e.g., Priest and Hunsacker 1969), which identified the performance superiority of males.

In terms of betting decision performance more specifically, Bruce and Johnson (1994) reported highly significant performance superiority for females in terms of returns to both bets and stakes, though the advantage disappeared when profits to bets and stakes were considered. This study covered all forms of bet type and, as such, it was speculated that the performance results may have been in large part accounted for by female preferences for bets with an insurance component, which would also help explain the distinction between returns to bets and stakes, on the one hand, and profits to bets and stakes, on the other hand. A further study offered a "like-for—like" comparison by focusing on "single" bets only. Singles are the simplest form of bet, where bet performance relates to one selection in a race. Strikingly, all five performance measures demonstrated a significant superiority of female performance.

Reflecting on the influences on betting performance in general, it is clear that the determinants of performance are multifaceted, not always straightforward and at times counterintuitive. The above offers just a glimpse of the types of factor that may come into play, variously specific to person, form of betting event, place, time or environmental aspects of the decision setting. Clearly there are often strong interactions between these types of variables so that the tasks of disentangling and identifying the fundamental influences on betting performance remain a challenge for researchers in the field.

4.3 The Favorite-Longshot Bias as a Behavioral Phenomenon

While much of the above discussion relates to aspects of individual betting behavior, it would be difficult to end this contribution without a little further discussion on one of the most pervasive and consistent phenomena in terms of aggregate behavior in betting markets—the favorite-longshot bias. At the same time, such is the coverage of this topic that it is difficult to offer insights into favorite-longshot bias and its causes that say anything new. This section is confined, therefore, to some observations relating to the origin of the bias. Specifically, it reports an alternative supply-side perspective on the causes of bias, which differs in material respects from the received version, followed by the introduction of the possibility that our understanding of the origins of the bias may be enriched by incorporation of demand-side factors, which are capable of investigation by studying the comparative behaviors of motivationally distinct subpopulations of bettors.

By far the most influential contributor to the causes of the bias is Hyun Song Shin (1991, 1992, 1993), whose focus on the possibility of insider activity in betting markets, to which bookmakers are vulnerable, sees the particular pattern of (mis)alignment between objective and subjective (implicit in odds) probabilities as arising from bookmakers' pricing behavior in the face of this vulnerability, which varies with the number of runners. Shin's contribution in this area remains dominant but continues to invite important questions.

First, to the extent that the bias is regarded as simply a supply-side phenomenon, does the supply-side influence relate exclusively to the adverse selection issue and the factors which affect vulnerability to adverse selection? Second, and more directly relevant to any discussion of betting motivation, is it reasonable to suggest that demand-side factors have no influence on the bias? An examination of each of these questions follows.

In addressing the nature of the supply-side influence on the favorite-longshot bias, while there is a fairly strong body of evidence to support the Shin view, to explain the bias wholly in terms of bookmaker pricing being influenced by fear of adverse selection seems too simple. Bookmakers have a variety of mechanisms independent of the prices they offer to limit their risk exposure. They can, for example, refuse to take bets, they can limit payouts, they can share risk with other bookmakers by "laying off bets" where liability on one horse becomes uncomfortable. And as a professional community, it is difficult to view bookmakers as a vulnerable group that is systematically at risk of exploitation by a diverse, differentially informed, and overwhelmingly amateur clientele. Equally, there are other aspects of bookmaker motivation, beyond fear of exposure to insiders, which might help to explain the observed bias in the set of prices. So, for example, long shots may be shortened in order to convey to recreational bettors that the chances of the horses in question are not as remote as an "honest" assessment of odds might imply. Similarly, odds relating to relative favorites may be lengthened to tempt more "serious" bettors to recognize the value in this part of the market. This may be particularly relevant where bookmakers seek to attract the high revenues associated with this subgroup of the betting population. The issue of collusive bookmaker behavior and its effects was the subject of a recent investigation by Bruce and David Marginson (2012) in the particular context of explaining the magnitude of the overround in betting markets.

The overround reflects the degree to which the sum of implied probabilities embodied in odds within a betting market exceeds one and offers a measure of bookmaker margin. As a phenomenon it can be an influential factor in determining the degree of favorite-longshot bias.

The rationale for exploring potential demand-side influences on the bias resides simply in the fact that prices, or menus of prices, are normally regarded as resulting from the interaction of supply and demand factors and that odds or odds menus within betting markets are essentially no different. So, for example, the ability of bookmakers, as suggested above, to influence bettors' activity toward particular areas of the market says something about the way in which bettors perceive odds menus, and this in turn may cast light on bettors' motivations or utility functions. To the extent that bettors persist in biasing their activity, or allowing their activity to be biased, toward less fertile (in financial terms) areas of the odds distribution, this suggests that their utility functions are, indeed, populated by factors other than financial return. In this context it has been suggested that, for more recreational bettors, some utility may attach to the holding of a betting slip which carries at least the theoretical possibility of a very substantial return on investment even if the probability implicit in the odds is both very small in absolute terms *and* smaller than the objective probability. This may suggest, in turn, some appetite for increased risk independent of that which is regarded as the price paid for a probability of higher return. Bruce et al. (2009) offered insights into the potential demand-side influences on bias by examining bias across subsets of an aggregate pari-mutuel betting population. Their results suggest that motivational distinctions between subpopulations affect the degree of bias and that relative transactions costs and the nature of information and decision context influence the motivational composition of these subpopulations. As with the alternative perspective on supply-side influences discussed above, these insights into demand-side factors suggest that there are still aspects of the favorite-longshot bias which remain under researched and that there are significant limitations to a simple, adverse selection-base explanation of the phenomenon.

Conclusion

This chapter has provided a personal perspective on the related issues of betting motivation and performance. Adopting this approach necessarily involves a subjective emphasis, which reflects the author's particular enthusiasm regarding certain topics, themes, and methods. As such, this contribution makes no claim to offer an exhaustive coverage of all aspects of motivation and behavior. Instead it provides an insight into a rich, diverse, and growing field of inquiry. While researchers will doubtless continue to strive for a deeper understanding of established behavioral curiosities, such as the favorite-longshot bias, the continuing and rapid development of new forms of engagement with betting markets, new betting media, and demographic changes in the betting clientele introduce new questions that sustain and advance the research agenda in this field. This exciting broad research agenda is susceptible to inquiry from diverse disciplinary and methodological perspectives, where different insights are increasingly recognized as complementary and mutually enriching.

References

Anderson, G., and Brown, R. I. F. 1984. Real and laboratory gambling, sensation-seeking and arousal. *British Journal of Psychology* 75(3): 401–410.

Baars, Bernard J. 1980. On eliciting predictable speech errors in the laboratory. In *Errors in linguistic performance: Slips of the tongue, ear, pen and hand*, edited by Victoria A. Fromkin. New York: Academic Press, 307–318.

Bruce, Alistair C., and Johnnie E. V. Johnson. 1992. Toward an explanation of betting as a leisure pursuit. *Leisure Studies* 11(3):201–218.

——. 1994. "Male and female betting behaviour: New perspectives. *Journal of Gambling Studies* 10(2):183–198.

——. 1995. Costing excitement in leisure betting. *Leisure Studies* 14(1):48–63.

——. 1996. Gender-based differences in leisure behaviour: Performance, risk-taking and confidence in off-course betting. *Leisure Studies* 15(1):65–78.

Bruce, Alistair C., and David Marginson. 2012. Power, not fear: A collusion-based account of betting market inefficiency. Working paper; https://research.mbs.ac.uk/accounting-finance/Portals/0/docs/Power%20not%20fear.pdf.

Bruce, Alistair C., Johnnie E. V. Johnson, and John D. Peirson. 2012. Recreational versus professional bettors: Performance differences and efficiency implications. *Economics Letters* 114(2):172–174.

Bruce, Alistair C., Johnnie E. V. Johnson, John D. Peirson, and JiejunYu. 2009. An examination of the determinants of biased behaviour in a market for state contingent claims. *Economica* 76(302):282–303.

Bruce and Marginson has now been accepted for publication in the International Journal of the Economics of Business and should therefore be referenced as 2013, forthcoming.

Cornish, Derek Blaikie. 1978. *Gambling: A review of the literature and its implications for policy and research.* London: H.M. Stationery Office.

Coval, Joshua D., and Tobias J. Moskowitz. 2001. The geography of investment: Informed trading and asset prices. *Journal of Political Economy* 109(4): 811–841.

Downes, David M., B. P. Davies, M. E. David, and P. Stone. 1976. *Gambling, work and leisure: A study across three areas.* London: Routledge & Kegan Paul.

Filby, Mike, and Lee Harvey. 1989. "Recreational betting: Individual betting profiles. *Leisure Studies* 8(3):219–227.

Gambling Commission. 2011. *Industry Statistics 2009/10.* Birmingham: U.K. Gambling Commission; www.gamblingcommission.gov.uk/gh-media/latest_news/2010/industry_statistics_2009_2010.aspx.

Gilovich, Thomas, and Christine Douglas. 1986. Biased evaluations of randomly determined gambling outcomes. *Journal of Experimental Social Psychology* 22(3):228–241.

Gilovich, Thomas, Dale Griffin, and Daniel Kahneman. 2002. *Heuristics and biases: The psychology of intuitive judgment.* New York: Cambridge University Press.

Johnson, Johnnie E. V, and Alistair C. Bruce. 1992. Successful betting strategies: Evidence from the U.K. off-course betting market. In *Gambling and commercial gaming: Essays in business, economics, philosophy and science*, edited by William R. Eadington and Judy A. Cornelius. Reno: Institute for the Study of Gambling and Commercial Gaming, College of Business Administration, University of Nevada, 635–656.

———. 1993. Gluck's Second Law: An empirical investigation of horserace betting in early and late races. *Psychological Reports* 72(2):1251–1258.

———. 1997a. An empirical study of the impact of complexity on participation in horserace betting. *Journal of Gambling Studies* 13(2):159–172.

———. 1997b. A probit model for estimating the effect of complexity on risk taking. *Psychological Reports* 80(3): 763–772.

Kahneman, Daniel, and Amos Tversky. 1979. Prospect theory: An analysis of decision under risk. *Econometrica* 47(2):263–291.

Keinan, Giora, Elchanan Meir, and Tsafrira Gome-Nevirovsky. 1984. Measurement of risk takers' personality. *Psychological Reports* 55(1):163–167.

Kopelman, Richard E., and Betsy L. Minkin. 1991. Toward a psychology of pari-mutuel behaviour: Test of Gluck's laws", *Psychological Reports* 68(2):701–702.

Langer, Ellen J. 1983. *The psychology of control.* London: Sage.

Letarte, Andree, Robert Ladouceur, and Marie Mayrand. 1986. Primary and secondary illusory control and risk-taking in gambling. *Psychological Reports* 58(1):299–302.

Levin, Irwin P., Mary A. Snyder, and Daniel P. Chapman. 1988. The interaction of experiential and situational factors and gender in a simulated risky decision-making task. *Journal of Psychology* 122(2):173–181.

Murphy, Patrick J., John Michael Williams, and Eric G. Dunning. 1990. *Football on trial: Spectator violence and developments in the football world.* London: Routledge.

Payne, John W. 1982. Contingent decision behaviour. *Psychological Bulletin* 92(2):382–402.

Priest, Robert F., and Phillip L. Hunsaker. 1969. Compensating for a female disadvantage in problem solving. *Journal of Experimental Research in Personality*, 4(1):57–64.

Saunders, D. M., and D. E. Turner. 1987. Gambling and leisure: The case of racing. *Leisure Studies* 6(3):281–289.

Shin, Hyun Song. 1991. Optimal betting odds against insider traders. *Economic Journal* 101(408):1179–1185.

———. 1992. Prices of state contingent claims with insider traders, and the favourite-longshot bias. *Economic Journal* 102(411):426–435.

———. 1993. Measuring the incidence of insider trading in a market for state-contingent claims", *Economic Journal* 103(420):1141–1153.

Sung, Ming-Chien, Johnnie E. V. Johnson, and John Peirson 2008. Understanding the weekend effect: Insights from a market for state-contingent claims. Department of Management Working Paper, University of Southampton.

Thaler, Richard H. 1991. *Quasi Rational Economics.* New York: Russell Sage Foundation.

CHAPTER 25

MOTIVATION IN BETTING MARKETS: SPECULATION, CALCULUS, OR FUN?

LES COLEMAN

Betting markets are diverse but share several distinguishing traits.[1] Generally they are open for short periods and have an endpoint that relates to an event whose occurrence (but not outcome) is known in advance. The events are conducted in public with strong spectator interest, and bettors can expect to have or develop skill that gives them an advantage. Bets are typically written by corporations or well-funded bookmakers but are predominantly purchased by individuals. Due to substantial operating fees and taxes, betting is a less-than-zero-sum transaction, and there are limited opportunities to sell securities or to arbitrage positions. Betting markets, then, are very different from financial markets that trade long-lived securities whose value is set by events conducted in private and which have sophisticated participants on both buy and sell sides and parallel derivatives markets.

The unique structure of betting markets and the asymmetry between participants on supply and demand sides raise the issue of the motivations of market participants. I use *motivation* with the meaning of a force that is internal or external to the bettor that triggers, directs, intensifies, or leads to betting (Lee et al. 2007). My specific research questions relate to what the literature tells us about the motivations of bettors and their implications for betting markets.

Background

Betting markets trade securities whose values are contingent on the outcome of a specified event, typically the result of a contest between individual humans or animals, sporting teams, political parties, or the like. In most cases the securities are issued over the counter by a bookmaker or pari-mutuel operator, though organized betting exchanges, such as Betfair, are growing in scale. Payment for a successful bet can be agreed in advance (fixed odds), be derived from other bets (e.g., the most favorable starting price odds offered on the successful outcome), or be established by the proportion of successful bets (as in a pari-mutuel system).

The typical assumption in finance about the value of any security is that it equals the present value of its expected cash flows discounted at a risk-adjusted rate (Damodaran 2002). This is less relevant to betting securities that relate to a single, near-term cash flow whose value is known or can be reasonably estimated. Moreover, the betting cash flow is *contingent* and will occur only in the event of particular outcomes among a number of possibilities. Bets, then, have more in common with options than bonds or shares because options represent a contingent claim and because their payout, too, can fluctuate.

The typical approach to valuing bets is to multiply the expected payout from a successful bet by its probability of occurrence. Consider an event that a bettor judges has probability p of occurrence, which indicates a theoretical payout from a \$1 bet of $\$\left(1+\frac{1}{p}\right)$. If the bettor can obtain a higher payout the bet has a positive expected return and vice versa. In other words, rational bettors should place bets when their expected probability of success is higher than that indicated by the odds on offer (i.e., $\textit{Payout offerred} > 1+\frac{1}{p}$) and arbitrage should ensure that odds converge to reflect the objective probability of success of each bet.

In practice, though, there is extensive evidence that betting markets are inefficient in the Fama (1970) sense. In particular, betting prices do not move randomly, or quickly price in all publicly available information, and they do appear to offer opportunities to profit from private information. The many examples of irrationality or inefficiency make it axiomatic that a significant portion of bettors are not rational in the neoclassical sense that they maximize expected utility. This is not surprising given that individuals appear to be vulnerable to cognitive biases, which dominate the demand side of betting markets. This allows rational bet sellers or large bettors to exploit the misjudgments of individual bettors. There are several comprehensive reviews of anomalies in betting markets (see Hausch and Ziemba 1995; Sauer 1998). Thus I will only briefly summarize several of the most important betting market biases, as they will support my contention of inefficiency and form the basis for discussion of bettor motivations.

Perhaps the most consistent of all betting market anomalies is the *longshot bias* in which high payout (that is, long odds or longshot) bets are priced above their expected value and short odds bets are priced below their expected value. As a result the expected return from short odds bets is greater than that from longshot bets, and expected return rises with the probability of a win. This has been extensively documented in horse racing (Coleman 2004) and in a variety of team sports, such as cricket, and individual contests, such as boxing and tennis (Cain, Law, and Peel 2003).

A second near-universal bias relates to location of the contest. According to the *home team underdog bias*, teams competing at home that have long odds of winning are more successful than the market expects. There is also a *home country bias* in contests with international participants such that odds on success by a competitor of a given nationality are lower in that nation's markets than in foreign markets. This bias is sufficiently strong to justify formation of large international syndicates that arbitrage bets in popular international contests, such as the various world cups.

Semi-strong inefficiency is also said to exist in betting markets. The most obvious example is that the payments for winning bets by pari-mutuel operators consistently exceed those offered on identical runners by bookmakers (Gabriel and Marsden 1990). This is not a true anomaly, however, as it is totally expected: bookmaker odds are incremental because they apply only to current bets, whereas pari-mutuel odds are cumulative, reflecting all bets to date. As odds on winners typically fall during the betting period, at any point in the betting (incremental) bookmaker odds will be less than (cumulative) pari-mutuel odds. Thus bookmaker starting price (SP) odds will be less than the mean odds offered by the pari-mutuel operator.

Weak form inefficiency is evidenced by *herding*, where odds will trend in the same direction for some time. In pari-mutuel markets this occurs when a disproportionate volume of bets is placed on one outcome; in bookmaker markets it reflects the market maker's decision and so can be part of a strategy to induce herding.

There is also consistent evidence of strong form inefficiency in betting markets. This exists due to chronic asymmetry in information among bookmakers, bettors, and contestants and offers opportunities to profit from private information, including manipulation of the betting market and contest outcome.

Given the spread of economic analyses of betting market efficiency and biases it is not surprising that some methodologies do not prove robust. Michael Cain, David Law, and David Peel (2001), for instance, found that the Gabriel and Marsden anomaly does not hold under different estimation methods. Another example is the influential methodology developed by Hyun Song Shin (1993) to calculate the extent of insider trading in bookmaker markets. Shin, following the assumption that bookmakers manipulate the supply side of the market to protect themselves against the risks of adverse selection, derived a value, z, that represents the proportion of these bets. My own study (Coleman 2007b) showed that near identical values of z are obtained in bookmaker markets and pari-mutuel markets. As the latter do not have a supply side and so should have a zero value of z, the methodology's premise is questionable. Similarly a number of studies assume all bettors are risk averse, despite evidence that they embrace risk (with its meaning of loss and higher volatility) and skew (e.g., Golec and Tamarkin 1998). More broadly, but just as unlikely, are common assumptions that there is competition between bookmakers (e.g., Cain, Law, and Peel 2001) and that skillful analysis of publicly available information can lead to profitable betting (e.g., Lessmann, Sung, and Johnson 2009). Examples such as these suggest that betting studies are particularly susceptible to what Eugene Fama (1991, 1575) has termed "the joint hypothesis problem" whereby models must rely on a priori assumptions that can be mis-specified and draw invalid conclusions.

The intuition behind this chapter is that biases and inefficiencies in betting markets mean that bettors are not motivated solely by maximization of expected value. Thus my research objective is to analyze bettors' other possible motivations. This extends previous studies which have usually involved only problem gamblers with a view to mitigating the issue (see, for instance, Raylu and Oei 2002). These typically seek addictive or psychological explanations, such as arousal effects, counters to depression, and masochism. While these are important issues, most bettors are not addicts. Thus a secondary research objective is to establish a framework to explain the motivations of nonproblem bettors and hence contribute to a more complete depiction of behavior that can help distinguish between those with and those without a betting problem.

Considerable research has been conducted on the motivation of gamblers, principally using slot machines and roulette and in both laboratory and naturalistic settings. An open-ended study asked college students to list their top five reasons for gambling and found the answers to be predominantly rational: money (22.1% of all motivations), enjoyment/fun (18.4), social reasons (13.3), and excitement (9.8). Less than 12 percent of all reasons related to such negative or pathological motivations as boredom, escape, or drinking (Neighbours et al. 2002). Such studies support conventional, rational explanations for betting, such as the financial objective of winning, a preference for hedonic fun and enjoyment, and the desire for stimulation and socialization. Another study conducted a factor analysis of 51 possible motives for gambling and found that the most important were excitement (including thrill and tension), socialization, avoidance of negative feelings (troubles, loneliness, anger, anxiety), monetary gain, and amusement (Lee et al. 2007). Other research, though, is less kind to bettors. Karim Benhsain, Alain Taillefer, and Robert Ladouceur (2004, 399) begin their analysis with the claim that "the majority of individuals behave and think irrationally when gambling."

To summarize studies of the motivation of gamblers and bettors, Lee et al. (2006) found three broad explanations. The first involves social motives, such as participation in gambling groups or escape from personal problems. The second is psychological and perceives gambling as a self-determined pursuit of excitement or achievement, including monetary reward. The third and final base views gambling as an experience that is consumed in the same way as tourism and watching movies.

To complement published sociological and psychological studies of bettor motivations, the following sections discuss motivation from a financial perspective, particularly in terms of bettors' attitudes toward risk, expected return, and socialization. I close with a brief discussion of the consequences for betting markets of different bettor motivations.

Risk Attitudes in Betting

For the average bettor, betting has a negative expected return of up to 30 percent. This means that betting is highly risky, with *risk* taking on the meaning of accepting the possibility of loss. Thus, with the exception of the few bettors whose skill delivers significantly more than the average return, it is hard to sustain a rational economic motive for betting. This means that betting must provide some nonfinancial benefit.

A common explanation is in bettors' attitudes toward risk, particularly as betting offers a range of risk-return trade-offs. This is analogous to the concept in financial markets of speculation, which—though often ill-defined—usually refers to the purchase of securities or goods without an investment horizon, and without the expectation of income, by a party who gains from their use or from a desire to hedge another risk. A typical example is purchase of commodities or derivatives that do not generate income but fluctuate widely in price. In conventional markets speculation involves investing in risky (often leveraged) securities that have a high probability of negative return but have an uncertain possibility of a disproportionate rise in price. Speculators are generally thought to be risk prone and profit by accepting risks from risk-averse hedgers.

Parallels to financial markets speculation can be seen in the actions of bettors who buy option-like bets, including lottery-style longshot bets. The latter may reflect risk embrace, but they can also meet a preference for skewness, or very high possible payout, that is attractive to risk-neutral or even risk-averse bettors (Golec and Tamarkin 1998). This preference of risk-averse bettors for lottery-like payouts is consistent with evidence in other areas of life, such as investment (Kumar 2009).

This points to the existence of at least two discrete groups of bettors (Coleman 2004). The first is informed or skilled, predominantly places low odds bets, and has a positive (or near zero) expected return: this group is risk neutral. The second, and larger, group is less skilled and informed, places longer odds bets largely in accordance with chance, and has a significantly negative expected return: these tend to be gamblers and risk lovers. The transition from risk aversion (positive expected return) to risk embrace (negative expected return) occurs around an objective probability of a positive outcome of about 0.2. This transition to risk embrace (or the overweighting of long shots) matches the finding of Malcolm Preston and Philip Baratta (1948) that probabilities of less than 0.25 are subject to systematic overestimation and is consistent with findings on risk transitions by Colin Camerer (1995) and Amos Tversky and Craig Fox (1995).

The importance of risk in bettors' motivation poses an interesting challenge to theories of decision-making under risk or uncertainty because it violates what Tversky and Daniel Kahneman (1992) described as a key element: that in the initial framing phase of a decision "transparently dominated prospects are eliminated." Bettors' psychological makeup causes them to irrationally interpret available information.

A risk-averse strategy is to follow changes in odds because those that reflect informed trades move toward a closing price that more closely represents the actual outcome than do opening odds at the start of betting. This is true of horse racing (Coleman 2007a), basketball (Gandar et al. 1998), and other sports.

Bettor Expectations of Return

Extending the assumption developed earlier that rational bettors test the probability of a win against odds on offer, information-driven expectations of return are important motives for bettors and bet sellers. Irrespective of where the odds are set on contest outcomes, if they rationally reflect available information the bet is only attractive to bettors with different information that indicates a higher probability of winning.

There are a number of sources of superior bettor-specific information. It can be a proprietary skill that more accurately calibrates the probability of a given outcome, most obviously by using a superior process to interpret publicly available information. In addition, some bettors have access to monopoly information. My guess is that it would be hard to find any bettor who does not believe that superior use of information, particularly information that is not public, provides a source of incremental return to betting on contests.

Looking first at superior processing of public information, experts abound in betting. Until the explosion of Internet-based wagering, virtually every newspaper published a form guide or tip sheet on racing, football, and other popular betting contests. This has prompted many analyses of betting experts' skill, particularly that of racing tipsters, and the literature has been reviewed by John Peirson (2011) and Leighton Vaughan Williams (2000). The general conclusion is that tipsters as a group are able to match the accuracy and financial return of the wagering market in racing. Because tipsters' predictions are made without the benefit of seeing the strength of others' expectations (for instance, through wagering market odds and changes in these odds) and well ahead of the races' commencement, it is easy to agree with Fergus Bolger and George Wright (1994) that they have expertise.

It is not so easy to establish, though, whether or not this apparent expertise comes from skill, as the sources of any skill are opaque. In an effort to identify the nature of expertise or superior betting skill, Stephen Ceci and Jeffrey Liker (1986) recruited 30 men who were longtime patrons of a harness racetrack in Delaware, had extensive knowledge of the industry, and purchased form guides ahead of race day. By comparing the men's ability to select starting favorites before any betting market formed, they divided the group into 14 experts and 16 nonexperts who, respectively, picked at least 9 out of 10 favorites and less than 5 favorites. Means for the two groups were not different when they were divided according to years of education (about 10), occupational prestige, measured IQ (100), and years of experience in betting (15–17). After exploring interactions, the researchers concluded that intelligence and learning do not determine betting skill. Rather, it is related to use of a greater number of predictor variables, which is equivalent to operating a more complex cognitive model.

There are more formalized evaluations of publicly available information. Michael Kaplan (2002), for instance, provided an insight into the quantity and variety of data collected by sophisticated betting syndicates. Stefan Lessmann, Ming-Chien Sung, and Johnnie Johnson (2009) developed a two-stage betting model in which the first stage estimates the probability a horse will win by calculating its objective performance level or ability and the second involves within-race ranking of horses' relative ability.

Another source of superior valuations comes from the ability to manipulate betting markets, which was graphically illustrated by Camerer (1998). He hypothesized that some bettors follow price signals and thus erroneous price signals could mislead uninformed bettors and distort odds to the advantage of a market manipulator. The author conducted a field experiment using races which had two runners that matched according to morning line odds and odds set in early betting. About 20 minutes before the races started, he randomly chose one of the matched horses and placed a win bet that was canceled about 15 minutes later. The second, unbet horse served as a control, and odds on both horses were recorded. Betting $500 on 50 races showed no substantial effect, and the experiment was modified to comprise two separate $500 bets at tracks with smaller pools that were made later, at around 10 minutes before the start, and canceled just before the start. This showed subsequent reduction in the odds of the bet horse, which was more pronounced in maiden races where there was less information available about the runners.

Confirmation that this kind of market manipulation may be significant comes from the fact that pari-mutuel operators have changed their rules to limit the ability to cancel bets in New Zealand (Camerer 1988, 460) and Australia (Templeton 2003).

A second source of manipulation is to affect the outcome of a contest rather than market odds on the outcome. This can be to benefit a team, such as in a round-robin tournament where it may be advantageous to play poorly against a weak opponent and allow them to progress to the next round ahead of a stronger opponent. Another advantage to losing can come toward the end of a season when poorly ranked teams can have an incentive to drop in the rankings to secure more favorable conditions in picking next season's players.

Sports can be corrupted by individuals for their own purposes. In cycling or motor racing where individuals compete within a team, the finishing order can be agreed upon in advance irrespective of individual performances. A specific example is provided in a study by Mark Duggan and Steven Levitt (2002, 1595) of Japanese Sumo contests over a decade, which found "overwhelming evidence that match rigging occurs in the final days of sumo tournaments." Tournaments involve 15 bouts, and wrestlers who win at least eight bouts are promoted in the rankings, which provides them with significant additional rewards. Duggan and Levitt found that wrestlers with seven wins in a tournament are victorious more often than expected (and their opponents win a disproportionate share of the next bouts, indicating payback) and are less successful at the end of their careers. In addition, manipulation disappears when the media are focused on match rigging.

Contest outcomes can also be manipulated to secure a higher return from gambling on the result either by participants or by gamblers who use part of their winnings to induce participants to rig the outcome. A good example of gambling corruption involves U.S. basketball, which offers spread betting in which a bet on the shorter priced team pays if the team wins by more than the spread (say 8.5 or 16.5 points) and a bet on the longer priced team pays if the team wins or loses by less than the spread. The colloquial term for winning by less than the spread is point shaving, which can promote manipulation because the team wins the game and secures championship points but bookmakers do not pay out. An analysis by Justin Wolfers (2006) of 44,120 National Collegiate Athletic Association (NCAA) men's basketball results in the period 1989–2005 concluded that as many as six percent of strong teams manipulated their performance downward and thus that gambling-related corruption affected around one percent of games. Richard Borghesi and William Dare (2009, 121) provided a list of games that were found to have been manipulated and revisited Wolfers's data to conclude that "strong favourites . . . win as frequently as expected, but by a margin less than anticipated."

Wolfgang Maennig (2005), Ian Preston and Stefan Szymanski (2003), and Stefan Winter and Martin Kukuk (2008) have provided a good range of examples of betting-related corruption in sport. Manipulation is more likely when it does not affect the outcome of the game (e.g., such within-game betting as which player makes the first score), when opponents face asymmetric rewards from winning (as is the case in leagues with relegation or advancement of teams from one season to the next), or with spread betting when the game can be won but the bet lost.

While the media in many countries carry frequent stories of betting-related corruption, there are surprisingly few proven cases. Borghesi (2008), for instance, reported that it has been more than 50 years since there has been a documented case of corruption in the four major American professional sports of basketball, football, baseball, and hockey. At the other extreme is cricket, against which numerous instances have been documented, often alleged to be tied to bookmakers in India (Maennig 2005).[2] In the middle is horse racing where there is a lot of anecdotal evidence of corruption but relatively few proven cases (Coleman 2007a). Corruption seems more common in countries where betting is weakly regulated and in sports where regulations are not enforced.

A further source of superior valuations is monopoly, or insider, information that is not generally available to other bettors, and this has stimulated a rich literature of economic analyses that seek to identify the incidence of insider trading. A review of a variety of betting markets by Cain, Law, and Peel (2003) using a methodology developed by Shin (1993) estimated that insiders place between two and eight percent of bets, which is consistent with results using a different technique by Les Coleman (2007a).

Betting as Socialization

A commonly accepted nonfinancial motivation for betting arises in such factors as socialization, fun, excitement, and other personal or recreational needs. Socialization motives can be strengthened when betting is conducted around spectator sports—cricket and football, horse and dog racing—or in entertainment complexes, such as racinos. Supporting the importance of betting for fun and excitement is the fact that this is the emphasis of most advertising for betting. Another possible explanation for this tilt in advertising is that men are more likely than woman to place a bet, and women may be attracted by its context. A related motivation is achievement where bettors exploit a developable skill, such as counting cards in blackjack. In each case bets are simply consumption goods and purchased in much the same way as a movie ticket or language training.

Betting has another personal attraction, which is that participants have total control over their decisions. This appeals to those with a preference for self-determination and also feeds into the illusion of control, or the belief that bettors can foster a win through their efforts, even in the face of conflicting evidence. This was a popular subject of research during the 1960s and 1970s, and a number of exotic experiments were performed to replicate the finding that people have most confidence in events they can control. Howell (1971) tested students who threw darts at a board and were rewarded in proportion to their score multiplied by a number obtained at random from spinning a roulette wheel. This resulted in multiple paired outcomes determined by the combination of a factor within the students' control (dart score) and a factor beyond their control (roulette wheel's result). After a familiarization period, students played for money and—when able to choose the criteria for a win—consistently preferred outcomes where the greatest uncertainty related to the dart; in other words, they preferred to back their own skill rather than trust chance.

A consequence of extensive betting for nonfinancial motives is that these bets will be distributed according to a relatively narrow range of factors (perhaps the leading jockey or trainer in horse racing or a home team) or factors that are not commonly associated with winning (contestant's name or colors). This can induce an element of randomness in bets that—in popular contests, such as the Super Bowl, or glamour races—can swamp biases. That is, social bettors have different valuations of bet outcomes.

Consequences for Betting Markets of Heterogeneous Bettor Motivations

In closing, let me discuss the consequences of bettors' varying motivations for betting markets.

Perhaps the most important economic question is where the inefficiencies and biases in betting markets arise. Obviously in pari-mutuel markets that do not have a supply side, anomalies arise in the pricing decisions of bettors. But what is the relative contribution of the different bettor populations? And, of course, in bookmaker markets there is the added complexity of a sophisticated supply side: so where do biases arise here?

The most obvious point to make is that many of the biases in betting markets are amenable to multiple explanations. Consider as an example the longshot bias, which is due to relative overpricing of low-probability winners. One common explanation is risk, so that some bettors prefer the higher variance of long shots. Alternatively the higher skew of long shots—where wins bring bragging rights—can be attractive even to risk-averse bettors. The bias may be perceptual, where bettors overestimate the probability of low-frequency events because they are more memorable. In any case, bettor populations are segmented such that some bettors pay too much for low-probability winners and induce the longshot bias. Other explanations for the longshot bias that are consistent with heterogeneous bettors is that informed money bets on favorites with long shots attracting less informed money from irrational, less-informed bettors who suspect that public information is incomplete or erroneously discount the attractiveness of favorites.

Most analyses of betting market biases look to demand-side causes,[3] but an intuitively obvious source of bias in betting markets with a supply side (that is, bookmaker markets, casinos, and lotteries) is the asymmetry between participants on the supply and demand sides. Usually there are few suppliers of bets, and their prices are sufficiently similar to suggest collusion (an alternative explanation of low, competitive prices under perfect competition is inconsistent with reported profitability of casinos, lotteries, and bookmakers). They are, then, oligopolies, which give opportunities to extract rent through biases in pricing. Another advantage is that it is expensive to obtain private information (or manipulate results) for most contests that attract bets, and the higher turnover of bet suppliers makes it easier for them to pay the high fixed costs for monopoly information on contests. Thus, biases and inefficiencies could arise if bet suppliers use monopoly power to distort prices to their advantage and are better able than bettors to form more accurate expectations of contest outcomes. This is a totally rational outcome because the volume of transactions in betting markets makes it very attractive for the most skilled and knowledgeable bettors to operate on the supply side.[4]

A strong indication that biases arise on the supply side comes from the fact that bet prices in markets with a supply side experience little change. Many are fixed, particularly games of chance, such as lotteries, roulette, or slot machines, which have a constant payout. Operators of these fixed-odds games have designed them to be attractive to bettors, despite a negative expected return. Even in the case of betting markets prices change relatively infrequently through the course of betting (Levitt 2004). In the case of U.S. National Basketball Association games, for example, the point spread changes by less than half a point in more than half the games; given that the mean opening line spread is 4.8 points, this means that the spread is relatively constant (Gandar et al. 1998). Moreover, except at the extremes, the odds offered differ little between bookmakers and usually are similar to those in the pari-mutuel market. Thus casinos, lotteries, and bookmakers are adept at setting prices—and hence expected returns—at a level that is attractive to bettors, and fragmented participants on the demand side of betting have no control over the bets' structure and little influence on bet prices. In addition, bet suppliers can leverage monopoly information and skill to bias odds in their favor.

As an example, knowing that bettors prefer local teams, bet suppliers could generate abnormal return by offering high odds on foreign teams and low odds on local teams.

In short, market makers construct attractive and profitable bets by catering explicitly to the motivations and preferences of bettors and by besting bettors at forecasting contest outcomes.

The implications of this are profound for betting market researchers. The latter usually follow the assumption of Richard Thaler and William Ziemba (1988) that wagering markets are mirrors of conventional markets, efficient, and thus well suited to financial studies. The conclusions above are quite different, as they highlight the heterogeneity of bet buyers and the large asymmetries between buyers and sellers of bets that make betting markets totally different from conventional markets. Betting markets have a negative expected outcome for bettors, whereas conventional markets have a positive expected outcome over time for investors. Absent IPOs, conventional markets have similar participants on buy and sell sides, whereas betting markets are oligopolies. Information and skill are diffused through conventional markets but are concentrated on the supply side of betting markets. Regulation of conventional markets is generally strong, whereas it appears much less robust in betting markets. Thus, models of betting markets that assume competition, efficiency, and behaviors similar to those in conventional markets can be mis-specified and hence, reach erroneous conclusions.

Notes

1. I use the terms *betting market* and *bettor* to emphasize that most of my material and discussion relate to contests other than those involving games of chance (such as roulette, lotteries, and slots) where participants are better described as gamblers. The expected return from both bets on contests and gambles on games of chance is negative but skewed, so a small proportion of players can expect to make money. However, because gambles involve random events, only bettors can be rationally expected to have some influence over the outcome.
2. A series of links to materials cataloging corruption in international cricket are provided by the Australian Broadcasting Corporation at www.abc.net.au/4corners/content/2010/s3047207.htm.
3. A prominent exception is the Shin (1993) methodology for calculating the amount of insider trading that assumes it can be observed in bookmaker-induced biases to protect against insider bettors.
4. Why does this situation not occur in conventional markets where it appears that the most sophisticated investors—fund managers—are generally unable to beat the market?

References

Benhsain, Karim, Alain Taillefer, and Robert Ladouceur. 2004. Awareness of independence in events and erroneous perceptions while gambling. *Addictive Behaviors* 29(2):399–404.

Bolger, Fergus, and George Wright. 1994. Assessing the quality of expert judgement: Issues and analysis. *Decision Support Systems* 11(1):1–24.
Borghesi, Richard. 2008. Widespread corruption in sports betting: Fact or fiction? *Southern Economic Journal* 74(4):1063–1069.
Borghesi, Richard, and William H. Dare. 2009. A test of the widespread-point-shaving theory. *Finance Research Letters* 6(3):115–121.
Cain, Michael, David Law, and David A. Peel. 2001. The incidence of insider trading in betting markets and the Gabriel and Marsden Anomaly. *Manchester School* 69(2):197–207.
——. 2003. The favourite-longshot bias, bookmaker margins and insider trading in a variety of betting markets. *Bulletin of Economic Research* 55(3):263–273.
Camerer, Colin F. 1995. Individual Decision Making. In *The handbook of experimental economics*, edited by John H. Kagel and Alvin E. Roth. Princeton, N.J.: Princeton University Press, 587–703.
——. 1998. Can asset markets be manipulated? A field experiment with racetrack betting. *Journal of Political Economy* 106(3):457–482.
Ceci, Stephen J., and Jeffrey K. Liker. 1986. A day at the races: A study of IQ, expertise, and cognitive complexity. *Journal of Experimental Psychology* 115(3):255–266.
Coleman, Les. 2004. New light on the longshot bias. *Applied Economics* 36(4):315–326.
——. 2007a. Just how serious is insider trading? An evaluation using thoroughbred wagering markets. *Journal of Gambling Business and Economics* 1(1):31–55.
——. 2007b. Measurement of insider trading in wagering markets. *Applied Financial Economics* 17(5):351–356.
Damodaran, Aswath. 2002. *Investment valuation: Tools and techniques for determining the value of any asset.* New York: John Wiley & Sons.
Duggan, Mark, and Steven D. Levitt. 2002. Winning isn't everything: Corruption in sumo wrestling. *American Economic Review* 92(5):1594–1605.
Fama, Eugene F. 1970. Efficient capital markets: A review of theory and empirical work. *Journal of Finance* 25(2):383–417.
——. 1991. Efficient capital markets: II. *Journal of Finance* 46(5):1575–1617.
Gabriel, Paul E., and James R. Marsden. 1990. An examination of market efficiency in British racetrack betting. *Journal of Political Economy* 98(4):874–885.
Gandar, John M., William H. Dare, Craig R. Brown, and Richard A. Zuber. 1998. Informed traders and price variations in the betting market for professional basketball games. *Journal of Finance* 53(1):385–401.
Golec, Joseph, and Maurry Tamarkin. 1998. Bettors love skewness, not risk, at the horse track. *Journal of Political Economy* 106(1):205–225.
Hausch, D. B. and Ziemba, W. T. (1995). Efficiency of sports and lottery betting markets. Published in R. A. Jarrow, V. Maksimovic and W. T. Ziemba (ed). *Handbooks in Operations Research and Management Science.* Elsevier Science. 9:545–580.
Howell, W. (1971). Uncertainty from internal and external sources. *Journal of Experimental Psychology* 89:240–243.
Kaplan, Michael. 2002. The high tech trifecta. *Wired Magazine*, March.
Kumar, Alok. 2009. Who gambles in the stock market? *Journal of Finance* 64(4):1889–1933.
Lee, Choong-Ki, Yong-Ki Lee, Bo Jason Bernhard, and Yoo-Shik Yoon. 2006. Segmenting casino gamblers by motivation: A cluster analysis of Korean gamblers. *Tourism Management* 27(5):856–866.

Lee, Heung-Pyo, Paul Kyuman Chae, Hong-Seock Lee, and Yong-Ku Kim. 2007. The five-factor gambling motivation model. *Psychiatry Research* 150(1): 21–32.

Lessmann, Stefan, Ming-Chien Sung, and Johnnie E. V. Johnson. 2009. Identifying winners of competitive events: A SVM-based classification model for horserace prediction. *European Journal of Operational Research* 196(2):569–577.

Levitt, Steven D. 2004. Why are gambling markets organised so differently from financial markets? *Economic Journal* 114(495):223–246.

Maennig, Wolfgang. 2005. Corruption in international sports and sport management: Forms, tendencies, extent and countermeasures. *European Sport Management Quarterly 5*(2): 187–225.

Neighbours, Clayton, Ty W. Lostutter, Jessica M. Cronce, and Mary E. Larimer. 2002. Exploring college student gambling motivation. *Journal of Gambling Studies* 18(4):361–370.

Peirson, John D. 2011. The economic analysis of sports betting by expert gamblers and insiders: A survey. In *Prediction markets: Theory and applications*, edited by Leighton Vaughan Williams. London: Routledge, chap. 13.

Preston, Ian, and Stefan Szymanski. 2003. Cheating in Contests. *Oxford Review of Economic Policy* 19(4):612–624.

Preston, Malcolm G., and Philip Baratta. 1948. An experimental study of the auction-value of an uncertain outcome. *American Journal of Psychology* 61(2):183–193.

Raylu, Namrata, and Tian P. S. Oei. 2002. Pathological gambling: A comprehensive review. *Clinical Psychology Review* 22(7):1009–1061.

Sauer, Raymond D. 1998. The economics of wagering markets. *Journal of Economic Literature* 36(4):2021–2064.

Shin, Hyun Song. 1993. Measuring the incidence of insider trading in a market for state-contingent claims. *Economic Journal* 103(420):1141–1153.

Templeton, S. 2003. Free rein. *Herald Sun* (Melbourne), March 26.

Thaler, Richard H., and William T. Ziemba. 1988. Parimutuel betting markets: Racetracks and lotteries. *Journal of Economic Perspectives 2* (2): 161–174.

Tversky, Amos, and Craig R. Fox. 1995. Weighing risk and uncertainty. *Psychological Review* 102(2):269–283.

Tversky, Amos, and Daniel Kahneman. 1992. Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty* 5(4):297–323.

Vaughan Williams, Leighton. 2000. Can forecasters forecast successfully? Evidence from UK betting markets. *Journal of Forecasting* 19(6):505–513.

Winter, Stefan, and Martin Kukuk. 2008. Do horses like vodka and sponging? On market manipulation and the favourite-longshot bias. *Applied Economics* 40(1):75–87.

Wolfers, Justin. 2006. Point shaving: Corruption in NCAA basketball. *AEA Papers and Proceedings* 96(2):279–283.

CHAPTER 26

EVIDENCE OF BIASED DECISION-MAKING IN BETTING MARKETS

DAVID MCDONALD, MING-CHIEN SUNG, AND JOHNNIE JOHNSON

Psychologists have long been aware of the limitations of normative models of judgment and decision-making. Herbert Simon's (1955) work on bounded rationality criticized rational models of decision-making for disregarding such factors as the individuals' limited cognitive capacity. Subsequently experimental psychologists confirmed through a series of experiments that decisions are systematically biased in many ways, with decision-makers adopting rules of thumb or "heuristics" in order to more rapidly solve complex problems (Kahneman et al. 1982). However, the vast majority of research in this area has involved experimental investigations conducted under controlled laboratory conditions. This has led to researchers questioning the generalizability of the results (e.g., Bruce and Johnson 2003; Levitt and List 2007). In particular it is well understood that laboratory experiments cannot replicate the richness and complexity of real-life situations. As Léon Festinger (1953, 141) noted: "In the most excellently done laboratory experiment, the strength to which the various variables can be produced is extremely weak compared to the strength with which these variables exist and operate in real life situations." Naturalistic environments offer an attractive alternative for examining decision-making behavior, featuring subjects who are experienced in the task at hand (whereas most laboratory experiments involve naive subjects with little domain knowledge) and who are not aware that their actions are being scrutinized (as in most laboratory experiments).

Betting markets constitute naturalistic decision-making environments that offer great potential for helping to understand individuals' decision-making behavior. These markets feature many of the aspects of real-world decision environments. In particular, they are associated with rich, dynamic information sets, offer strong incentives to participants for success, require the commitment of the individual's own resources, and involve repeated trials, offering significant potential for learning. This chapter provides a survey of previous studies that have employed betting markets of various kinds to investigate the decisions made by bettors with particular reference to systematic biases that were first identified in the laboratory.

The remainder of this chapter is structured in four main parts. First we summarize the debate over the generalizability of laboratory findings and identify the ways in which naturalistic environments offer an alternative for studying the extent to which individuals' decisions are biased. In particular we outline the usefulness of betting markets and review a range of studies that have demonstrated that bettors are in many ways rational and well-calibrated decision-makers. Secondly, we briefly discuss the widely documented favorite-longshot bias, with particular attention to studies concerned with the psychological factors that may cause the bias. Third, we address two decision biases, anchoring and herding, each of which involve judgments of some unknown quantity being unduly influenced by external stimuli. Finally, we survey studies that have investigated biases that result from a failure of individuals to recognise randomness: the gambler's fallacy and the hot hand fallacy.

1 Betting Markets as a Naturalistic Environment in which to study Decision-Making

1.1 The Generalizability of Findings from Laboratory Studies

At the heart of this discussion is the distinction between experiments conducted under controlled conditions in artificial laboratory settings and analysis of data obtained from naturalistic environments, such as casinos, lotteries, and markets for betting on horse races or other sports.

While experiments can be carried out under controlled conditions in artificial "real-world" environments, we define a naturalistic environment to be one that "has not been artificially manipulated (i.e., a nonexperimental setting)" (Johnson and Bruce 2001, 266). This distinction is crucial, and there is a long-running debate concerning the relative merits of the two alternative methodologies when employed in experimental psychology (e.g., Ebbesen and Konečni 1980; Hogarth 1981; Funder 1987, Bruce and Johnson 2003) or economics (Harrison and List 2004; Levitt and List 2007). Levitt and List (2007) pointed out that a critical assumption in experimentation is that results generalize to the broader population. This generalizability or "external validity" has been seriously questioned because of significant variations in observed behavior between laboratory and naturalistic environments (e.g., Ebbesen and Konečni 1980; Koehler 1996). The factors that have been identified as limiting the generalizability of laboratory experiments (cf. naturalistic studies) include the following:

1. *Context*: The context in which decisions are evaluated is of central importance. Laboratory environments often present simplified versions of tasks that may be more complex in real-world environments. As a result laboratory experiments may unintentionally omit variables that are influential in the natural setting. It has been reported that significant differences in behavior may depend only on small changes to the experimental conditions (Ayton and Wright 1994), and Glenn Harrison and John List (2004, 1010) noted that

> although it is tempting to view field experiments as simply less controlled variants of laboratory experiments, we argue that to do so would be to seriously mischaracterize them. What passes for "control" in laboratory experiments might in fact be precisely the opposite if it is artificial to the subject or context of the task.

In addition, there are things that the experimenter cannot control, such as past experiences or social norms, which can affect the results (Levitt and List 2007). Furthermore, laboratory experimentation is often conducted under a condensed time frame rather than the extended period within which interaction occurs in naturalistic settings. In the real world, cognitive processes are "trained" over time and individuals develop strategies that can handle redundant and unreliable data. However, these strategies prove inappropriate when tackling the normal "static" tasks set in laboratory experiments. Biases in judgment recorded in the laboratory may simply be a response to that particular laboratory stimulus, and those same biases may not occur under ordinary circumstances (even while resulting from the same cognitive processes). For example, when mistakes are made in visual perception tasks in the laboratory, it is usually assumed that the mechanisms that result in the error generally produce correct judgments in real life (Funder 1987).[1]

2. *Experience*: Laboratory-based studies typically use university students, who may be inexperienced in tackling the kind of tasks with which they are presented. It is possible for inexperienced subjects to misinterpret the problem, whereas this is far less likely among participants experienced with the task. Robin Hogarth (1981) highlighted the importance of feedback in making correct decisions over the continuous time period often associated with real-world decision-making tasks. This is not possible in many "one-shot game" laboratory studies where there is no potential for the participants to learn from their mistakes. Laboratory participants often lack expertise in the tasks presented to them, so they fail to apply the correct strategies. Even worse, they frequently carry "baggage": behavior learned in the outside world entirely unsuited to the problem at hand (e.g., Burns 1985). Furthermore, a number of studies demonstrate large differences between the decision strategies of experts and novices in terms of the way they think, the information set and the nature of the decision models they employ, and the speed and accuracy of their problem solving (e.g., Larkin et al. 1980).

3. *Scrutiny*: Participants in laboratory experiments, who are generally aware that they are being investigated, may be keen to project a particular image (even if they have no idea of the purpose of the experiment). The student volunteers studied in most investigations are more likely to be "scientific do-gooders" (e.g., interested in the research or seeking approval from the experimenter) with unusually high awareness of the moral implications of their decisions (Levitt and List 2007). Scrutiny may therefore exaggerate the importance of pro-social behaviors, such as altruism and fairness. Conversely, the anonymity that is often present in real settings may allow decision-makers to feel that they are able to avoid being judged morally.

4. *Incentives*: Laboratory experiments are usually conducted with relatively trivial rewards for success. However, in the real world, decision-makers are often involved in high-stakes environments where they must commit their own or others' resources and where the results of their decisions can have significant personal consequences. These high-stakes environments can, therefore, involve a meaningful degree of risk. This can lead to a marked difference in risk-taking behavior between laboratory and real-world environments (Yates 1992). For example, the lack of excitement and low arousal levels in laboratory studies may lead to behaviors that would not be present in real settings (Anderson and Brown 1984).

The issues discussed above may all limit the potential for generalizing the biased behavior often found among laboratory participants to the wider population. However, to discard laboratory findings outright would be naive (Hogarth 1981). Rather, data gathered in the laboratory and under naturalistic conditions have their own strengths and weaknesses, and these data should be considered complementary (Keren and Wagenaar 1985). For instance, naturalistic work suffers from the inability to use control groups and difficulties associated with the replication of results. In addition, laboratory-based investigations are usually more cost-effective and afford the possibility of isolating specific variables.

1.2 Betting Markets as Valuable Naturalistic Environments

Betting markets, whether markets for bets on horse races, sports, or lotteries, offer an ideal naturalistic environment in which to explore biased decision-making. A key pragmatic advantage is the availability of extensive, rich, and detailed quantitative data relating to bettors' decisions. Since markets are finite in nature, there is a continually expanding set of "completed" markets, that is, a time period during which betting continues up to a defined endpoint, at which time all bets are settled in an unambiguous manner.[2] Furthermore, there is potential for comparative analysis across different types of event or bet, according to such recognized criteria as the quality (e.g., Smith et al. 2006), time of day (e.g., McGlothlin 1956), or complexity (e.g., Johnson and Bruce 1998) of the event. Thus it is possible to control for some aspects of the decision setting. Most importantly, betting markets include many of the factors regarded as distinctive to naturalistic decision-making (Orasanu and Connolly 1993): uncertain dynamic environments, poorly-structured problems, high stakes, time stress, action/feedback loops, and multiple players. Each element of the decision-making event (i.e., the bet) is unique: no two horse races or football matches are the same. Thus the outcome is uncertain, and the information relating to that outcome is often (as it is in many real-world decision environments) ambiguous, vague, or redundant. For example, it is not obvious how to combine the various factors that might enable one to predict participants' performance. The dynamic nature of betting markets is evidenced by the constantly updating prices as bettors with diverging opinions participate in the market.

Bettors, like many decision-makers in real-world environments, often risk meaningful amounts of money while under stress from time pressures (the window of opportunity in a betting market may last only minutes, or even seconds). A further important feature of these markets is the repetitive nature of betting. Since events take place regularly and often, there is potential for gaining familiarity and expertise with the task. Betting markets involve action-feedback loops; once bets have been placed and a market is closed and decided, bettors receive relatively unambiguous feedback on the success of their decisions, and this can be incorporated into future decisions (Goodman 1998). Also, betting markets involve multiple players, and it has been shown that the interaction between individuals in markets can significantly reduce errors (Wallsten et al. 1997). This results from a variety of causes, not least the fact that different individuals use different decision-making procedures and have diverse information gathering skills. As a result, their reaction to the same information may vary. Consequently, the final prices that emerge in these markets take into account a wide range of information and the forecasts of many individuals, and studies show that combining diverse forecasts generally leads to significantly more accurate predictions (e.g., Grant and Johnstone 2010; Vlastakis, Dotsis, and Markellos 2009). In addition, betting markets are not subject to several of the limitations of laboratory investigations listed above. For example, bettors are unaware that their decisions may be scrutinized, as they are not directly volunteering to take part in an experiment; instead, betting patterns are analyzed in such a way as to observe their decisions unobtrusively.

1.3 Analyzing Decision-Making Using Betting Market Data

The operation of betting markets is straightforward, which helps in the analysis of decision-making behavior. Specifically, in a betting market individuals are able to place bets on a set of outcomes of a particular event. For instance, in the simplest of markets for betting on a horse race with n runners, n different bets are available, one for each horse to win the race. After the market has closed and the race has taken place, each bet pays a return, £r_i, for each £1 staked if horse i wins the race but pays nothing otherwise. While the returns, r_i (usually referred to as the "odds" against each outcome), are determined differently according to the type of market and event, generally they depend on the relative amounts bet on each outcome by all the market participants. Consequently, bettors have an incentive to continue to place money on each outcome until the returns reflect the market's best estimate of that outcome's probability of occurring (Figlewski 1979). Therefore, a typical approach to assessing decisions in betting markets is summarized (with reference to horse race betting) by R. M. Griffith (1949, 290) as follows:

> the odds on the various horses in any race are a functioning of the proportion of the total money that is bet on each and hence are socially determined. On the other hand, the objective probability for winners from any group of horses is given a posteriori by the percentage of winners. Thus the odds express (reciprocally) a psychological probability while the percentage of winners at any odds group measures the true probability; any consistent discrepancy between the two may cast light not only on the specific topics of horse-race betting and gambling but on the more general field of the psychology of probabilities.

So, the "socially determined" prices in betting markets reflect the "subjective probabilities" assigned to each possible outcome by the bettors, in aggregate. The results of the event then determine the "objective probabilities." A comparison of subjective and objective probabilities thus allows an evaluation of any biases in bettors' decisions. If the betting is such that the relative volumes of betting on each outcome introduce a systematic bias, this can be detected by researchers.

A drawback of most betting market research is that, for ethical and/or practical reasons, it is usually not possible to obtain information relating to the decisions of individual bettors. Instead, subjective probabilities are an aggregation of opinions of all bettors. Hence it is possible that "certain biases present in an individual bettor's decisions are being counterbalanced by opposite biases in other bettors' decisions" (Johnson and Bruce 2001, 280). Colin Camerer (1987, 982) noted that a common argument for the rationality of market participants is that "random mistakes of individuals will cancel out" but also offered the counterargument that "biases found by psychologists are generally *systematic*—most people err in the same direction." Thus the best we can hope for in betting market research is evidence of systematic bias.

A further weakness of employing betting market data to examine decision behavior is that psychologically significant biases also hold an economic significance. Consequently, if some bettors (even a small group) become aware of an overall disparity between subjective and objective probabilities, they can potentially profit by betting against the bias. This could reduce the extent to which any systematic bias that exists among bettors is detectable from aggregate betting market data. Fortunately for researchers, transaction costs ensure that it is rarely possible to entirely arbitrage away biases.

1.4 Calibration of Bettors' Judgments

Given the above discussion, it might be expected that bettors display significantly less biased judgment in their natural domain than that demonstrated among naive participants in laboratory experiments. Indeed, a number of studies have investigated bettors' rationality and calibration. Specifically, Richard Rosett (1965) found that horse race bettors are generally sophisticated and rational agents who will not forgo combinations or sequences of bets when such bets offer a higher probability of winning for the same return or a higher return for the same probability of winning. Furthermore, results reveal a high correlation between expected returns and realized winning probabilities, suggesting that bettors are familiar with their decision-making environment and are able to accurately forecast risky outcomes.[3] Rosett (1965, 596) noted that

> if these gamblers behave as though they know statistical prediction methods and the probability calculus, it seems reasonable to suppose that, in a variety of other circumstances, human beings can be expected to respond appropriately to risky situations merely after having had sufficient experience with them.

Johnnie Johnson and Alistair Bruce (2001) also investigated the calibration of horse race bettors' subjective probability judgments. They found that bettors' subjective probabilities are not significantly different from the observed objective probabilities. They noted that while there is substantial evidence of poor calibration among decision-makers, this may reflect on the specific laboratory experiments involved. For example, James Shanteau (1992) suggested that task characteristics may account for differences observed in the quality of experts' judgments; specifically, more competent performance is likely if the decisions involve stimuli that are relatively constant, the tasks undertaken are repetitive, and decision aids are widely available. Furthermore, it has been empirically observed that violations of rationality are reduced under the multiple-play conditions that exist at the racetrack (e.g., Keren and Wagenaar 1987). Johnson and Bruce's (2001) study therefore suggests that bettors are skilled in a similar way to weather forecasters, who are also required to make frequent risky forecasts (Murphy and Brown 1984). Arthur Hoerl and Herbert Fallin (1974) also found no significant difference between subjective and objective probabilities in horse races. They argued that this was due to the high incentives available for successful gambling.

Not only are bettors well calibrated in general, but they are able to constantly adapt to uncertain and dynamic information. Johnson, Raymond O'Brien, and Ming-Chien Sung (2010) investigated how bettors respond to changing information. They set out to test Gerd Gigerenzer's (2000) assertion that evolution has equipped individuals to process probabilistic information from frequencies observed in a natural environment. They investigated the extent to which horse race bettors accounted for post position bias (an advantage/disadvantage afforded to the horses depending on their position in the starting stalls for the race), a factor shown to be a particularly important determinate of race outcome at the racetrack examined. Despite the fact that track managers employed a variety of procedures to change the bias (even between two consecutive races on the same day and often unannounced) bettors were able to account for most of the dynamic and changing information through regular outcome feedback, over a period of 6 years. This important finding may be accounted for by the fact that (i) bettors have a strong motivation to make accurate probability judgments, as their own financial resources and often their peer group esteem depend on the outcome of their decisions (Saunders and Turner 1987), and (ii) those who frequently make probability judgments are often better calibrated (Ferrell 1994). It has also been shown that bettors' calibration is generally improving over time (Smith and Vaughan Williams 2010) and that expert bettors employ complex mental models encompassing a wide range of variables and interactions between these variables (Ceci and Liker 1986).

In summary, naturalistic environments, and betting markets in particular, offer rich, complex settings in which to examine decision-making biases that have been observed in the laboratory. Due to a number of factors, such as learning, outcome feedback, and incentives, bettors appear to be more rational, well calibrated, and able to adapt to dynamic information than participants in laboratory studies. However, there are a number of ways in which bettors are biased; the first, and most widely documented of these, is the favorite-longshot bias, which is the focus of the next section.

2 Favorite-Longshot Bias

By far the most widely reported departure from rationality reported in the betting literature is that of the favorite-longshot bias (FLB). Reported over many decades and in many jurisdictions around the world, the bias is the phenomenon whereby returns to bets are such that the chances of low-/high-probability events (long shots/favorites) are over-/under-estimated.

2.1 Laboratory Evidence of the Bias

Malcolm Preston and Philip Baratta (1948) provided early laboratory evidence of the bias. They were concerned that "rational" theories of behavior could not universally explain peculiarities in the way people approached "wagering games" (i.e., games in which participants are required to bet on an uncertain outcome). They hypothesized that players might apply a scale of "psychological" probabilities to outcomes that are not necessarily the same as the mathematically correct probabilities of those outcomes. In order to investigate this possibility they carried out games with both undergraduate students and faculty members (the latter were more experienced in the fields of mathematics, statistics, and psychology). The game required the participants to compete against each other, bidding for the chance to win a given prize with a given probability. They found that the players tended to pay too generously for outcomes with low probabilities and not high enough for outcomes with high probabilities. This result was independent of the value of the prizes. The indifference point, where the psychological and mathematical probabilities corresponded, was found to be about 0.20. Moreover, the faculty members also displayed the bias (though to a lesser extent than the undergraduates) despite in many cases appearing to actively employ mathematics when forming their decisions. This suggests that expertise only partially eliminates the bias. The experimental findings of Preston and Baratta have since been confirmed in numerous laboratory experiments (e.g., Yaari 1965; Rosett 1971; Lichtenstein et al. 1974; Piron and Smith 1994).

2.2 Evidence and Explanations for the Bias

The first naturalistic evidence of the FLB was from the psychologist R. M. Griffith (1949). He was inspired by the laboratory evidence of Preston and Baratta (1948) but keen to test the results in a complex, non-laboratory environment. Employing U.S. racetrack data, Griffith found that horses with low probabilities of winning were systematically overvalued while horses with high probabilities of winning were systematically undervalued. This result was consistent with that of Preston and Baratta, with a similar indifference point of about 0.20. William McGlothlin (1956) replicated (and expanded upon) Griffith's study with a larger dataset. His data also confirmed the existence of the FLB.

In the decades that followed the original studies a significant body of evidence for the bias emerged in betting markets around the world (e.g., in the United States: Ali 1977; Asch, Malkiel, and Quandt 1982; Thaler and Ziemba 1988; in the United Kingdom: Dowie 1976; Vaughan Williams and Paton 1997; in Australia: Tuckwell 1983; in New Zealand: Gandar, Zuber, and Johnson 2001).[4] The emphasis in the research then shifted toward attempting to explain the origins of the bias. As a result, a broad range of explanations have been offered, including, for example, the "bragging rights" associated with holding a winning longshot ticket (Thaler and Ziemba 1988) or the additional excitement derived from longshot betting (Bruce and Johnson 1992). Robert Henery (1985) suggested that bettors may discount a fixed proportion of their losing bets, leading them to believe that longshot bets are more attractive. Alternatively the bias may arise from particular characteristics of the market itself, such as the cost of obtaining information and transaction costs (Hurley and McDonough 1995) or the defensive pricing policies adopted by bookmakers (Shin 1991). In this chapter we simply provide an overview of the significant debates concerning the origins of the FLB from the perspective of bettors' decision behavior; for more comprehensive explorations see surveys by Richard Thaler and William Ziemba (1988), Raymond Sauer (1998), Leighton Vaughan Williams (1999), Bruno Jullien and Bernard Salanié (2008), and Marco Ottaviani and Peter Sørensen (2008).

2.3 Do Bettors Love Risk, or Do They Misestimate Probabilities? Expected Utility Theory versus Prospect Theory

One strand of the FLB literature in particular warrants attention because it has led to an important intellectual debate concerning the relative merits of two prominent competing theories for explaining decision-making in wider fields: *expected utility theory* and *prospect theory*. The building block for this debate is the "representative bettor." Martin Weitzman (1965) introduced Mr. Avmart, a fictitious person who represents the "social average" of all bettors. Weitzman's (1965, 26) innovation was to infer the preferences of the "most typical" bettor from the population of bettors

> instead of concentrating on individuals and trying to derive utility generalizations from their experimental behavior, more nearly the converse approach was attempted. A plethora of data concerning the collective risk actions of parimutuel bettors was employed in investigating utility aspects of the behavior of a hypothetical member of the group.

Weitzman was concerned primarily with constructing Mr. Avmart's utility of wealth curve (the mathematical representation of preferences over various monetary outcomes and the basis of expected utility theory). He found that the FLB in the data was best explained by a convex utility of wealth curve, indicating that the average bettor is locally risk loving (i.e., the average bettor prefers the riskier, low-probability outcomes). Richard Quandt (1986) extended the analysis by showing that the bias is the natural result of equilibrium in a market where the average bettor is risk loving. The findings of Mukhtar Ali (1977) and Shahid Hamid, Arun Prakash, and Michael Smyser (1996) also supported this hypothesis.

However, there are alternative scenarios that can explain the biased decisions of the representative bettor. So, for instance, Joseph Golec and Maurry Tamarkin (1998) showed that the FLB can arise if bettors are risk averse in general but with a preference for skewness of returns. An alternative explanation stems from the motivation behind the original Preston and Baratta (1948) study. In this study it was supposed that the "psychological" probabilities assigned to uncertain outcomes were systematically biased in such a way that small/large probabilities are over-/under-estimated. If this is the case, then the FLB can be explained solely with reference to bettors' systematic misestimation of probabilities (i.e., bettors need not be locally risk loving). This was formalized in Daniel Kahneman and Amos Tversky's (1979) *prospect theory* (later extended and renamed cumulative prospect theory; see Tversky and Kahneman 1992). The important feature of prospect theory for this discussion is that objective probabilities are transformed into subjective decision weights that allow for biases in the estimation of probabilities.

Hence there are now two broadly competing sets of theories regarding the explanations for the bias in terms of the representative bettor: are bettors unbiased in their estimation of probabilities but risk loving, or are bettors risk-neutral but biased in their estimation of probabilities? Unfortunately there is no straightforward answer. As Menahem Yaari (1965, 278) commented,

> at first blush it seems as though one cannot, by looking at empirical data, choose between the two hypotheses (distortion of utility versus distortion of probability) because utility and probability are two purely theoretical components of an integral decision process. Thus, the two hypotheses are empirically indistinguishable, and choosing between them is a matter of taste.

However, some researchers have made progress in this regard. Golec and Tamarkin (1995) noted that risk love cannot explain the relatively unfair returns for the low-risk, low-return, side bets offered by some bookmakers. Instead they suggested that overconfidence (which is consistent with bettors overestimating small probabilities) better explains the FLB. Jullien and Salanié (2000) found that prospect theory (cf. expected utility theory) better explains the bias for standard bets, though computational limitations of this approach restricted their analysis. Ian Bradley (2003) adapted the prospect theory approach of Jullien and Salanié by accounting for bet size and found an even better fit to the data.

More recently Erik Snowberg and Justin Wolfers (2010) set out to test the competing theories using a novel approach and a large dataset of all the horse races run in North America from 1992 to 2001 (over 865,000 races). They first estimated the parameters of the two models (the expected utility/risk-love model and the prospect theory/misestimation of probabilities model) by fitting the models to standard "win" bets (bets that a horse will finish in first place). They then examined compound exotic bets, such as the exacta, a bet that two horses will finish a race in first and second place in a specific order. Snowberg and Wolfers reasoned that because bettors would bet in the same manner in the exotic and win betting pools, the same models should apply for each bet type. Accordingly, they used the fitted models to predict expected market prices in the exotic betting pools and compared their predictions with the actual prices on offer. They found that the misestimation of probabilities model predicted exotic bet prices more accurately than the risk-love model. Snowberg and Wolfers concluded that, with respect to the representative bettor, prospect theory explained the FLB more effectively than expected utility theory.

An important issue in this debate is the validity of the assumption that bettors' decisions can be averaged by the representative bettor. In the third section of this chapter we show that the distinction between different types of bettors (on the basis of the quality of the information they hold or how they handle this information) is crucial to fully understanding some other biases in betting behavior. Russell Sobel and Travis Raines (2003) demonstrated this by differentiating between "serious" and "casual" bettors. They identified serious bettors as those who attend the racetrack on week nights, bet larger sums, and bet to a greater extent on more complicated types of bet. Casual bettors, conversely, attend primarily on weekends and bet smaller sums on simpler types of bets. Sobel and Raines found evidence of the FLB, but the bias was significantly reduced in those races that involved a higher proportion of serious bettors.

2.4 The Late-Race Effect

A curious element of the nature of the FLB is its apparent tendency to vary in a systematic manner over the course of a day's betting activity. In horse race betting markets in particular it has been found that the extent of the bias appears to increase significantly in the last race or the last few races of the day. This phenomenon has become known as the late-race effect. Early evidence of this pattern was uncovered by McGlothlin (1956), who was investigating the stability of the FLB over the course of the day. He found that the bias was present in the data as a whole but that bettors did not underbet favorites in the seventh race of the day (out of eight). McGlothlin argued that the seventh race was usually the feature race, involving more coverage and scrutiny of favorites, so it might be expected that bettors would prefer the favorites in these races. However, he found that, in the eighth and last race of the day, bettors underbet favorites to a greater extent than in any other race. He suggested that bettors might avoid bets on favorites in the last race because winning such bets would not recoup earlier losses (the track take ensured that most bettors would finish the day out of pocket). Rather, McGlothlin suggested that they preferred to bet on long shots, hoping for a lucky win in order to end the day in profit.

Over time, as more evidence of the late-race effect emerged, it was explained in terms of the risk-loving attitudes of the representative bettor. For example, Ali (1977), who found a greater degree of the FLB in the last race than in the first two races of the day, posited that this demonstrated that bettors, who were on average risk loving, became more risk loving as the day progressed. Similarly, Asch, Malkiel, and Quandt (1982) replicated McGlothlin's (1956) results, though in their study the extent of the bias was greater in the last two races of the day. Mary Ann Metzger (1985) also found evidence of the effect but only if the first race of the day was excluded from the analysis. The late-race effect soon passed into betting lore, with Richard Kopelman and Betsy Minkin (1991) describing how an avid racing enthusiast known as "Gluck" espoused the rule: "The best time to bet the favourite is in the last race." Kopelman and Minkin's analysis confirmed that there was a sound economic basis for Gluck's rule.

More recent evidence has thrown the existence of the late-race effect into question. Johnson and Bruce (1993) found that bettors in U.K. betting shops tended to place more bets on favorites in the last race and suggested that this might be due to a "break-even" effect whereby bettors seek to recover their losses by betting on outcomes that have at least a moderate success of actually occurring. This hypothesis is supported by evidence that decision-makers tend to exhibit loss aversion after a series of prior losses (Thaler and Johnson 1990). Similarly, Lawrence Brown, Rebecca D'Amato, and Randy Gertner (1994) observed a greater prevalence of the FLB in the last race of the day than in earlier races, but the difference was not statistically significant. Sobel and

Raines (2003) found that (having controlled for the differences in race grade) there was a slight increase on betting on long shots in the last two races of day (especially the last race), with no corresponding decrease on betting on favorites. However, they also found that the general trend over the latter half of the evening (i.e., over the last 7 or 8 races of the 15 in each meeting) is for bettors to begin to prefer favorites and shun long shots. They note that this could be explained by casual bettors leaving over the course of the evening (resulting in the remaining more serious bettors betting more on the favorites). Finally, Snowberg and Wolfers (2010) found no significant difference in the extent of the FLB in the last race of the day (in a dataset of over 850,000 races), suggesting that the late-race effect has now been eliminated.

From the contrasting evidence discussed above, it appears that bettors' increasing risk love over a day's betting cannot fully explain the late-race effect. Johnson and Bruce (1993) considered that their converse result (a decreasing FLB in the last race) could be explained by a "break-even" effect. However, a similarly plausible explanation is used by other authors to explain the opposite effect (an increasing FLB in the last race). Furthermore, it is not clear that expected utility theory is an adequate explanation. As Thaler and Ziemba (1988, 171) asked, "why should a reduction in wealth increase the tendency for risk seeking?" Camerer (2001) pointed out that expected utility theory cannot explain why the same bettor leaves the racetrack one day, arrives again the next, and adopts a completely different risk attitude. Thaler and Ziemba proposed that the effect can be explained by "mental accounting" whereby bettors partition their wealth into separate accounts and do not attempt to recoup losses in one account with funds from another. So the late-race effect could be explained by bettors opening a mental account at the beginning of the day and closing it at the end, with an increasing desperation to break even as the day progresses (Camerer 2001). Finally, the relative paucity of evidence for the effect in recent years could be attributed to a learning effect among bettors, as those who are aware of the effect are able to arbitrage it away should it reappear.

In summary, the FLB, while proving to be an interesting riddle for researchers, admirably demonstrates the value of naturalistic environments, betting markets in particular, in the study of decision-making. While some potentially unrealistic simplifications (such as the representative bettor) must sometimes be made when seeking explanations, the quality and quantity of betting market data have enabled the development of a large body of research on the nature of preferences and perceptions of risk under uncertainty.

3 Anchoring and Herding

Betting market research has largely focused on the FLB, but some studies have investigated whether bettors make biased decisions in other ways. In particular, anchoring and herding represent biased behavior whereby decision-makers alter their decisions to account for external stimuli. Thus when employing the anchoring and adjustment heuristic, decision-makers unnecessarily alter their judgments to reflect an initially provided estimate. Herding arises when decision-makers neglect their own information and alter their judgments to reflect those of others. This section details the findings of these studies.

3.1 Anchoring and Adjustment

Laboratory research suggests that when making a numerical estimate individuals, in an attempt to simplify the decision-making process, tend to start from an initial value and make "adjustments" upward or downward from it (e.g., Tversky and Kahneman 1974). However, this often results in a bias whereby the decision is "anchored" on the initial estimate and adjustments are not sufficient. This is known as the anchoring and adjustment heuristic. For example, Tversky and Kahneman asked participants pairs of questions, such as:

(a) Is the percentage of African countries in the United Nations higher or lower than 25?
(b) What do you think the exact percentage is?

They found that the figure given in (a) (i.e., 25 in the above example) significantly influenced the participants' responses to (b), even when the figure was randomly generated by spinning a wheel of fortune in the participants' presence. Higher/lower random numbers were associated with higher/lower estimates.

Anchoring has mainly been studied in controlled laboratory conditions. The few studies that have been conducted in naturalistic environments (e.g., among auditors: Bhattacharjee and Moreno 2002; and among real estate agents: Northcraft and Neale 1987) have generally concluded that anchoring does seem to occur in information-rich, real-world settings. However, these studies have used questionnaires or artificial problems. Consequently the advantages of studying anchoring in betting markets are that participants are making estimates that matter to them in a familiar, real-world environment without the use of questionnaires or artificial problems and that they do not alter their normal behavior (because they do not know they are being observed).

In the first study to investigate whether bettors anchor their judgments excessively, Shuang Liu and Johnson (2007) were primarily concerned with whether or not participants in horse race betting markets employed factors relating to previous performance of horses, jockeys, and trainers as anchors. For example, if a jockey had won his or her previous race, do bettors overestimate the chance that he or she will also win the current race? Previous finishing positions are not anchors in the *traditional* sense, since bettors are not specifically required to make direct comparisons between initial values and final judgment. Rather this study attempted to find evidence of *basic* anchoring, where decision-makers can be influenced by anchor values even when not asked to consider them directly (Wilson, Houston, and Etling 1996). Liu and Johnson investigated, using betting market data from Hong Kong, various explanatory variables that represent possible anchors (such as whether the horse won its previous race). However, the only significant explanatory variable was one that summarized a horse's finishing position over its career; this variable showed that bettors tend to *under*estimate horses that have a strong finishing record. Consequently it appears that bettors tend to ignore some useful information relating to the horses' potential (or are unable to effectively employ such a complicated variable). However, the key finding was that no other explanatory variables were significant, indicating that bettors do not anchor their judgments on previous performances.

It is possible that Liu and Johnson's (2007) results failed to identify the anchoring that does occur in betting markets since any bias created by the anchoring of most bettors could be arbitraged away by the remainder of bettors. For instance, it is well known (e.g., Benter 1994) that large betting syndicates, attracted by the unusually large betting volumes and strict regulation in Hong Kong (which helps to eliminate malpractice and insider trading), use sophisticated computer models to make considerable profits in this market.

Johnson, Adi Schnytzer, and Liu (2009) extended the analysis of Liu and Johnson (2007) in two ways. First, noting that bettors in Hong Kong often spend considerable time reviewing race results, they expected that barrier position (the stall position from which the horse starts the race) would be a significant anchor for bettors. Second, decision-makers with a higher level of expertise tend to be less susceptible to anchoring effects (e.g., Northcraft and Neale 1987), so they expected that more experienced bettors would be less prone to anchoring. They found that bettors as a whole did not anchor excessively on barrier position over all their data but that bettors overestimated the advantage offered by a good barrier position in one of the two racetracks under investigation. However, they found that expertise significantly reduced the extent of anchoring displayed by bettors (they used early and late betting as a proxy for inexpert and expert bettors, respectively). In summary, the two anchoring studies conducted in betting markets indicate that anchoring in real-world environments may be a more complex phenomenon than has been found in laboratory studies, suggesting that further research may be required to fully understand its influence on decisions in real-world environments.

3.2 Herding

Herding occurs when decision task participants neglect their own information and adjust their actions to be more representative of the actions of others.

Early theoretical models rationalized herding behavior as information cascades where decisions are made sequentially by different agents who each hold their own private information (e.g., Banerjee 1992; Bikhchandani, Hirshleifer, and Welch 1992; Avery and Zemsky 1998). The validity of the information is inherently uncertain, and as a result, individuals may be rational in disregarding some of their private information when the information held by other agents appears to conflict with their own. Hence strictly speaking, herding behavior in itself may not be "biased" decision-making. However, a biased outcome results from the combined effect of herding by multiple participants. In particular this behavior can lead to expected returns differing significantly from their "rational" value.[5]

Many empirical herding studies have been conducted in the laboratory. In general these studies have found that participants display herd behavior but to a lesser extent than theoretical models predict. However, evidence has generally been inconclusive (Spiwoks, Bizer and Hein 2008).[6]

Herding might be expected in betting markets because there is a belief that certain bettors have access to privileged information. It has been found that betting on a horse or team that subsequently attracts a high degree of betting interest during the course of the market (known as a "market mover" or "plunger") is, on average, profitable (e.g., Crafts 1985). The problem, of course, is that it is difficult to identify such opportunities before the fact, and this is where bettors with access to privileged information can gain an advantage. Bettors with superior information are often referred to as "insiders" in the literature because of the presumption that their information is not in the public domain (e.g., a racehorse owner may have knowledge of secret training programs). However, there are also some bettors who use only publicly available information but expertly combine all the information in such as a way as to form highly accurate opinions of the competitors' chances; these bettors are often referred to as "informed" bettors. The presence of insiders and informed bettors in betting markets is widely reported (e.g., Crafts 1985), and consequently, herding behavior may ensue when "uninformed" bettors interpret a significant price movement as a signal that a competitor is being backed by insiders or informed bettors and alter their bets accordingly.

The first study that investigated whether bettors herd is that of Camerer (1998). He tested whether bettors might respond to privileged information signals by placing large early bets in pari-mutuel pools at U.S. racetracks and recording subsequent betting patterns. The purpose of this field test was to investigate whether markets could be manipulated. However, by observing the reactions of bettors to the temporary bets (Camerer subsequently canceled the early bets), Camerer was also able to infer the relative proportions of "opinion bettors" and "full/partial rational expectations bettors." Opinion bettors do not take the current odds into account; instead they bet solely based on their own subjective probabilities of the relative chances of the horses. On the other hand, full rational expectations bettors believe that current odds fully reflect all available information and so always bet in proportion to the odds. Partial rational expectations bettors occupy the middle ground, believing that odds do to some extent reflect the information but also that other bettors do not react to this information. Since opinion bettors completely ignore price movements, they never herd. Full rational expectations bettors will herd to some extent; if odds move from, say, 20/1 to 12/1 after a "fake" signal (i.e., following one of Camerer's early bets), they will bet as if the horse is a genuine 12/1 shot. Partial rational expectations bettors will herd to the greatest extent;

they might bet a fake 12/1 down to 10/1. Camerer conducted two studies as part of his experiment. In the first study he placed 50 temporary $500 bets early in the market. He found that while his bets did temporarily distort the odds, after canceling his bets the odds returned to their expected levels (based on "control" horses with similar odds on which he did not bet), indicating that bettors were not responding to the fake signals. In a second study Camerer increased his bet size to $1,000 and targeted smaller racetracks and "maiden" races (for horses that had never won a race). He detected a weakly significant herding effect whereby bettors were more likely to respond in the maiden races. However, overall the results still resolutely showed that bettors did not display herding behavior. There remains an important caveat: although Camerer's bets made up of about 7 percent of the pool in the second study, they still may not have been large enough to induce herding.

In a later study David Law and David Peel (2002) argued that the apparent lack of herding in Camerer's (1998) study probably arose because while the bets were sufficiently large to temporarily distort the markets, there was little incentive for bettors to herd on the initial price movement since pari-mutuel bettors cannot lock in profits. To counter this, they conducted an empirical test for herding in U.K. bookmaker markets for horse racing. They argued that since the returns to a bet with a bookmaker are known at the time of bet placement, bettors might be more likely to herd in these markets. They noted that while an initial price movement could be due to informed trading, a further price movement may result from further informed trading or herding. Using the Hyun Song Shin (1991) measure of the degree of insider (or informed) trading, they were able to identify those large price movements that resulted from the trading of those with access to privileged information (the Shin measure increased over the duration of the market) or from herding (the Shin measure decreased). Law and Peel were particularly interested in those horses that opened at shorter odds than forecasted that then attracted significant betting interest. Significant positive returns of 10.2 percent could be made by betting on horses with these characteristics whose odds plunged as a result of informed trading; returns were significantly negative otherwise, at −10.9 percent. Consequently Law and Peel (2002) were able to demonstrate that herding led to biased prices, with negative/positive returns being reported when price movements were due to herding/informed betting.

Schnytzer and Avichai Snir (2008) examined herding in bookmaker markets for horse racing in both the United Kingdom and Australia. They developed a theoretical model which showed that herding leads to odds that overestimate a horse's chances of winning. Specifically, if a horse that is not attracting bets suddenly attracts a high degree of betting interest, that horse's chances are likely to be overestimated due to herding. However, noting that early plunges in odds suggest trading by bettors with privileged information, Schnytzer and Snir (2008, 3) hypothesised that, due to the limited budgets of insiders, "a short time later, when the odds on those runners are lengthened again, those insiders are either unable or unwilling to place bets of sufficient significance to affect prices, even when the odds on those runners have drifted back to initial levels or even further." This may arise because informed traders place most of their bets early in the market to secure profits. Schnytzer and Snir considered two possible situations: either odds increase early in the market and then decrease or odds decrease early in the market and then increase. In the former, the late betting interest on the horse is considered to be evidence of herding, since the horse attracted little interest in the early market, and the final odds are expected to overestimate the horse's chances of winning. In the latter the early plunge followed by a lack of betting interest in the late market was considered evidence of cash-constrained informed betting; that is, the final odds are expected to underestimate the winning horse's chances. The results demonstrated that for horses attracting early but not late betting interest, positive/negative returns of 15.3 percent/−10.3 percent were possible from a simple betting strategy for the Australian/U.K. races (though an 8.5% return was possible for U.K. races if stricter criteria were applied). On the other hand, only highly negative returns (as low as −27.2% in the Australian races) were possible for horses that lacked interest in the early market but were the subject of herding in the late market. These results confirmed that bettors herd and that this can lead to highly biased outcomes.

In summary, studies of anchoring and herding in betting markets have offered mixed conclusions. Camerer (1998) was unable to induce herding behavior with his "fake" signals, but other studies have found evidence of significant herding by bettors when insider trading is prevalent. However, it appears that bettors do not anchor their judgments to the extent that has been reported in the laboratory. This may result from the fact that bettors are making decisions in an environment with which they are familiar (cf. naive subjects in unfamiliar laboratory settings) and in which they have learned (e.g., through repeated trial and improvement) to handle appropriately the redundant information and decision-relevant cues. Equally, while many bettors may herd to a significant extent, the actions of informed bettors, who arbitrage on the herding behavior of others, may serve to suppress the observable effects of herding.

4 The Gambler's Fallacy and the Hot Hand Fallacy

The gambler's fallacy and the hot hand fallacy both involve a misunderstanding of the nature of randomness. The application of these fallacies often results in systematically biased behavior. The gambler's fallacy is defined as the belief that an event's probability of occurring is reduced after that event has occurred, even if the event is independent from one trial to the next (Rabin 2002). Pierre Simon Laplace ([1825] 1995, 92) gave the following examples from lotteries and coin tossing:

> when one number has not been drawn in the French lottery, the mob is eager to bet on it. They fancy that, because the number has not been drawn for a long time, it, rather than the others, ought to be drawn on the next draw…. It is, for example,

> very unlikely that in a game of *heads or tails* one will get *heads* ten times running. This unlikeliness, which surprises us even when the event has happened nine times, leads us to believe that *tails* will occur on the tenth toss.

The gambler's fallacy is the conviction that the coin, which is known, objectively, to be fair, is more likely to land heads than tails after the "streak" of nine tails. This belief is demonstrated in laboratory experiments where participants are asked to invent a random sequence, such as repeated tosses of a coin. The results show that people tend to produce sequences containing too many alternations in the outcome relative to genuine randomness (Falk and Konold 1997). The representativeness heuristic has been proposed as an explanation: the gambler believes that small samples must be representative of the population, so if unexpected sequences occur, a correction is expected (Tversky and Kahneman 1971). As Tversky and Kahneman (1974, 1125) noted: "chance is commonly viewed as a self-correcting process in which a deviation in one direction induces a deviation in the opposite direction." Since nine tails in a row is an extremely unlikely event, the observer committing the gambler's fallacy expects that the next toss should be heads in order to make the sequence of 10 tosses seem less unusual. A commonly cited example of this phenomenon is that of the Monte Carlo casino where, during a roulette game in 1913, black occurred 26 times in a row. During this streak customers bet increasing amounts on red, and the casino profited as a result (Lehrer 2009).

The hot hand fallacy involves mistaken convictions that run contrary to the gambler's fallacy. In particular, this fallacy involves the belief that if a player or team is on a winning (or losing) streak this streak will continue longer than should be expected in a random sequence. So in a game where the objective is to obtain tails on the toss of a coin, a gambler who has achieved the unlikely feat of landing tails nine times in a row believes that he or she is on a "hot streak" and therefore expects that the coin has a greater probability of showing tails than heads on the next toss.

Thomas Gilovich, Robert Vallone, and Tversky (1985) found that many basketball players and fans believed that a player would be more likely to score on a shot if he had scored (cf. missed) on the previous shot. However, they found no evidence to support this claim in either real games or controlled shooting experiments. The hot hand has been attributed to the illusion of control, which is the misplaced perception that gamblers have an element of control over random events (Langer 1975). In fact, it has been shown that some gamblers believe that luck is separate from chance and that their good fortune allows them to operate outside the laws of probability while they are on winning streaks (Wagenaar and Keren 1988). Gilovich, Vallone, and Tversky (1985) suggested that, as with the gambler's fallacy, bettors may be employing the representativeness heuristic. In this case, long runs are deemed too unusual for the representative sequence, so bettors infer that the sequence generating process is no longer random (e.g., a basketball player who shoots an usually high run of on-target shots is said to be "in the zone" or a roulette table or die is assumed to be biased). It is possible that, while a general belief in the hot hand may be misplaced, an accurate belief in the hot hand in specific instances motivates people to believe in its universality (see Bar-Eli, Avugos, and Raab 2006 for many examples of genuine hot hand effects).

The remainder of this section details the findings of studies that have investigated the two fallacies in naturalistic environments.

4.1 Evidence of the Gambler's Fallacy in Betting Markets

Charles Clotfelter and Philip Cook (1993) undertook one of the early studies using real betting data to investigate the gambler's fallacy. The U.S. state of Maryland runs a "daily numbers" draw lottery where a three-digit number between 000 and 999 is picked at random and the bettor wins if he or she selects this number. Clotfelter and Cook found that betting volumes on a number decreased in the days after the number was drawn before returning to original levels after 84 days. It was postulated that bettors could be reducing their bets on numbers that had been drawn previously because they thought that that number was less likely to appear again. However, Clotfelter and Cook were unable to eliminate a "wealth effect" from their data: bettors who regularly bet a particular number might stop betting altogether because they had achieved their financial goals. This could lead to a natural reduction in betting volumes on a winning number in the days and weeks after its appearance. A more significant caveat with Clotfelter and Cook's study was noted by Dek Terrell (1994): the Maryland lottery has fixed payouts (winners are always paid $500 on a $1 bet), so choosing numbers based on the gambler's fallacy does not reduce the expected return to the bettor.

Rachel Croson and James Sundali (2005) studied 18 hours of roulette play in a real casino, during which more than one hundred players placed thousands of bets. They found evidence of the gambler's fallacy after streaks of around five or more similar outcomes (e.g., five red numbers in a row). However, Croson and Sundali (2005, 200) pointed out a similar concern to that existing in the Clotfelter and Cook (1993) study: "Since the house advantage on (almost) all bets at the wheel is the same, there is no economic reason to bet one way or another (or for that matter, at all)."[7]

These studies highlight an important issue: while the gambler's fallacy is anecdotally known to be a common belief among gamblers, it does not always result in biased behavior. For example, in roulette the returns to bets on each outcome are independent of the bets placed by the customers. Therefore, the decision of which outcome to bet on is irrelevant. The gamblers in the Monte Carlo casino were not necessarily wrong to bet on red rather than black (though they might have bet more than they could afford). In such cases it is plausible that belief in the fallacy only adds to the excitement of the game.

In circumstances where acting on the fallacy results in a systematic bias that leads to a lower expected return for the bettor, it might be expected that the fallacy would be eliminated (e.g., by a learning process). However, there are a number of examples of the gambler's fallacy resulting in a systematic bias. These studies have necessarily needed to be creative in order to identify situations where one might expect evidence of the gambler's fallacy. For example, Metzger (1985) found evidence that horse race bettors tend to believe that streaks of favorites and long shots winning should cancel out. So, if a series of long shots wins, they bet more on favorites and vice versa. Terrell and Amy Farmer (1996) thought that bettors at greyhound racing events might believe that the starting positions of the winning dogs should be more random than it appears. Thus they might underestimate the winning chances of a dog starting in a given position from winning if the winner of the previous race also started from that position. Their calculations revealed that this was the case, with a positive return of $1.09 per dollar bet for a strategy of betting on dogs starting from the same position as the winner of the previous race. Terrell (1998) extended the study of Terrell and Farmer (1996) with a larger dataset but found significant evidence of the fallacy only in one of the two years in their data.

Terrell (1994) conducted a similar investigation to Clotfelter and Cook (1993) but in a pari-mutuel New Jersey lottery where payouts are shared between all the bettors who choose the winning number. Hence if many gamblers avoid numbers that have recently appeared, the expected return to these gamblers is reduced. As expected, the extent of the gambler's fallacy was lower in this case. However, there was still a tendency to avoid numbers that had recently appeared. Terrell also found that if the results of Clotfelter and Cook were converted to a pari-mutuel system there would be frequent occurrences when the payout would exceed $500, giving a positive expected return to bettors. This is not the case in New Jersey, so bettors appear to bet more evenly to avoid forgoing the increased potential winnings, and this diminishes the potential to exploit the fallacy. An alternative explanation for the results is that bettors simply prefer not to bet on a recently seen number in the same way that they prefer certain numbers (such as 777). Similarly, George Papachristou (2004) found only marginal evidence of the gambler's fallacy in the pari-mutuel lottery in the United Kingdom.

4.2 Evidence of the Hot Hand Fallacy in Betting Markets

As indicated above, the hot hand fallacy is also a mistaken perception of randomness. However, as with the gambler's fallacy, this mistaken belief does not necessarily impose economic penalties. Camerer (1989) examined the economic significance of the hot hand fallacy by investigating whether this mistaken belief is represented in gamblers' betting decisions. He categorized basketball teams based on their current winning or losing streak (in games) and then compared the actual results with the point spreads offered by bookmakers.[8] If bettors believe in the hot hand, point spreads will overestimate the chances of teams currently on winning streaks against the spread while underestimating the chances of teams on losing streaks. The results showed that the performance of teams on winning streaks is worse than predicted by point spreads and that teams on losing streaks perform better than predicted. However, the results were only marginally statistically significant.

William Brown and Sauer (1993, p. 1377), highlighted the importance of the following critical assumption in Camerer's (1989, p. 1257) study: "the hot hand is belief in a myth." Camerer was effectively testing two alternatives: either bettors believe in a mythical hot hand or they do not. However, there is evidence that genuine hot hand effects exist (Bar-Eli, Avugos, and Raab 2006). Consequently there is a third alternative: bettors believe in a genuine hot hand.[9] In this case, while bettors will move point spreads to account for the hot hand effect, so teams' performance levels will also change. Brown and Sauer considered all three alternatives in basketball point spread markets but found only mixed results. They could not reject the hypothesis that the hot hand is real and that bettors correctly account for it, but they could also not reject the hypothesis that bettors believe in a mythical hot hand.

In a further study on the hot hand in point spread markets for basketball, Dale Oorlog (1995) found strong evidence against the hypothesis that gamblers believe in the hot hand. Oorlog devised a number of betting strategies to account for possible hot hand effects, but none were profitable. Christopher Avery and Judith Chevalier (1999) investigated U.S. football betting markets and also found a small bias as a result of the hot hand fallacy, but, again, the magnitude of the effect was small.

Additional mixed evidence for the hot hand fallacy was provided by Gregory Durham, Michael Hertzel, and J. Spencer Martin (2005). These authors found that point spreads over-/underestimated U.S. college football teams on short winning/losing streaks against the spread, which is consistent with the hot hand fallacy. However, the point spreads suggested that bettors expected longer winning or losing streaks to end rather than continue. Similarly, Rodney Paul and Andrew Weinbach (2005) reported that betting against basketball teams on short winning streaks was profitable while betting against teams on longer winning streaks was not. Moreover, they found no hot hand effect for teams on losing streaks and suggested that this might be because bettors derived additional utility from betting on teams on winning streaks.

4.3 The Paradox of the Hot Hand and Gambler's Fallacies

An important consideration is that the hot hand and gambler's fallacies appear at first to be opposite effects. While bettors may believe that long runs in the results of players or teams will continue (the hot hand), they simultaneously believe that long runs should end (the gambler's fallacy). This begs the question: how can these two apparently opposite effects be explained?

One proposed explanation for both fallacies is the representativeness heuristic (Tversky and Kahneman 1971) in which decision-makers believe that sequences should be representative of the generating process. Decision-makers apply the "law of large numbers" too readily; that is, they believe in the "law of small numbers." In other words, while the relative frequencies of outcomes approximate the generating process in the long run, people believe that this should also be the case in the short run. So the gambler's fallacy is explained because people believe that unusually long streaks are not representative and so predict an alternation to make the sequence more representative. The hot hand is explained because people tend to over-infer from short sequences in a random process and decide that there is some underlying nonrandom process generating the sequence (Rabin 2002).

It is potentially problematic to explain opposite phenomena with the same principle. However, a solution was provided by Peter Ayton and Ilan Fischer (2004; see also Burns and Corpus 2004), who tested whether the type of random process employed to generate the result was consequential in whether decision-makers displayed the hot hand or the gambler's fallacy. They hypothesized that when outcomes reflect human performance people believe in the hot hand, whereas when outcomes reflect inanimate mechanisms people believe in the gambler's fallacy. This might explain why winning streaks of basketball and roulette players are perceived to exhibit long-run tendencies but outcomes of roulette games and lotteries are not. They conducted an experiment where they asked participants to play a simulated roulette-style game. Participants were first required to choose between red and blue and second were asked to rate their confidence in their prediction. The results confirmed that while people are more likely to predict an alternation after a long run of either color they are also more confident in their own ability after a long run of successful predictions. Ayton and Fischer (2004, 1374) concluded that while the sequences of outcomes (red or blue) and predictions (win or lose) are each identical independent processes, "the two sequences are psychologically perceived quite differently; subjects *simultaneously* exhibited both … the hot hand and the gambler's fallacy." In a second experiment they found that participants were more likely to attribute random sequences with low/high alternation rates to human performance/inanimate mechanisms. This line of experimentation goes some way to unravel the problematic nature of explaining two apparently opposite effects with the same heuristic.

In summary, there is evidence from a diversity of naturalistic betting environments that the decisions of bettors are consistent with the gambler's fallacy. However, the extent of the fallacy is reduced when it results in biased decisions, suggesting that bettors are sensitive to its economic significance. Research examining the hot hand fallacy in betting markets has been inconclusive. None of the above studies found irrefutable evidence that bettors believe in the hot hand and that market odds are biased in accordance with this belief. If there is a hot hand effect in markets, it generally is so small as to be economically insignificant.

5 Conclusion

This chapter has shown that while many biases in decision-making have been demonstrated in laboratory-based studies, there are numerous reasons for suggesting that these findings may not translate to the real world. Betting markets provide a valuable naturalistic setting in which to explore biased decision-making because participants are making decisions in a situation that is more representative of the environments in which day-to-day decisions are made. We have argued that bettors display significantly less biased judgments in their natural domain than do naive participants in laboratory experiments. To support this view we have cited a number of examples related to rationality and calibration of subjective probability judgments. Furthermore, we have shown that there is only mixed evidence that bettors anchor their judgments on available information, engage in herding behavior, or believe in the hot hand or gambler's fallacies. Even the FLB, which has been the focus of the majority of research in betting markets, is no longer observable in some markets.

The primary conclusion of this chapter is that while systematic biases reported in the laboratory have been found in naturalistic betting markets, the extent and generality of these biases in these real-world environments are often significantly less. The context of the decision task, the incentives offered, the lack of scrutiny involved, and the experience of the decision-makers all contribute to an explanation for this conclusion. Another consideration is the importance of aggregation. It is costly and ethically challenging to obtain betting market datasets from which it is possible to discern individual biases. In a more typical dataset individual biases may be eliminated by aggregation of the opinions of a diverse range of bettors. Moreover, even a systematic bias that is attributed to a large portion of the betting population can be reduced by the unbiased actions of a wealthy few, as there is always a strong economic motivation to capitalize on the biases of others.

A drawback of the heuristics and biases approach to decision-making in general is highlighted by our discussion of the hot hand and gambler's fallacies. There is the initial problem of explaining two apparently opposite biases with the same heuristic, though subsequent research has clarified that there are two separate situations when people use either of these fallacies. On the other hand, it can be impossible to narrow down multiple explanations for one bias to the single, most-valid explanation. Thus a wide range of explanations has been proposed for the FLB. Similarly the hot hand fallacy could be explained by the illusion of control or by the representativeness heuristic or by extrapolation of genuine hot hand effects. As Willem Wagenaar (1988, 115–116) has argued, the heuristics and biases approach

> does not specify rules telling us which heuristic will be applied in a given situation. Even worse, from the individual differences among gamblers, it is obvious that several heuristics could be chosen in one and the same situation, and that these heuristics lead to opposite behaviors . . .. There are so many heuristics, that it will be virtually impossible to find behaviors that cannot be accounted for.

Hence while there is some evidence of biased behavior in betting markets, explaining its prevalence is another matter altogether.

There are additional issues associated with betting market research that may lead one to question the generalizability of the conclusions drawn from such studies. For example, bettors may be unrepresentative of the wider public since they are predominantly older males (Dipboye and Flanagan 1979), and there may be some self-selection effects (indeed, it is not obvious as to why some people gamble and some do not; see Rachlin 1990). We must also retain some skepticism about generalizability from betting markets to other economic settings (Levitt and List 2007). Just as laboratory research should recognize that generalizability of findings is limited, additional research on biased decision-making in betting markets should acknowledge that laboratory experimentation is often the first available evidence that heuristics are being employed or biased outcomes are occurring. Without either the theoretical background or the controlled elegance of laboratory research, naturalistic research might be confounded by the vast array of potential variables involved and the often unintuitive nature of real-world decision-making. The way forward appears to be a tandem approach with betting market studies being informed by results from laboratory experiments and the latter being designed to examine the causes of phenomena that the former highlight. In this manner the true nature and real-world characteristics of behavioral biases may be revealed.

Notes

1. As a further example of the importance of context in decision-making, consider the following problem. There are four cards on the table, each with a letter on one side and a number on the other. The rule is, "If there is a vowel on one side of a card, then there is an even number on the other side." The cards show A, D, 4, and 7. Which cards must be turned over in order to determine whether the rule is true or false? This is known as Wason's four-card selection task (Wason 1968), and usually less than 10 percent of people respond with the correct answer of A and 7 (most neglect to choose 7 or unnecessarily include 4). However, when this problem is reframed in terms of certain social contexts, such as asking subjects to test the rule "If a person is over 18, they can drink alcohol" and replacing the cards with "16 years old," "22 years old," "Coke," and "beer," the correct answer ("16 years old" and "beer") is given by most respondents even though the problem is logically identical to the first, more abstract, task (e.g., Cox and Griggs 1982).
2. This is a particular advantage of betting markets over other types of financial market for naturalistic research. The payoffs in betting markets are entirely unambiguous, so there is a time when all uncertainty is resolved. This is not the case in regular financial markets, where prices continuously represent the current expectation of future prices.
3. An exception holds for objective probabilities of less than 0.05, which is the favorite-longshot bias detailed in the second part of this chapter.
4. Exceptions have been reported in the horse race betting markets in Hong Kong (Busche and Hall 1988; Busche 1994), the market at one U.S. racetrack (Swidler and Shaw 1995), and exchange betting markets in the United Kingdom (Smith, Paton, and Vaughan Williams 2006).
5. In financial markets the results can be catastrophic, with herd behavior exacerbating asset-price bubbles and crashes, and bank runs (Devenow and Welch 1996).
6. Evidence from financial markets is similarly inconclusive (Sias 2004).
7. Croson and Sundali also found evidence of the hot hand fallacy: 80 percent of bettors quit playing after losing a bet while only 20 percent quit after winning. Moreover, bettors tended to place more bets after winning than after losing.

8. The point spread market is a betting market in which a bet wins if the home team wins by a specified margin of points (the point spread) or, if the point spread is negative, the home team loses by less than the point spread (this is known as the team winning "against the spread").
9. There is a fourth alternative—that bettors are unaware of a genuine hot hand effect—but this hypothesis is not tested by Brown and Sauer.

References

Ali, Mukhtar M. 1977. Probability and utility estimates for racetrack bettors. *Journal of Political Economy* 85(4):803–815.

Anderson, G., and R. I. F. Brown. 1984. Real and laboratory gambling, sensation-seeking and arousal. *British Journal of Psychology* 75(3):401–410.

Asch, Peter, Burton G. Malkiel, and Richard E. Quandt. 1982. Racetrack betting and informed behavior. *Journal of Financial Economics* 10(2):187–194.

Avery, Christopher, and Judith Chevalier. 1999. Identifying investor sentiment from price paths: The case of football betting. *Journal of Business* 72(4):493–521.

Avery, Christopher, and Peter Zemsky. 1998. Multidimensional uncertainty and herd behavior in financial markets. *American Economic Review* 88(4):724–748.

Ayton, Peter, and Ilan Fischer. 2004. The hot hand fallacy and the gambler's fallacy: two faces of subjective randomness? *Memory & Cognition* 32(8), 1369–1378.

Ayton, Peter, and George Wright. 1994. Subjective probability: What should we believe? In *Subjective probability*, edited by George Wright and Peter Ayton. Chichester, U.K.: Wiley, 163–183.

Banerjee, Abhijit V. 1992. A simple model of herd behavior. *Quarterly Journal of Economics* 107(3):797–817.

Bar-Eli, Michael, Simcha Avugos, and Markus Raab. (2006). Twenty years of "hot hand" research: Review and critique. *Psychology of Sport and Exercise* 7(6):525–553.

Benter, William. 1994. Computer based horse race handicapping and wagering systems: A report. In *Efficiency of racetrack betting markets*, edited by Donald B. Hausch, Victor S. Y. Lo, and William T. Ziemba. San Diego, Calif.: Academic Press, 183–198.

Bhattacharjee, Sudip, and Kimberly Moreno. 2002. The impact of affective information on the professional judgments of more experienced and less experienced auditors. *Journal of Behavioral Decision Making* 15(4):361–377.

Bikhchandani, Sushil, David Hirshleifer, and Ivo Welch. 1992. A theory of fads, fashion, custom, and cultural change as informational cascades. *Journal of Political Economy* 100(5):992–1026.

Bradley, Ian. 2003. The representative bettor, bet size, and prospect theory. *Economics Letters* 78(3):409–413.

Brown, Lawrence D., Rebecca D'Amato, and Randy Gertner. 1994. Racetrack betting: Do bettors understand the odds? *Chance* 7(3):17–23.

Brown, William O., and Raymond D. Sauer. 1993. Does the basketball market believe in the hot hand? Comment. *American Economic Review* 83(5):1377–1386.

Bruce, Alistair C., and Johnnie E. V. Johnson. 1992. Toward an explanation of betting as a leisure pursuit. *Leisure Studies* 11(3):201–218.

——. 2003. Investigating betting behavior: A critical discussion of alternative methodological approaches. In *The economics of gambling*, edited by Leighton Vaughan Williams. London: Routledge, 224–246.
Burns, Bruce D., and Bryan Corpus. 2004. Randomness and inductions from streaks: "Gambler's fallacy" versus "hot hand." *Psychonomic Bulletin & Review* 11(1):179–184.
Burns, Penny. 1985. Experience and decision making: a comparison of students and businessmen in a simulated progressive auction. In *Research in Experimental Economics*, vol. 3, edited by Vernon L. Smith. Greenwich, Conn.: JAI Press, 139–157.
Busche, Kelly. 1994. Efficient market results in an Asian setting. In *Efficiency of racetrack betting markets*, edited by Donald B. Hausch, Victor S. Y. Lo, and William T. Ziemba. London: Academic Press, 615–616.
Busche, Kelly, and Christopher D. Hall. 1988. An exception to the risk preference anomaly. *Journal of Business* 61(3):337–346.
Camerer, Colin F. 1987. Do biases in probability judgment matter in markets? Experimental evidence. *American Economic Review* 77(5):981–997.
—— 1989. Does the basketball market believe in the "hot hand"? *American Economic Review* 79(5):1257–1261.
—— 1998. Can asset markets be manipulated? A field experiment with racetrack betting. *Journal of Political Economy* 106(3):457–482.
—— 2001. Prospect theory in the wild: Evidence from the field. In *Choices, values, and frames*, edited by Daniel Kahneman, and Amos Tversky. New York: Cambridge University Press, 288–300.
Ceci, Stephen J., and Jeffrey K. Liker. 1986. A day at the races: A study of IQ, expertise, and cognitive complexity. *Journal of Experimental Psychology: General* 115(3):255–266.
Clotfelter, Charles T., and Philip J. Cook. 1993. The "gambler's fallacy" in lottery play. *Management Science* 39(12):1521–1525.
Cox, James R., and Richard A. Griggs. 1982. The effects of experience on performance in Wason's selection task. *Memory & Cognition* 10(5):496–502.
Crafts, N. F. R. 1985. Some evidence of insider knowledge in horse race betting in Britain. *Economica* 52(207):295–304.
Croson, Rachel, and James Sundali. 2005. The gambler's fallacy and the hot hand: Empirical data from casinos. *Journal of Risk and Uncertainty* 30(3):195–209.
Devenow, Andrea, and Ivo Welch. 1996. Rational herding in financial economics. *European Economic Review* 40(3–5):603–615.
Dipboye, Robert L., and Michael F. Flanagan. 1979. Research settings in industrial and organizational psychology: Are findings in the field more generalizable than in the laboratory? *American Psychologist* 34(2):141–150.
Dowie, Jack. 1976. On the efficiency and equity of betting markets. *Economica* 43(170): 139–150.
Durham, Gregory R., Michael G. Hertzel, and J. Spencer Martin. 2005. The market impact of trends and sequences in performance: New evidence. *Journal of Finance* 60(5): 2551–2569.
Ebbesen, Ebbe B., and Vladimir J. Konečni. 1980. On the external validity of decision making research: what do we know about decisions in the real world? In *Cognitive processes in choice and decision behavior*, edited by Thomas S. Wallsten.. Hillsdale, N.J.: Erlbaum, 21–45.
Falk, Ruma, and Clifford Konold. 1997. Making sense of randomness: Implicit encoding as a basis for judgment. *Psychological Review* 104(2):301–318.

Ferrell, William R. 1994. Calibration of sensory and cognitive judgments: a single model for both. *Scandinavian Journal of Psychology* 35(4):297–314.

Festinger, Léon. 1953. Laboratory experiments. In *Research methods in the behavioral sciences*, edited by Léon Festinger, and Daniel Katz. New York: Holt, Rinehart & Winston, 136–172.

Figlewski, Stephen. 1979. Subjective information and market efficiency in a betting market. *Journal of Political Economy* 87(1):75–88.

Funder, David C. 1987. Errors and mistakes: Evaluating the accuracy of social judgment. *Psychological Bulletin* 101(1):75–90.

Gandar, John, Richard Zuber, and R. Stafford Johnson. 2001. Searching for the favourite-longshot bias down under: An examination of the New Zealand pari-mutuel betting market. *Applied Economics* 33(13):1621–1629.

Gigerenzer, Gerd. 2000. *Adaptive thinking: Rationality in the real world.* New York: Oxford University Press.

Gilovich, Thomas, Robert Vallone, and Amos Tversky. 1985. The hot hand in basketball: On the misperception of random sequences. *Cognitive Psychology* 17(3):295–314.

Golec, Joseph, and Maurry Tamarkin. 1995. Do bettors prefer longshots because they are risk-lovers, or are they just overconfident? *Journal of Risk and Uncertainty* 11(1):51–64.

———. 1998. Bettors love skewness, not risk, at the horse track. *Journal of Political Economy* 106(1):205–225.

Goodman, Jodi S. 1998. The interactive effects of task and external feedback on practice performance and learning. *Organizational Behavior and Human Decision Processes* 76(3):223–252.

Grant, Andrew, and David Johnstone. 2010. Finding profitable forecast combinations using probability scoring rules. *International Journal of Forecasting* 26(3):498–510.

Griffith, R. M. 1949. Odds adjustments by American horse-race bettors. *American Journal of Psychology* 62(2):290–294.

Hamid, Shahid S., Arun J. Prakash, and Michael W. Smyser. 1996. Marginal risk aversion and preferences in a betting market. *Applied Economics* 28(3):371–376.

Harrison, Glenn W., and John A. List. 2004. Field experiments. *Journal of Economic Literature* 42(4):1009–1055.

Henery, Robert J. 1985. On the average probability of losing bets on horses with given starting price odds. *Journal of the Royal Statistical Society: Series A (General)* 148(4):342–349.

Hoerl, Arthur E., and Herbert K. Fallin. 1974. Reliability of subjective evaluations in a high incentive situation. *Journal of the Royal Statistical Society: Series A (General)* 137(2):227–230.

Hogarth, Robin M. 1981. Beyond discrete biases: Functional and dysfunctional aspects of judgmental heuristics. *Psychological Bulletin* 90(2):197–217.

Hurley, William, and Lawrence McDonough. 1995. A note on the Hayek hypothesis and the favorite-longshot bias in parimutuel betting. *American Economic Review* 85(4):949–955.

Johnson, Johnnie E. V., and Alistair C. Bruce. 1993. Gluck's second law: An empirical investigation of horserace betting in early and late races. *Psychological Reports* 72(2):1251–1258.

———. 1998. Risk strategy under task complexity: A multivariate analysis of behavior in a naturalistic setting. *Journal of Behavioral Decision Making* 11(1):1–17.

———. 2001. Calibration of subjective probability judgments in a naturalistic setting. *Organizational Behavior and Human Decision Processes* 85(2):265–290.

Johnson, Johnnie E. V., Raymond O'Brien, and Ming-Chien Sung. 2010. Assessing bettors' ability to process dynamic information: Policy implications. *Southern Economic Journal* 76(4):906–931.

Johnson, Johnnie E. V., Adi Schnytzer, and Shuang Liu. 2009. To what extent do investors in a financial market anchor their judgments excessively? Evidence from the Hong Kong horserace betting market. *Journal of Behavioral Decision Making* 22(4):410–434.

Jullien, Bruno, and Bernard Salanié. 2000. Estimating preferences under risk: The case of racetrack bettors. *Journal of Political Economy* 108(3):503–530.

——. 2008. Empirical evidence on the preferences of racetrack bettors. In *Handbook of sports and lottery markets*, edited by Donald B. Hausch and William T. Ziemba. Amsterdam: Elsevier, 27–49.

Kahneman, Daniel, and Amos Tversky. 1979. Prospect theory: An analysis of decision under risk. *Econometrica* 47(2):263–291.

Kahneman, Daniel, Paul Slovic, and Amos Tversky, eds. 1982. *Judgment under uncertainty: Heuristics and biases.* Cambridge: Cambridge University Press.

Keren, Gideon B., and Willem A. Wagenaar. 1985. On the psychology of playing blackjack: Normative and descriptive considerations with implications for decision theory. *Journal of Experimental Psychology: General* 114(2):133–158.

——. 1987. Violation of utility theory in unique and repeated gambles. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 13(3):387–391.

Koehler, Jonathan J. 1996. The base rate fallacy reconsidered: descriptive, normative, and methodological challenges. *Behavioral and Brain Sciences* 19(1):1–17.

Kopelman, Richard E., and Betsy L. Minkin. 1991. Toward a psychology of parimutuel behavior: Test of Gluck's Laws. *Psychological Reports* 68(2):701–702.

Langer, Ellen J. 1975. The illusion of control. *Journal of Personality and Social Psychology* 32(2):311–328.

Laplace, Pierre Simon. [1825] 1995. *A philosophical essay on probabilities.* Reprint (translated from the fifth French edition of 1825 by A. I. Dale), New York: Springer-Verlag.

Larkin, Jill, John McDermott, Dorothea P. Simon, and Herbert A. Simon. 1980. Expert and novice performance in solving physics problems. *Science 208*(4450), 1335–1342.

Law, David, and David A. Peel. 2002. Insider trading, herding behavior and market plungers in the British horse-race betting market. *Economica* 69(274):327–338.

Lehrer, Jonah. 2009. *The Decisive Moment: How the brain makes up its mind.* Edinburgh: Canongate.

Levitt, Steven D., and John A. List. 2007. What do laboratory experiments measuring social preferences reveal about the real world? *Journal of Economic Perspectives* 21(2):153–174.

Lichtenstein, Sarah, Paul Slovic, Baruch Fischhoff, Mark Layman, and Barbara Combs. 1974. Judged frequency of lethal events. *Journal of Experimental Psychology: Human Learning and Memory* 4(6):551–578.

Liu, Shuang, and Johnnie E. V. Johnson. 2007. Searching for anchoring effects in a naturalistic environment: evidence from the Hong Kong horserace betting market. *Journal of Gambling Business and Economics* 1(1):69–84.

McGlothlin, William H. 1956. Stability of choices among uncertain alternatives. *American Journal of Psychology* 69(4):604–615.

Metzger, Mary Ann. 1985. Biases in betting: An application of laboratory findings. *Psychological Reports* 56(3):883–888.

Murphy, Allan H., and Barbara G. Brown. 1984. A comparative evaluation of objective and subjective weather forecasts in the United States. *Journal of Forecasting* 3(4):369–393.

Northcraft, Gregory B., and Margaret A. Neale. 1987. Experts, amateurs, and real estate: an anchoring-and-adjustment perspective on property pricing decisions. *Organizational Behavior and Human Decision Processes* 39(1):84–97.

Oorlog, Dale R. 1995. Serial correlation in the wagering market for professional basketball. *Quarterly Journal of Business and Economics* 34(2):96–109.

Orasanu, Judith, and Terry Connolly. 1993. The reinvention of decision making. In *Decision making in action: Models and methods*, edited by Gary A. Klein, Judith Orasanu, Roberta Calderwood, and Caroline E. Zsambok. Chichester, U.K.: Wiley, 3–20.

Ottaviani, Marco, and Peter N. Sørensen. 2008. The favorite-longshot bias: an overview of the main explanations. In *Handbook of sports and lottery markets*, edited by Donald B. Hausch and William T. Ziemba. Amsterdam: Elsevier, 83–101.

Papachristou, George. 2004. The British gambler's fallacy. *Applied Economics* 36(18):2073–2077.

Paul, Rodney J., and Andrew P. Weinbach. 2005. Bettor misperceptions in the NBA: The overbetting of large favorites and the "hot hand." *Journal of Sports Economics* 6(4):390–400.

Piron, Robert, and L. Ray Smith. 1994. Testing risklove in an experimental racetrack. *Journal of Economic Behavior & Organization* 27(3):465–474.

Preston, Malcolm G., and Philip Baratta. 1948. An experimental study of the auction-value of an uncertain outcome. *American Journal of Psychology* 61(2):183–193.

Quandt, Richard E. 1986. Betting and equilibrium. *Quarterly Journal of Economics* 101(1): 201–208.

Rabin, Matthew. 2002. Inference by believers in the law of small numbers. *Quarterly Journal of Economics* 117(3):775–816.

Rachlin, Howard. 1990. Why do people gamble and keep gambling despite heavy losses? *Psychological Science* 1(5):294–297.

Rosett, Richard N. 1965. Gambling and rationality. *Journal of Political Economy* 73(6):595–607.

——. 1971. Weak experimental verification of the expected utility hypothesis. *Review of Economic Studies* 38(116):481–492.

Sauer, Raymond D. 1998. The economics of wagering markets. *Journal of Economic Literature* 36(4):2021–2064.

Saunders, D. M., and D. E. Turner. 1987. Gambling and leisure: The case of racing. *Leisure Studies* 6(3):281–299.

Schnytzer, Adi, and Avichai Snir. 2008. Herding in imperfect betting markets with insider traders. *Journal of Gambling Business and Economics* 2(2):1–15.

Shanteau, James. 1992. Competence in experts: The role of task characteristics. *Organizational Behavior and Human Decision Processes* 53(2):252–266.

Shin, Hyun Song. 1991. Optimal betting odds against insider traders. *Economic Journal* 101(408):1179–1185.

Sias, Richard W. 2004. Institutional herding. *Review of Financial Studies* 17(1):165–206.

Simon, Herbert A. 1955. A behavioral model of rational choice. *Quarterly Journal of Economics* 69(1):99–118.

Smith, Michael A., and Leighton Vaughan Williams. 2010. Forecasting horse race outcomes: New evidence on odds bias in UK betting markets. *International Journal of Forecasting*, 26(3):543–550.

Smith, Michael A., David Paton, and Leighton Vaughan Williams. 2006. Market efficiency in person-to-person betting. *Economica* 73(292):673–689.

Snowberg, Erik, and Justin Wolfers. 2010. Explaining the favorite-longshot bias: Is it risk-love or misperceptions? *Journal of Political Economy* 118(4):723–746.

Sobel, Russell S., and S. Travis Raines. 2003. An examination of the empirical derivatives of the favourite-longshot bias in racetrack betting. *Applied Economics* 35(4):371–385.

Spiwoks, Markus, Kilian Bizer, and Oliver Hein. 2008. Informational cascades: A mirage? *Journal of Economic Behavior & Organization* 67(1):193–199.

Swidler, Steve, and Ron Shaw. 1995. Racetrack wagering and the "uninformed" bettor: A study of market efficiency. *Quarterly Review of Economics and Finance* 35(3):305–314.

Terrell, Dek. 1994. A test of the gambler's fallacy: Evidence from pari-mutuel games. *Journal of Risk and Uncertainty* 8(3):309–317.

——. 1998. Biases in assessments of probabilities: New evidence from greyhound races. *Journal of Risk and Uncertainty* 17(2):151–167.

Terrell, Dek, and Amy Farmer. 1996. Optimal betting and efficiency in parimutuel betting markets with information costs. *Economic Journal* 106(437):846–868.

Thaler, Richard H., and Eric J. Johnson. 1990. Gambling with the house money and trying to break even: The effects of prior outcomes on risky choice. *Management Science* 36(6): 643–660.

Thaler, Richard H., and William T. Ziemba. 1988. Parimutuel betting markets: Racetracks and lotteries. *Journal of Economic Perspectives* 2(2):161–174.

Tuckwell, R. H. 1983. The thoroughbred gambling market: Efficiency, equity and related issues. *Australian Economic Papers* 22(40):106–118.

Tversky, Amos, and Daniel Kahneman. 1971. Belief in the law of small numbers. *Psychological Bulletin* 76(2):105–110.

——. 1974. Judgment under uncertainty: heuristics and biases. *Science* 185(4157):1124–1131.

——. 1992. Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty 5*(4):297–323.

Vaughan Williams, Leighton. 1999. Information efficiency in betting markets: A survey. *Bulletin of Economic Research* 51(1):1–30.

Vaughan Williams, Leighton, and David Paton. 1997. Why is there a favorite-longshot bias in British racetrack betting markets? *Economic Journal* 107(440):150–158.

Vlastakis, Nikolaos, George Dotsis, and Raphael N. Markellos. 2009. How efficient is the European football betting market? Evidence from arbitrage and trading strategies. *Journal of Forecasting* 28(5):426–444.

Wagenaar, Willem A. 1988.*Behavior Paradoxes of gambling behavior.* London: Lawrence Erlbaum Associates.

Wagenaar, Willem A., and Gideon B. Keren. 1988. Chance and luck are not the same. *Journal of Behavioral Decision Making* 1(2):65–75.

Wallsten, Thomas S., David V. Budescu, Ivo Erev, and Adele Diederich. 1997. Evaluating and combining subjective probability estimates. *Journal of Behavioral Decision Making* 10(3):243–268.

Wason, P. C. 1968. Reasoning about a rule. *Quarterly Journal of Experimental Psychology* 20(3):273–281.

Weitzman, Martin. 1965. Utility analysis and group behavior: An empirical study. *Journal of Political Economy* 73(1):18–26.

Wilson, Timothy D., Christopher E. Houston, and Kathryn M. Etling. 1996. A new look at anchoring effects: Basic anchoring and its antecedents. *Journal of Experimental Psychology: General* 125(4):387–402.

Yaari, Menahem E. 1965. Convexity in the theory of choice under risk. *Quarterly Journal of Economics 79*(2):278–290.

Yates, J. Frank, ed. 1992. *Risk Taking Behavior.* Chichester, U.K.: John Wiley.

CHAPTER 27

BEHAVIORAL FINANCE AND POINT SPREAD WAGERING MARKETS

GREG DURHAM

> [T]HE point-spread betting market may be a fruitful place [in which] to conduct research about behavioral theories that could apply in conventional market settings.
>
> — John, Rick, Thomas, and Ben (1988)

1 INTRODUCTION

THE objective of this chapter is to review and extend an emerging body of research that tests theories, models, and general predictions of human behavior—most of which are rooted in psychology—while using the point spread wagering market as the setting. The bridge between behavioral psychology and sports betting is provided by the growing field of behavioral finance, a relatively new subdiscipline of behavioral economics. Behavioral psychology is a component of behavioral finance, and financial markets and the point spread market are strikingly similar. As will be explained in section 4, a clear settling-up point associated with each wager makes the point spread market a powerful, particularly simple setting in which to test theories of investor behavior. Furthermore, because of numerous similarities between the point spread market and financial markets, any findings from the betting market should have useful extensions for wider financial audiences.

This chapter proceeds with discussions of the concept of rational behavior, of the two general levels of behavioral-finance research (individual investor level and market level), and of the conditions that must exist in order for irrational behavior to affect asset prices. Section 3 documents the evolution from classical financial economics to the current theoretical environment, which is characterized by a much greater openness to the possibility—or reality—that individuals may not always act in rational manners and may not always act solely to maximize expected utility of wealth. This section also summarizes a few select cognitive biases and types of human sentiment that seem to regularly affect bettor behavior in point spread markets.

Section 4 addresses the mechanics of point spread wagering and touts the usefulness and advantages of point spread betting markets as research laboratories. Section 5 presents a thorough review of academic studies that have tested for evidence of rational (or irrational) behavior in point spread wagering markets. This survey is restricted to point spread betting (as opposed to odds betting or pari-mutuel betting), since point spread markets are most similar to the more conventional markets wherein financial assets are traded. Section 6 concludes and offers suggested directions for future research.

2 Rationality, Irrationality, and Behavioral-Finance Research

Behavioral finance encompasses traditional finance, psychology, and perhaps even sociology. Behavioral finance represents a divergence from the classical assumptions of traditional finance in that, unlike the latter discipline, it allows for the legitimate possibility that various agents in the financial marketplace do not always act in rational, unbiased, utility-maximizing ways (where utility is defined in the traditional terms of wealth and risk). In the specific context of investing rational behavior involves (i) processing new information and updating beliefs in proper Bayesian fashion and (ii) making decisions that are consistent with the classically presumed investor's objective function of maximizing expected utility in mean-variance fashion. Irrationality is any behavior that is inconsistent with either, or both, of these two complementary specifications of rational behavior.

Some studies of investor behavior examine investing activity at the individual level, either in experimental settings or by using data from individual brokerage accounts. Robert Bloomfield and Jeffrey Hales (2002), for example, found that human subjects (MBA students) respond to historical stock-performance sequences in ways consistent with an investor-behavior model developed by Nicholas Barberis, Andrei Shleifer, and Robert Vishny (1998).[1] In a study that uses proprietary data from household stock-trading accounts, Brad Barber and Terrance Odean (1999) found evidence of excessive trading (suggesting that investors are overconfident) and of holding losing stocks too long (suggesting a desire to avoid the feeling of regret). Examining data from the same proprietary source, Barber and Odean (2001) also found that men and women both trade too frequently—men more so—at a cost of lower returns than what would have been realized using simple buy-and-hold strategies.

However, most studies look at market-level data (stock prices and returns) in testing for evidence consistent with irrational behavior. For example, Werner DeBondt and Richard Thaler (1985) found long-run reversals in stock returns, suggesting that investors overreact to firms that perform well, thereby driving these firms' stock prices to artificially high levels only to have prices drop to normal levels after the long-term run-ups.[2] Narasimhan Jegadeesh and Sheridan Titman (1993) found that stock returns are positively autocorrelated (i.e., they exhibit momentum) over shorter horizons, suggesting that investors underreact to information and that information thus slowly incorporates into prices. As a final example, Andrea Frazzini (2006) found positive post-announcement drift in stock returns following announcements of good news, suggesting that happily disposed individuals sell their stocks immediately following the good news, thereby putting instant downward pressure on prices and leading to the post-announcement drift.

These three noted studies are representative examples among countless others that utilize market prices in investigations of investor behavior. If investors depart from rational behavior in predictable, systematic ways and if any pricing errors caused by these departures are not fully corrected by rational traders, then researchers should be able to detect the effects that various types of irrational behavior have on asset prices. The challenge, unfortunately, is that often more than one behavioral bias can explain an observed pricing anomaly, and, therefore, distinguishing among multiple explanations is difficult. Even when rationality can be rejected, the source of departure from rationality may not be immediately apparent.

Implications of irrational behavior will only appear at aggregate price levels if limits to arbitrage are sufficiently large so as to prevent rational traders from fully correcting any mispricing that emerges due to irrationality. Brad DeLong, Shleifer, Larry Summers, and Robert Waldmann (1990) explained that arbitrageurs face a short-run risk that irrational traders might cause prices to deviate further away from fundamental values instead of the arbitrageurs transacting to eliminate any pricing inefficiencies. If an arbitrageur were forced to liquidate its position before prices are corrected, the arbitrageur's round-trip transaction would yield negative profits. Shleifer and Vishny (1997) proposed the possibility that arbitrageurs may face capital constraints that prevent the arbitrageurs from fully offsetting irrational trading behavior's effects on prices. The large body of empirical research showing asset-pricing anomalies strongly implies that arbitrage is, indeed, limited in many cases.

For at least half a century, psychologists have been documenting human behaviors that are seemingly irrational; the irrationality is often attributable to predictable biases that have been shown to systematically impair the human decision-making process. The list of cognitive biases is long and broad, and includes conservatism, reliance on the representativeness heuristic, overconfidence, the illusion of knowledge, biased self-attribution, aversion to regret, and aversion to losses, among others. More recently the more general concept of sentiment has also worked its way into the asset-pricing and efficient-market literature: many of the types of sentiment that are hypothesized to affect investors in the financial marketplace derive from the cognitive biases just mentioned. Academicians are attuned to these various psychological impairments and are examining the ways by which they affect investor behavior as well as the effects that any irrational behavior has on asset prices and returns.

3 Behavioral Biases and Irrational Preferences in Betting

As already noted, the field of behavioral finance permits human decision-making behavior that violates the principles of either Bayes' theorem, the maximization of expected utility, or both. Behavioral finance researchers embrace the possibility that an individual's ability to behave rationally—in formulating beliefs or in maximizing utility—is often impaired by behavioral biases or by irrational preferences. This section begins with a recap of the evolution of alternative utility functions and behavioral models for a representative investor who might behave normally, where the definition of *normal* has itself evolved over time, perhaps to where it means rational most of the time and irrational sometimes. The section continues with a brief overview of a handful of cognitive biases that seem to repeatedly affect bettor behavior in point spread wagering markets and concludes with a discussion of various types of sentiment that have been hypothesized to exist in point spread markets.

3.1 Departures from the Expected Utility Framework

John Von Neumann and Oskar Morgenstern (1944) (hereafter VNM) developed an expected-utility-of-wealth function that accurately represents two axioms of rational preferences and choice (namely, completeness and transitivity) as well as two axioms of choice under uncertainty (continuity and independence). The VNM expected utility function relies on an underlying, continuous utility of wealth function that is increasing in wealth and reflects diminishing marginal utility of wealth. (This underlying utility function is often referred to as a Bernoulli utility function.) The concavity of the Bernoulli function captures the risk aversion of the representative individual who derives utility of wealth.

The decreasing marginal utility of wealth feature of the Bernoulli function is in conflict with the acceptance of a fair gamble and in even greater conflict with acceptance of a unfair gamble. In response to this limitation of the VNM–Bernoulli model, a succession of models of expected utility of wealth followed. One prominent model that emerged is Harry Markowitz's (1952) utility function that is primarily concave but locally convex ("locally" meaning at or near an individual's current wealth level); the Markowitz model can accurately explain an expected-utility maximizer's preference for gambles. Until the late 1970s all of the hypothesized versions of expected utility continued to adhere to the underlying assumption of rationality, still defined in terms of the completeness and transitivity axioms.

Prospect theory, developed by Daniel Kahneman and Amos Tversky (1979), represents the first and most enduring diversion away from rationality. This theory posits a utility function with a reference point relative to which gains and losses are then measured, in contrast to the Bernoulli function which begins at a wealth level of zero. Furthermore, the prospect theory value function is S-shaped: it is concave over the range of gains and convex over the range of losses. The other main departure of prospect theory from VNM utility theory is that the decision-maker is modeled as using decision weights instead of probabilities in calculating expected value. These decision weights are increasing in, but do not represent estimates of, corresponding probabilities. Also, the decision weights are higher (lower) than the true probabilities attached to extreme outcomes with very low (high) probabilities.

Kahneman and Tversky proposed a two-step decision process. The first step is what they call the "editing phase," wherein the decision-maker evaluates a set of probabilistic outcomes (i.e., a gamble) with the goal of organizing and formulating the alternatives so as to simplify the subsequent comparison across, and eventual choice from among, the alternatives. The editing phase involves the establishment of the aforementioned reference point, around which the S-shaped value function will be centered and relative to which various possible outcomes will be coded as gains and losses. The second step is the "evaluation stage," which involves the assignment of the above-mentioned decision weights as well as subjective values (defined relative to the already established reference point) to the various possible outcomes. To conclude the evaluation phase, the decision-maker selects the prospect with the highest expected value.

Human behavior prescribed by prospect theory can often result in departures from expected-utility theory, including violations of the transitivity property or of the principle of dominance, but the theory has remained robust in explaining so many of the empirical and experimental findings that are in conflict with the previous rationality-based utility functions.

A second theoretical framework that also allows for irrational behavior and is specifically developed with investors in mind is the one advanced by Hersh Shefrin and Meir Statman (1984). In response to widely observed tendencies of investors to prefer cash dividends, to sell winner stocks too quickly, and to hold "loser" stocks too long, Shefrin and Statman developed a framework that uses prospect theory as one of its four fundamental tenets along with the additional behavioral concepts of mental accounting, aversion to regret (and its converse, desire for pride), and self-control. Three additional models of investor behavior will be introduced shortly, each within the respective subsection dedicated to the cognitive biases on which it is based.

Having now established prospect theory as a solid foundation for explaining many types of irrational behavior, with the Shefrin and Statman framework as a viable offshoot, the discussion proceeds to brief overviews of cognitive biases that are rooted in prospect theory and are particularly relevant to point spread bettors.

3.2 Conservatism and the Representativeness Heuristic

First proposed by Ward Edwards (1968), conservatism is defined as a slowness to fully revise beliefs in the presence of new information. When individuals are impaired by conservatism while updating expectations, they underweigh relevant new information and over-rely on older data, as compared to how a purely rational Bayesian would respond. Individuals who exhibit the conservatism bias seem unwilling, or are unable, to have full faith in the newest information. In a markets context the conservatism bias is suggestive of investor underreaction.

The representativeness heuristic, developed by Kahneman and Tversky (1972, 430), is an evaluation mechanism under which an individual assesses the probability of an uncertain event by the degree to which it "(i) is similar in its essential properties to the presumed parent population [and] (ii) reflects the salient features of the process by which it is generated." A person distracted by reliance on the representativeness heuristic[3] will not formulate probability estimates in Bayesian fashion, instead overweighting the representative description and underweighting any statistical evidence, such as sample size.

Barberis, Shleifer, and Vishny (1998) (hereafter BSV) developed a theoretical model of investor behavior involving a representative agent who believes that, at any instant, stock price performance is being governed by one of two regimes when in fact it follows a random process. With each new event (either a continuation or reversal in performance), the agent updates beliefs in proper Bayesian fashion about which regime is in place: a steadily trending regime or a mean-reverting regime. Belief in a continuation regime generates the same behavior as does the representativeness heuristic; belief in a reversal regime creates the same effects as those predicted by conservatism. The BSV model is consistent with two pervasive empirical anomalies that have been the focus of much research in finance: short-run momentum and long-run reversals in stock returns.

3.2.1 *Overreaction*

Experimental research in psychology shows that individuals react to new information more excessively than is appropriate, even more so when the new news is striking. This cognitive error, known as overreaction, follows closely from the representativeness bias. For the purposes of this survey chapter, overreaction applies specifically to the cognitive exercise of updating probability estimates in the face of new information and given an already existing information set. The mere name of this bias necessitates a benchmark for defining what an appropriate reaction would be. Whereas the appropriate reaction (or revision in beliefs) would be that which is dictated by Bayes's rule, an overreacting decision-maker will overestimate the statistical importance of the piece of new information.

Dale Griffin and Tversky (1992) (hereafter G&T) hypothesized that in the face of new evidence about the likelihood of a future event or about the characteristics of an unobservable total population, individuals over-rely on the "strength" (or saliency) of the evidence and under-rely on the same evidence's "weight" (or statistical informativeness). One implication, then, of G&T's hypothesis is that when evidence has high strength and low weight, people overreact in a manner consistent with representativeness. A second implication within this same framework is that conservatism would occur in the presence of low-strength, high-weight evidence. People are unimpressed by the low strength and react mildly to the evidence, lesser than what Bayesian updating would suggest.

Manifestations of overreaction are numerous in financial markets. Its market-level effects have emerged in various forms, including excess volatility in stock prices (as shown by Robert Shiller (1981) or positive cumulative abnormal returns on portfolios of stocks that were previously losers (as documented by DeBondt and Thaler (1985)). Another example is the prevalence of unjustifiably high stock prices for companies with recently high sales growth (or with high price-to-earnings ratios) relative to prices on these stocks' counterparts, known as value stocks (as documented by Josef Lakonishok, Shleifer, and Vishny (1994)). In another study of investor reactions, Barberis, Shleifer, and Vishny (1998) found evidence in support of Griffin and Tversky (1992): investors overreact to high-strength, low-weight earnings surprises and underreact to low-strength, high-weight surprises.

3.3 Overconfidence, the Illusion of Knowledge, and the Self-Attribution Bias

Overconfidence is a belief about a pending outcome that is greater than the underlying characteristics describing the outcome can justify. Overconfidence could emerge in the form of a person believing that the precision of an information set is greater than it actually is; people tend to underestimate the amount of uncertainty that truly defines the information set. Or, it could be in the form of a person overestimating his or her ability to successfully complete a task (including the processing of information or, say, correctly predicting an outcome).

One factor that can contribute to the overconfidence bias is the illusion of knowledge. Individuals tend to assume that their level of knowledge increases with the size of an information set; however, they fail to realize that any knowledge to be gained from any extra information is constrained by their general inabilities to properly interpret and understand information. In addition, overconfidence as a personality trait may derive from the more deeply rooted cognitive bias known as biased self-attribution whereby individuals tend to attribute successes to their own talents and skills while blaming failures on random bad luck. An investor's confidence may rise when an outcome turns in his or her favor, yet under this bias it will not fall commensurately following an unfavorable outcome. Hindsight bias—by which an individual, after an outcome is observed, overestimates his or her ex-ante ability to have predicted the outcome—also may contribute to overconfidence.

As demonstrated by Kent Daniel, David Hirshleifer, and Avanidhar Subrahmanyam (1998) with a theoretical model, the presence of investors impaired both by overconfidence (about the informativeness of their private information sets) and by the self-attribution bias can explain the well-documented empirical anomalies of momentum and reversals in returns. Another financial-market implication of overconfidence is that it can cause investors to trade too frequently, at a cost of realized returns that fall short of benchmarks, as per Barber and Odean (1999).

3.4 Aversion to Regret

Regret is an unpleasant intellectual and emotional sense that is affiliated with the knowledge that a different past decision would have yielded a current outcome that is more desirable than the outcome that did emerge from the actual decision taken. As documented by Kahneman and Tversky (1982), the uneasiness of this feeling will cause individuals to try to avoid it, even to the point of not taking any action at all out of fear that in hindsight the action will yield an outcome that is suboptimal. As experimental evidence, Ilana Ritov (1996) found that when human subjects are faced with a choice, and if they know that the outcome of the unselected alternative will be made known after the choice is made, their selection process will result in a different choice than if the unselected alternative's outcome will remain unresolved. In other words, when the probability of knowing the outcome of the unselected option is high (or certain), the probability of regret is higher. In turn, people will select the options that are less likely to yield worst outcomes in order to avoid eventual regret.

3.4.1 *Aversion to Losses*

Closely related to individuals' aversion to regret is their general aversion to losses. Shefrin and Statman (1985) have shown that—due to this emotional preference—investors tend to keep stocks of which prices have recently dropped (in contrast to selling stocks that recently experienced price increases). In other words, investors want to avoid the feeling of regret that accompanies selling a stock at a loss while they welcome the pride associated with selling a stock at a gain. Such behavior is in direct conflict with the objective of wealth maximization because of the tax advantages (disadvantages) associated with negative (positive) capital gains.[4] Using proprietary brokerage account data, Odean (1999) found a strong tendency for investors to hold their losing stocks for too long, suggesting a desire to avoid the feeling of regret after eliminating other alternate explanations for not selling "loser" stocks.

3.5 Sentiment

Sentimental investing is a type of irrational behavior in which factors that are uninformative nonetheless affect investing decisions. Sentiment may be purely emotions-based, and the investor may even be cognizant of the fact that the bases for such sentiment are not informative, yet still make investment decisions using these factors. Loyalty and admiration are examples of such factors. Also, such exogenous factors as familiarity with firms' products or services, geographic locations of corporate headquarters, high visibility within the news media and popular press, or having stocks listed on major exchanges or having stocks added to major stock indices are all potential sources of investor sentiment.[5] Other sentiment may relate to characteristics of an investment or pieces of news that are thought to be informative when in actuality they are not, so that investors trade based on these attributes as if they were informative (as suggested by Fischer Black (1986)).

4 Point Spread Betting Markets

With the backdrop of behavioral finance, potential departures from rationality, and common cognitive errors and other sources of sentiment in place, the focus of this chapter now turns to point spread wagering. The goal here is to establish that the point spread market is a viable research setting. This section will explain how a point spread market functions and will emphasize the usefulness and advantages of this market as a laboratory in which to perform tests of rationality.

4.1 Market Conventions and Mechanics

In any point spread wagering market, the asset at stake for a given game is a proposition that the favored team will defeat the underdog team by an amount greater than the point spread.[6] A bet on the favored team wins if that team wins the game by an amount greater than the spread while a bet on the underdog wins if the underdog team loses by an amount less than the spread or wins the game outright. As per the "11-for-10 rule," a winning bet pays an amount equal to $(1 + 10/11)$ (the original bet) and a losing bet pays zero. If the favored team wins the game by an amount exactly equal to the point spread, all wagers are refunded.[7]

An important participant in point spread markets is the bookmaker (or sports book), the entity that facilitates wagers on either team in any given contest. The bookmaker wants its expected post-game wealth level to be independent of the game's outcome.[8] If equal dollar amounts are wagered on both teams in a game, the sports book is guaranteed a commission equal to 1/22 of the combined total wagers. In the presence of this "11-for-10" commission, a winning rate of 52.38 percent is required for a bettor to break even.[9]

The point spread is the mechanism that the bookmaker uses to try to satisfy its objective of balancing the books for a given contest. In the absence of any sentimental (or irrational) betting, the bookmaker maintains a spread that is equal to the median of the distribution of possible outcomes, given whatever information set exists at the time. Such a point spread makes the probability of winning a wager on the favored team and the probability of winning a wager on the underdog team both equal to 50 percent, ignoring the small probability associated with a push. On the other hand, if the bookmaker anticipates or observes a sentimental (or irrational) clientele of bettors favoring one team, the point spread will be different from the median possible outcome and the aforementioned probabilities will both deviate from 50 percent. In the presence of this biased spread, more than half of the rational bettors' money will be on the sentimentally undesirable team, serving to offset the dollars wagered by irrational bettors on their preferred team and allowing the bookmaker to maintain a balanced book.

The first point spread that is posted by the sports book is called the opening spread; it is the spread at which betting commences. Bets on a game arrive during the betting period, and a game's spread should change any time the bookmaker realizes, or even anticipates, an imbalance in dollars bet on the two contestants.[10] The imbalance may emerge due to the arrival of either new information or a biased clientele of bettors, or it could emerge due to randomness. Betting continues until the instant when the actual contest begins; the spread that exists when the betting market closes is called the closing spread.

The length of time that the point spread market is open for wagering on a particular game differs across the four most common markets. Betting on almost all National Football League (NFL) games commences on Monday morning, when opening spreads are posted at the casinos. Nearly all games are played on Sundays, so that for each NFL game the market is typically open for about 6.5 days. The market for wagering on college football games usually opens on Sunday nights and stays open until the respective kickoffs, a large majority of which occur on Saturday. At the other extreme, the market for wagering on a basketball game—either professional or college—is typically open for just half a day, right up until tip-off. The brevity of these markets is due to the higher number and daily frequency of basketball games.

Besides opening spread, closing spread, and actual outcome, two other key variables in studies of sports betting markets are forecast error and change in spread. A game's forecast error is calculated as the difference between the actual outcome and the point spread; depending on the desired test it is alternately specified using either opening spread or closing spread. The change in spread for a game is straightforward: it is the difference between the closing and opening spreads.

4.2 Betting Markets as Useful, Advantageous Laboratories

Point spread wagering markets parallel financial markets in numerous ways. Both markets are characterized by the presence of rational investors, informed agents, arbitrageurs, irrational traders, and sentimental participants. Like arbitrageurs and informed traders in financial markets, professional bettors stand ready to exploit any arbitrage opportunities that might arise in sports gambling markets.[11] Information about point spread wagers and information about stocks are both widely disseminated.

Sports betting "wise guys" saturate the news services and gambling trade publications with their "wisdom," analogously to those expert stock pickers who offer their predictions for capital markets. The bookmaker for point spread wagers has a near-perfect analogue in the form of the market maker for stocks and other securities. The bookmaker uses the point spread to balance wagers on the two teams in a contest, just as the market maker uses price to manage order flow for a stock. And, just to be clear, bettors are sports betting's analogue to investors and traders. Thus, because of these highlighted similarities, any findings using point spread markets can be useful for wider financial audiences.

Furthermore, point spread markets possess one key feature that makes them an ideal setting for empirical studies: every wager reaches its terminal value in a fairly short period of time. The existence of a clear settling-up point is important for studies of rationality, since it allows for direct measurement of fundamental value for each wager. A game's perfectly observable outcome can be compared with the point spread(s) for the game so that bettors are easily able to determine which point spread wagers are winners. The single-payoff feature stands in contrast to stocks, which by definition have an infinite stream of possible future payoffs and which thereby create difficulty in connecting changes in prices to revisions of expected future cash flows. Christopher Avery and Judith Chevalier (1999) emphasized the usefulness of bets' relatively short lives. The short life of any given wager reduces the likelihood that new information will enter the marketplace during trading and cause spreads to change, thereby allowing researchers to focus more acutely on point spread movements caused by sentiment or by other behavioral biases.[12]

Occasionally academics argue that point spread betting markets are too different from more traditional financial markets due to either the negative expected payoff on a wager or the entertainment value of betting. However, in spite of their expected wealth being negative, and while likely deriving pleasure and other consumption value from betting, nearly all bettors are also heavily focused on maximizing their expected utility of wealth, just as stock market investors are. John Conlisk (1993) presented a model in which expected utility of wealth is the predominant component of a gambler's preference function, augmented with a second component that captures a relatively smaller additional taste for gambling.[13] Occasional skeptics notwithstanding, point spread betting markets are useful laboratories in which to perform tests of investor rationality because of their numerous parallels to, and advantages relative to, conventional financial markets.

5 Empirical Implications of the Various Biases in Point Spread Markets

While the implications of the behavioral biases and the alternate specifications of utility functions are generally more apparent at the individual level, most of these biases—if they characterize a sufficiently large portion of the betting public—can impact aggregate prices (i.e., point spreads). In his late-1990s survey of the economics of wagering markets, Raymond Sauer (1998, 2031) anticipated such impacts and predicted that "future papers on wagering markets will certainly be motivated by the behavioral approach." As of yet, no studies have tested for evidence of behavioral biases at the level of individual bettors, largely due to the scarcity of betting data at the individual account level. Instead, all of the behavioral-focused studies of point spread markets have tested for evidence only at the market level—these studies are still sparse enough that I can summarize nearly all of them here.

5.1 Typical Tests for Irrational Behavior

Each study presented in this section employed one or more types of tests for rationality. One typical test is to formulate betting strategies based on anticipated irrational behavior and then compare the profitability of these strategies against the break-even rate of 52.38 percent. Another type of test is whether mean forecast errors are statistically significantly different from zero or statistically significantly different across groups of wagers. Another common tool is an ordinary least squares (OLS) regression analysis in which the dependent variable is actual outcome and the independent variables are any combination of point spreads (opening or closing), fundamental variables (to proxy for teams' skills, abilities, and general performance attributes), and other variables that reflect various potential sources of irrational behavior. Probit analyses are a different way to test whether the same types of explanatory variables can explain wager outcomes, with the dependent variable taking a value of $+1$, 0, or -1, depending on whether a wager wins, is a push, or loses.

A more recent set of papers test for the effects of irrationality on the spread-formation process: the dependent variable in an OLS regression analysis is change in spread and the independent variables are different proxies for the presence of sentiment and other behavioral biases. Greg Durham and Muku Santhanakrishnan (2012) established that change in spread is the appropriate analogue to realized return in financial markets. This point is particularly important for any researchers who are interested in testing financial economic theories using a point spread market as the setting yet who may want to position their studies within the finance literature as opposed to the sports economics literature. Tests of investor behavior commonly utilize percentage return as the variable of interest; tests of bettor behavior can similarly employ change in spread.

5.2 Informed Bettors and Potential Limits to Arbitrage

One such-positioned study is by John Gandar, Bill Dare, Craig Brown, and Rick Zuber (1998). While not addressing behavioral aspects of betting, this research is still noteworthy for two reasons: (1) it establishes the presence of informed bettors in at least one betting market and (2) it is the first study to employ the intra-period change in point spread, albeit as a conditioning variable and not as a dependent variable. Examining a sample of wagers on National Basketball Association (NBA) games, Gandar et al. hypothesized that change in spread (from opening spread to closing spread) reflects the betting activity of rational traders who possess better information than what the bookmaker possesses when setting the opening spread. If their hypothesis is true, the degree of initial mispricing (i.e., the size of the forecast error in the opening spread) should be positively related to the size of the subsequent change in spread, and the change in spread should be in the direction of the initially undervalued team. Sorting games into bins based on intra-day changes in spreads, Gandar et al. found that the profitability of betting on the initially undervalued teams (i.e., teams in whose directions spreads subsequently move) at opening spreads[14] is a positive function of change in spread, consistent with rational bettors' transactions serving to correct errors in bookmakers' opening forecasts. Information-based trading is an important facet of the NBA point spread market.

In contrast, an earlier study by Gandar, Zuber, Thomas O'Brien, and Ben Russo (1988) suggested an insufficient presence of informed bettors. These authors tested for evidence consistent with rationality in the point spread betting market for NFL games. While statistical tests do not allow for rejection of rationality, the authors did find three different behavioral-based betting strategies (all implemented at closing spreads) to be abnormally profitable. These strategies' profitability suggests that the certain types of irrational betting are prevalent enough to create biases in spreads that are not fully offset—or corrected—during the course of betting by rational, informed participants, perhaps due to the limits to arbitrage discussed in section 2. The role of informed traders is limited.

5.3 Market Effects of Conservatism and Representativeness

To date, conservatism and representativeness are the two cognitive biases that appear most prominently in studies of bettor behavior in point spread wagering markets. Perhaps the prevalence of these biases is not surprising since both of these biases involve processing information as it arrives, often in sequential or successive pieces, and then updating beliefs about an often unobservable parent population, group, or process. Bettors impaired by conservatism will underreact to the true statistical importance of the most recent information so that prices (or point spreads) will only gradually incorporate the new information as bettors gradually respond. Bettors affected by the representativeness heuristic will overreact to recent information, will tend to want to detect patterns amid randomness, and will often over-extrapolate past performance into the future.

5.3.1 *Bettor Responses to Strength and Weight*

Gandar et al. (1988) examined a strategy of betting against favored NFL teams that, in their most recently preceding games, covered the point spread by 10 points or more. The potential profitability of this strategy is consistent with a hypothesis by Griffin and Tversky (1992) that, upon encountering a new piece of information about an uncertain future event or parent population, individuals over-rely on the "strength" of a new piece of evidence and under-rely on the same evidence's "weight." One implication of this hypothesis is that in response to high-strength, low-weight evidence people overreact in a manner suggestive of representativeness. Gandar et al.'s finding that this betting strategy is abnormally profitable suggests that bettors are impaired by representativeness; bettors do seem to overreact to large lagged forecast errors in spite of the fact that a wager's payoff is independent of the amount by which the wagered-upon team covers the spread.

In a closely related study Durham and Santhanakrishnan (2008) tested predictions of Griffin and Tversky's (1992) behavioral hypothesis with an interest in discovering whether bettors might be impaired with conservatism or representativeness. For a sample of college football wagers, the authors used length of a team's current winning streak[15] and the average amount by which the team had covered the spread during the same streak as measures of weight and strength, respectively. Change in spread was their measure of bettors' reaction. Since a wager's payoff for depends on whether the wagered-upon team covers the spread, irrespective of the magnitude by which it covers the spread, the weight of the wager's outcome is simply whether the bet wins or not. The amount by which the team covers (or fails to cover) the spread can be thought of as the strength of the outcome. The authors found that, holding weight constant, bettors overreact more to high-strength games than to low-strength games. They also found that bettors underreact more to games involving high weight than to games low weight, for games of similar strength. Both findings are consistent with G&T's predictions of ways by which individuals impaired by conservatism and representativeness would affect prices (or point spreads).

5.3.2 *Momentum, Streaks, and the Hot Hand*

Philip Gray and Stephen Gray (1997) are among the first researchers to acknowledge behavioral finance research in a study that used betting markets. As part of a thorough examination of market efficiency for the point spread wagering market for NFL games, Gray and Gray accurately drew the analogy of momentum in teams' performances against the spread with momentum in stock prices. They acknowledged the well-documented behavioral phenomenon whereby individuals perceive runs in random series, though they did not mention the representativeness bias by name. They constructed dummy variables to capture season-long, as well as recent, team performance. Using a probit analysis, Gray and Gray found that bettors appear to underreact to teams' season-long performance to date (suggestive of conservatism) but that these same bettors seem to overreact to recent form (i.e., to win–loss records in teams' most recent four games). These findings together suggest a profitable (though untested) betting strategy of betting on teams that have recently performed poorly yet have performed well across the season overall.

Durham, Michael Hertzel, and Spencer Martin (2005) analyzed data on college football wagering to conduct various tests of the BSV model of investor behavior described in section 3. The authors examined the numbers of continuations and reversals in teams' performance histories of various lengths and found that football bettors do not seem to update beliefs in the same manner as BSV's representative investor. Durham, Hertzel, and Martin also found that bettors believe that a team's performance is more likely to reverse as the team's streak (winning or losing) grows longer. This finding contradicts the BSV model's implication that investors become more certain about the presence of a continuation regime as the number of successive wins (or losses) increases. Durham, Hertzel and Martin's two main findings do not support BSV's theoretical model of investor behavior.

One potential outcome from reliance on the representativeness heuristic is that belief in the phenomenon known as the hot hand will emerge. In the respective realms of gambling and sports, the belief that a roulette wheel, a craps table, an athlete, or a team is "hot" is quite common; thus the manifestation of such a belief in sports wagering markets is not surprising. Bettors will perceive streaks, or patterns in performance, when performance is, in fact, random. Colin Camerer (1989) and Rodney Paul and Andrew Weinbach (2005a) found evidence that wagers on NBA teams on current winning streaks tend to lose more frequently than the expected 50 percent, which means that point spreads are higher than actual outcomes for games involving such streaking teams. These findings are consistent with bettors' mistaken belief in the so-called hot hand: bettors overreact to winning streaks and believe that a winning pattern has emerged when in fact performances relative to spreads should be random.[16] Bettors overbet on these teams that have been winning recently, causing wagers to be mispriced and creating a profitable betting strategy of betting against these recent winners. Although not specifically mentioned by either Camerer or Paul and Weinbach, bettors impaired by representativeness is a viable explanation for these findings. Camerer makes a more general claim: that a persistent misunderstanding of randomness can lead to belief in the hot hand (i.e., in winning more frequently than a random process would suggest).

5.4 Market Effects of Bettor Sentiment

Avery and Chevalier (1999) (hereafter A&C) examined intra-week changes in NFL point spreads, focusing on possible sources of sentiment that have natural stock market analogues. The authors found that bettors do seem to follow "expert" forecasts despite the general inability of prognosticators to predict which teams will cover the spreads. NFL bettors also appear to overreact to, and cause spreads to move in the direction of, teams that perform well in the recent term (i.e., teams exhibiting short-run momentum). Also, bettors' sentiment for teams that are highly visible or highly recognizable (proxied for by membership in a dominant, large-market football conference or by appearance at the top of the standings) seems to cause spreads to move during the week. Overall, A&C's results suggest that bettors are affected by these various types of sentiment and that sentimental wagering does affect price paths or point spread paths. A&C also found sentiment-based biases in opening spreads, suggesting that bookmakers attempt to incorporate anticipated sentiment into opening spreads; this bias is in addition to that caused by the systematic intra-week sentimental betting. The two biases together lead to a strategy of betting against sentimentally popular teams at closing spreads, a strategy that is marginally profitable after accounting for transaction costs.

Following A&C's lead, Durham and Tod Perry (2008) found evidence that sentiment also affects the dynamic spread-formation process for college football betting. Their findings suggest that bettors naively follow the advice of experts, believe in the hot hand, prefer to bet on teams that are visible (in terms of major -conference membership or appearance at the tops of standings), and bet on teams about which they are most avid (or loyal). The fan-avidity variable marks this study's novel contribution to the literature and relies on the premise that college football bettors exhibit strong emotional commitment and loyalty to particular teams (Wayne Root and Wilbur Cross (1989)). Durham and Perry constructed a measure of fan avidity for each team by using survey results from a national market research firm that asked fans to indicate their favorite college football teams. As noted, spreads do move in the direction of the teams to which fans are more loyal. However, betting strategies designed to exploit fan avidity and the other hypothesized types of sentiment are not profitable except under extreme filtering rules.

5.5 Other Biased Behavior without Psychological Basis

A general phenomenon—albeit not rooted in any specific psychological biases—is that bettors appear to systematically overbet on heavy favorites,[17] creating profitable wagering strategies in all of the markets tested to date. For example, Paul, Weinbach, and Christopher Weinbach (2003) found that betting on big underdogs in college football is abnormally profitable (in the absence of bookmaker commissions) and that betting on big underdogs when they are at home yields abnormal profits in excess of commissions. Similarly Paul and Weinbach (2005a) showed that betting on big underdogs in the NBA is abnormally profitable (without commissions) and is abnormally profitable (in excess of commissions) when the sample is reduced to big-underdog teams playing at home. And in a study of college basketball wagering, Paul and Weinbach (2005b) found that a strategy of betting on big underdogs generates a winning percentage that is statistically significantly different from 50.00 percent. However, in contrast to the big-home-underdog phenomenon in the first two markets, abnormal profits come from betting on heavy-underdog visiting teams.

While this overarching finding that people tend to overbet on heavily favored teams is striking in that it seems to repeat across all point spread markets, only one of the authors' explanations hints at either a cognitive bias or sentiment as the reason. Paul, Weinbach, and Weinbach (2003) suggested an asymmetry in information available about the heavily favored teams and information available about their opponents (in college football), a suggestion that is reminiscent of Robert Merton's (1987) investor recognition hypothesis. Avery and Chevalier (1999) and Durham and Perry (2008) found that bettors are attracted to teams that are members of the prominent conferences or that generally appear at the top of the win–loss standings each year. Avery and Chevalier were the first to propose that these attributes are proxies for teams that bettors may have read or heard a lot about in the media and that bettors may be sentimentally predisposed to bet on such teams. The apparent overbetting on heavy favorites seems to support this notion.

A number of other ad hoc betting rules have been found to be abnormally profitable for different sports across different time periods. However, because the potential behavioral explanations for any of these findings are not readily apparent, these rules are not included in this survey.

6 Conclusion and Directions for Future Research

Behavioral finance has made great strides in recent years in terms of gaining acceptance in finance theory, research, and practice. Prospect theory has proven to be robust in explaining many empirical and experimental findings in financial markets, findings that were considered to be anomalies under the assumption of rational behavior. Multiple theories of investor behavior can explain the well-documented phenomena of short-run momentum and long-run reversals in returns, though to find tests that will distinguish among the competing theories can be difficult.

The point spread betting market might help to reduce at least some of the difficulties and challenges facing all tests of investor (or bettor) behavior, since the asset-pricing question is simplified by a clear settling-up point for each bet and by the absence of any systematic risk. Many researchers recognize the advantages of this simpler market and have used it to find evidence consistent with bettors being impaired by representativeness, belief in the hot hand, and various types of sentiment.

Looking ahead, one significant development will be if any researchers can gain access to proprietary betting data at the individual account level. Having such data would allow for tests that are sports betting's analogues to all of the work performed by Barber and Odean (and other coauthors) using stock trading data from household brokerage accounts. Anyone equipped with such individual-level, point spread betting data could then test for such sentimental preferences as bettors betting on teams situated in the same geographic region as where they live. Or, dollar-volume data could be aggregated and examined to see whether betting volumes follow patterns similar to those of point spreads.

Another useful, relatively new type of data are the percentages of dollars wagered on each team in a point spread wager, data that are currently available from at least two services: Sports Insights and Sportsbook.com, as per Paul and Weinbach (2011). Paul and Weinbach are the first researchers to successfully employ such data, though thus far the data have been used primarily to reject the long-believed notion that the objective function of the bookmaker is to "balance the books" and to reject claims of point shaving in the sport of college basketball. Perhaps these same betting-proportions data can be examined to see whether the same sources of sentiment mentioned in this survey seem to affect the proportions of dollars wagered in games which have net sentiment for one team.

For researchers who are interested in positioning their sports-betting studies more squarely amid the behavioral finance literature, change in point spread should likely be an important variable of interest. To the extent that bookmakers cannot perfectly anticipate irrational behavior when setting opening spreads and if the clientele of arbitrageurs is insufficient to correct all pricing errors, irrationality should affect the dynamic intra-period spread-formation process. Changes in spreads will reflect these effects.

With the establishment of change in spread as an important new variable of interest, the emergence of new data types (such as proportions of dollars wagered on the two teams in each contest and intra-period point spreads besides only opening and closing spreads), and increasingly easier access to such data, the quote at the beginning of this chapter from Gandar, Zuber, O'Brien, and Russo (1988) resonates even more soundly now than when it was written in 1988. An updated version might read like this: The point spread betting market is likely an even more fruitful place than it has ever been for conducting research about behavioral theories that could apply in conventional market settings.

Notes

1. This model is discussed in more detail in subsection 3.2.
2. For firms that perform poorly, the authors find inverse responses: overreaction drives prices too low and then prices are eventually corrected.
3. For ease of exposition in the remainder of this chapter, either "the representativeness bias" or simply "representativeness" will be used regularly in place of the more cumbersome—but correct—phrase "reliance upon the representativeness heuristic."
4. Furthermore, Odean (1999) has shown that "winner stocks" generally outperform "loser stocks" across various-length time periods after the winners are sold and the losers are kept, or held.
5. Christo Pirinsky and Qinghai Wang (2006) found strong evidence suggesting that stock price formation has a geographic component whereby a region's residents seem to invest

disproportionately in firms that are headquartered in their own region. Lily Fang and Joel Peress (2009) showed that the degree of mass media coverage of a company affects the company's stock returns. These last two factors (or events) also relate to greater visibility for a firm and its stock. Merton's (1987) investor recognition hypothesis shows how a firm's expected stock returns are affected by investors' degree of familiarity with the company's products or services. Stephen Foerster and Andrew Karolyi (1999) documented the negative effects on returns of American Depository Receipts after they list on U.S. exchanges. Bala Dharan and David Ikenberry (1995) found evidence consistent with the investor recognition hypothesis following stock listings on the American Stock Exchange and the New York Stock Exchange. Honghui Chen, Gregory Noronha, and Vijay Singal (2004) found evidence of increases (decreases) in investor awareness following index additions (deletions).

6. The point spread can take any value in half-point increments. On rare occasions the point spread can equal 0, in which case the "favorite" and "underdog" labels are dropped and the game is labeled as a "pick 'em" game.
7. Introducing some sports betting lingo, the three possible outcomes for a given wager are: the favorite "covers the spread," the underdog "covers the spread," or the bet is a "push."
8. While the literature provides some evidence in support of the bookmaker's objective being different from "balancing the books" (see, for example, Steven Levitt 2004; Paul and Weinbach 2007; Paul and Weinbach 2011), the more widespread belief (or understanding) is that the bookmaker prefers to maximize order flow while minimizing exposure to games' outcomes. Avery and Chevalier (1999) offered a nice explanation with multiple reasons for why a bookmaker's goal should be to minimize its exposure to a game's outcome against the spread. For additional explanations, please see the literature review by Sauer (1998).
9. Solving for p in the following equation which sets expected profit equal to \$0 yields the necessary winning percentage to break even: $p \cdot \$10 + (1 - p) \cdot -\$11 = \$0$.
10. Point spread changes are frequent and nontrivial. For a sample of 7,904 wagers on professional basketball games, Gandar et al. (1998) found that changes in spread (from opening spread to closing spread) are different from zero for 79.5 percent of the observations. For this same sample the mean absolute value of the change in spread is equal to 1.08, and the standard deviation of change in spread is 0.80. For Avery and Chevalier's (1999) sample of 2,366 wagers on professional football games, the change in spread is non-zero for 70.2 percent of the games. The mean absolute value of the change in spread is 0.68, and the standard deviation is roughly 0.96. For Durham and Perry's (2008) sample of 4,584 college football wagers, the spread changes for 84.5 percent of the games. The mean absolute change in spread is 1.44, and the standard deviation is 2.73.
11. As noted by Paul, Weinbach, and Weinbach (2003), among others, sports books have limits on the maximum sizes of bets. These limits may prohibit informed bettors from placing wagers that are large enough to correct many of the mispricings caused by irrational betting. These limits can perhaps be circumvented if a bettor places a number of bets at different casinos or offshore.
12. Gandar et al. (1988) reported that only about five percent of the changes in spreads for NFL games are caused by the arrival of new information; the figure is probably similar, if not lower, for college football games due to privacy rules that prevent the release of certain types of information about student athletes. For both professional and college basketball games, this percentage will be even lower because the length of time for which markets

for wagering on basketball games are open is about one-tenth of that for football game wagering.

13. The utility-of-wealth function and the taste-for-gambling function are additive. The first function is a classic Von Neumann–Morgenstern expected utility function, where the underlying utility function is increasing and concave in wealth. The second function is increasing and concave in the size of the wager and is increasing in the probability of winning the wager.
14. This betting strategy is obviously not implementable, since it involves placing wagers at opening spreads, conditional on subsequent changes in spreads.
15. Here and throughout the remainder of this chapter the reader can presume that all winning streaks, losing streaks, and performance are always defined from a wagering perspective (i.e., wins and losses relative to spreads), not from the perspective of teams winning and losing games outright.
16. William Brown and Sauer (1993) verified that the hot hand is not some fundamentally grounded improvement in team performance; for Camerer's sample, the hot hand is, indeed, a myth.
17. The definition of *heavy favorite* depends on the sport. For college football, to be heavily favored is to be favored by more than 28 points. For college basketball, a heavy favorite is favored by 20 or more points; for NBA basketball, the threshold is 10 points.

References

Avery, Christopher, and Judith Chevalier. 1999. Identifying investor sentiment from price paths: The case of football betting. *Journal of Business* 72(4):493–521.

Barber, Brad M., and Terrance Odean. 1999. The courage of misguided convictions: The trading behavior of individual investors. *Financial Analysts Journal* 55:41–55.

——. 2001. Boys will be boys: Gender, overconfidence, and common stock investment. *Quarterly Journal of Economics* 116(1):261–292.

Barberis, Nicholas, Andrei Shleifer, and Robert W. Vishny. 1998. A model of investor sentiment. *Journal of Financial Economics* 49(3):307–343.

Black, Fischer, 1986. Noise. *Journal of Finance* 41(3):529–543.

Bloomfield, Robert, and Jeffrey Hales. 2002. Predicting the next step of a random walk: Experimental evidence of regime-shifting beliefs. *Journal of Financial Economics* 65(3):397–414.

Brown, William, and Raymond D. Sauer 1993. Does the basketball market believe in the "hot hand"? Comment. *American Economic Review* 83(5):1377–1386.

Camerer, Colin F. 1989. "Does the basketball market believe in the "hot hand"? *American Economic Review* 79(5):1257–1261.

Chen, Honghui, Gregory Noronha, and Vijay Singal. 2004. "The price response to S&P 500 Index additions and deletions: Evidence of asymmetry and a new explanation. *Journal of Finance* 59(4):1901–1929.

Conlisk, John. 1993. The utility of gambling. *Journal of Risk and Uncertainty* 6(3):255–275.

Daniel, Kent, David Hirshleifer, and Avanidhar Subrahmanyam. 1998. Investor psychology and security market under- and overreactions. *Journal of Finance* 53(6):1839–1885.

DeBondt, Werner F. M., and Richard Thaler. 1985. Does the stock market overreact? *Journal of Finance* 40(3):793–805.

DeLong, J. Bradford, Andrei Shleifer, Lawrence H. Summers, and Robert J. Waldmann. 1990. Noise trader risk in financial markets. *Journal of Political Economy* 98(4):703–738.

Dharan, Bala G., and David L. Ikenberry. 1995. The long-run negative drift of post-listing stock returns. *Journal of Finance* 50(5):1547–1574.

Durham, Gregory, and Tod Perry, 2008. "The impact of sentiment on point spreads in the college football wagering market. *Journal of Prediction Markets* 2(1):1–27.

Durham, Gregory, and Mukunthan Santhanakrishnan. 2008. The relative importance of strength and weight in processing new information in the college football betting market. *Journal of Prediction Markets* 2(2):13–28.

——. 2012. Point-spread wagering markets' analogue to realized return in financial markets. *Journal of Sports Economics* 13(5):554–566.

Durham, Gregory R., Michael G. Hertzel, and J. Spencer Martin. 2005. The market impact of trends and sequences in performance: New evidence. *Journal of Finance* 60(5):2551–2570.

Edwards, Ward. 1968. Conservatism in human information processing. In *Formal representation of human judgment*, edited by Benjamin Kleinmuntz. New York: John Wiley & Sons, 17–52.

Fang, Lily, and Joel Peress. 2009. Media coverage and the cross-section of stock returns. *Journal of Finance* 64(5):2023–2052.

Foerster, Stephen, and G. Andrew Karolyi. 1999. The effects of market segmentation and investor recognition on asset prices: Evidence from foreign stocks listing in the United States. *Journal of Finance* 54(3):981–1013.

Frazzini, Andrea. 2006. The disposition effect and under-reaction to news. *Journal of Finance* 61(4):2017–2046.

Gandar, John, Richard Zuber, Thomas O'Brien, and Benjamin Russo. 1988. Testing rationality in the point-spread betting market. *Journal of Finance* 43(4):995–1008.

Gandar, John, William Dare, Craig Brown, and Richard Zuber. 1998. Informed traders and price variations in the betting market for professional basketball games. *Journal of Finance* 53(1):385–401.

Gray, Philip K., and Stephen F. Gray. 1997. Testing market efficiency: Evidence from the NFL sports betting market", *Journal of Finance* 52(4):1725–1737.

Griffin, Dale, and Amos Tversky. 1992. The weighing of evidence and the determinants of confidence. *Cognitive Psychology* 24(3):411–435.

Jegadeesh, Narasimhan, and Sheridan Titman. 1993. Returns to buying winners and selling losers: Implications for stock market efficiency. *Journal of Finance* 48(1):65–91.

Kahneman, Daniel, and Amos Tversky. 1972. Subjective probability: A judgment of representativeness. *Cognitive Psychology* 3(3):430–454.

——. 1979. Prospect theory: An analysis of decision under risk. *Econometrica* 47(2):263–291.

——. 1982. The psychology of preferences. *Scientific American* 246(1):160–173.

Lakonishok, Josef, Andrei Shleifer, and Robert W. Vishny. 1994. Contrarian investment, extrapolation, and risk. *Journal of Finance* 49(5):1541–1578.

Levitt, Steven, D. 2004. Why are gambling markets organized so differently from financial markets? *Economic Journal* 114(495):223–246.

Markowitz, Harry. 1952. The utility of wealth. *Journal of Political Economy* 56(4):151–158.

Merton, Robert, C. 1987. A simple model of capital market equilibrium with incomplete information. *Journal of Finance* 42(3):483–510.

Odean, Terrance. 1999. Are investors reluctant to realize losses? *Journal of Finance* 53(5): 1775–1798.

Paul, Rodney, and Andrew Weinbach. 2005a. Bettor misperceptions in the NBA: The overbetting of large favorites and the "hot hand." *Journal of Sports Economics* 6(4):390–400.

———. 2005b. Market efficiency and NCAA college basketball gambling. *Journal of Economics and Finance* 29(3):403–408.

———. 2007. Does Sportsbook.com set pointspreads to maximize profits? Tests of the Levitt model of sportsbook behavior. *Journal of Prediction Markets* 1(3):209–218.

———. 2011. Investigating allegations of pointshaving in NCAA basketball using actual sportsbook betting percentages. *Journal of Sports Economics* 12(4):432–447.

Paul, Rodney, Andrew Weinbach, and Christopher Weinbach. 2003. Fair bets and profitability in college football gambling. *Journal of Economics and Finance* 27(2):236–242.

Pirinsky, Christo, and Qinghai Wang. 2006. Does corporate headquarters location matter for stock returns? *Journal of Finance* 61(4):1991–2015.

Ritov, Ilana. 1996. Probability of regret: Anticipation of uncertainty resolution in choice. *Organizational Behavior and Human Decision Processes* 66(2):228–236.

Root, Wayne Allyn, and Wilbur Cross. 1989. *Betting to Win on Sports*. New York: Bantam.

Sauer, Raymond D. 1998. The economics of wagering markets. *Journal of Economic Literature* 36(4):2021–2064.

Shefrin, Hersh M., and Meir Statman. 1984. Explaining investor preference for cash dividends. *Journal of Financial Economics* 13(2):253–282.

Shiller, Robert, J. 1981. Do stock prices move too much to be justified by subsequent changes in dividends? *American Economic Review* 71(3):421–436.

———. 1985. The disposition to sell winners too early and ride losers too long: Theory and evidence. *Journal of Finance* 40(3):777–790.

Shleifer, Andrei, and Robert W. Vishny. 1997. The limits of arbitrage. *Journal of Finance* 52(1):35–55.

Von Neumann, John, and Oskar Morgenstern. 1944. *Theory of games and economic behavior.* Princeton, N.J.: Princeton University Press.

This page intentionally left blank

SECTION VI

PREDICTION MARKETS AND POLITICAL BETTING

This page intentionally left blank

CHAPTER 28

A SIMPLE AUTOMATED MARKET MAKER FOR PREDICTION MARKETS

DAVID JOHNSTONE

1 Introduction

This note describes a simple betting mechanism by which a robot market maker quotes prices for trades (buy or sell) in a binary security. The security traded has finite expiration time, at which time it is worth either $V = 1$ or $V = 0$. Its value at expiry is a random (uncertain) variable determined by the occurrence or nonoccurrence of a prespecified real-world event (e.g., a stock market increase over a given time interval). In finance securities such as this trade fall under the name of "binaries" or "digital options." In economics they are known as Arrow securities.

The market maker proposed here is more easily understood than the well-known Hanson (2003, 2007) market maker and extensions such as Othman et al. (2010). Its derivation does not involve any mention of a probability scoring rule or utility function. Its logic is based on an easily understood generalization of a pari-mutuel betting market. It differs nonetheless from a pari-mutuel market in that traders who bet for or against $V = 1$ (or $V = 0$), by either buying or selling a security that pays out under one of these outcomes, are aware at the time of trading what payout they will get under each of these two possible outcomes. Trades are therefore effectively fixed-odds bets; however, unlike conventional fixed-odds betting, all trades are executed at prices that vary continuously with trade size. More specifically, trade prices are worse for larger trades, as is typical of trading stocks and other financial securities.

Apart from its straightforward derivation, a notable advantage of this market maker algorithm is that the opening security price can be set naturally at any point between 0 and 1. This is useful when the market maker's ex ante probability of a given outcome, say $V = 1$, is not the inbuilt Hanson binary asset opening price of 0.5.

1.1 Prediction Markets

The Hanson market maker is widely used in online prediction markets operated by Inkling Markets, Consensus Point, Yahoo!, and Microsoft. Details are provided by Chen and Pennock (2010) and Othman et al. (2010). The basic advantage of an automated market maker in prediction markets is that traders (buyers and sellers) can trade at any time, unlike a common double auction where there may be no current offers on the other side of the trade. Automated prediction markets run at low cost and with minimum human oversight. Helpful and interesting surveys of the literature on prediction markets are provided by Wolfers and Zitzewitz (2004), Pennock et al. (2001), Chen et al. 2005, Chen and Pennock (2010) and Agrawal et al. (2011).

The theory and application of prediction markets developed in the field of computer science rather than in finance or economics, but the theoretical overlap between fields is striking and valuable. Note, for example, the connections drawn by Yiling Chen and David Pennock (2010) between prediction markets, market microstructure, and efficient markets. Automated prediction markets can also be viewed as a computerized form of experimental markets with trading conducted on-screen against potentially unknown individuals rather than across a table.

Automated prediction markets have potentially wide application in behavioral finance and behavioral economics. They provide a realistic market trading environment in which to test traders' natural versus rational instincts (in any chosen judgment or choice problem). Many prediction markets are designed such that the price of the security traded represents a "market probability" of some well-defined event. In this way they resemble the over-the-counter markets for binaries that are now widely traded by "Wall Street" style financial market makers, such as IG Markets.

One of the aims behind the advent of prediction markets was to design a mechanism that would extract and merge probabilistic opinion on important but uncertain phenomena, such as election results and other business, political, environmental, medical, and social phenomena. There is a very extensive empirical literature indicating in many contexts that market consensus probabilities are as accurate, and often more accurate, than such conventional forecasts as those made by experts or opinion polls (e.g., Berg, Nelson, and Rietz 2008). This is a generalization of the longer held empirical proposition that bookmakers' odds are better predictors of outcomes than those of at least the great majority of individual gamblers. See Chen and Pennock (2010) and Agrawal et al. (2011) for a partial survey of related literature.

2 A Generalized Pari-Mutuel Betting Market

In a theoretic commission-free pari-mutuel binary betting pool gamblers bet on either $V = 1$ or $V = 0$ and the total pool of money wagered is paid out to those who bet on the realized event in proportion to the amounts that they bet (the others lose their money). In a more general form of betting pool, each gambler i bets a pair, (X_i, p_i), representing an amount of cash X_i and a stated probability of $p_i = \Pr_i(V = 1) = 1 - \Pr_i(V = 0)$. Bets are therefore of the form "\$200 at probability 0.2." The payout or return

$$\begin{cases} X_i(p_i/q) & \text{if } V = 1 \\ X_i(1-p_i)/(1-q) & \text{if } V = 0, \end{cases} \tag{28.1}$$

where $q = \sum X_i p_i / \sum X_i$ is the value-weighted average (pari-mutuel) probability of $V = 1$. It is immediately evident that the total money payout under this unusual market construction is

$$\sum X_i = \begin{cases} \sum X_i(p_i/q) & \text{if } V = 1 \\ X_i(1-p_i)/(1-q) & \text{if } V = 0, \end{cases} \tag{28.2}$$

as occurs in a common pari-mutuel market. A conventional pari-mutuel betting market is simply the special case of (1)–(2) in which all gamblers i are constrained to state $p_i \in \{0, 1\}$, implying that when they lose their bet they lose all of it.

Numerical example. Consider four gamblers who bet respectively $(X_i, p_i) = (600, 0.5)$, $(200, 0.9)$, $(150, 0)$, and $(50, 1)$, on event $V = 1$. Note that the third gambler gives event $V = 1$ zero probability and, in effect, makes a conventional bet of amount 150 on $V = 0$. The total betting pool is 1,000 and the pari-mutuel market probability of $V = 1$ is

$$q = \frac{600(0.5) + 200(0.25) + 150(0) + 50(1)}{1000} = 0.4.$$

The respective payouts to each of the four gamblers, in the event of $V = 1$ or $V = 0$, are as shown in table 28.1.

Table 28.1 Payout to Gambler i Given $V = 1$ and $V = 0$

i	If $V = 1$	If $V = 0$
1	$600(0.5/0.4) = 750$	$600(0.5/0.6) = 500$
2	$200(0.25/0.4) = 125$	$200(0.75/0.6) = 250$
3	$150(0/0.4) = 0$	$150(1/0.6) = 250$
4	$50(1/0.4) = 125$	$50(0/0.6) = 0$
	1,000	1,000

3 Market Maker Equations

There are two securities. The first is a binary paying value $V = 1$ to its owner in the event of state E (and zero otherwise), and the second is identical except that it pays value $V = 1$ in the event of state *not-E*. The following description deals with pricing the first of these securities. There is no mechanical bond between the two security prices. Rather they are priced separately using the same MM (market maker) logic, meaning that the no-arbitrage condition (i.e., prices that sum to one) is satisfied only if traders make that happen via their trading decisions. In many applications there may be no need for simultaneous trade in both securities. Instead a trader can bet on E (*not-E*) by buying (selling, perhaps short-selling) only the E security.

For convenience of expression, one unit of the E security is described as "a share in E" or simply as a "share." All trades (buys and sells) in this security are expressed in terms of buying or selling some given number of shares, and all prices quoted by the MM are for a given parcel or number of shares. One share in E has value V at expiry.

Suppose that MM initiates the market by lodging a discretionary cash amount $B(B > 0)$ at given probability p $(0 < p < 1)$. At any time before expiry, MM will have sold some number of shares and bought some number of shares. Let the sum of money proceeds (prices) received from buyers (i.e., from selling shares) equal M_b. This is how much the buyers stand to lose to MM in the event of *not-E*. Now consider the traders who have sold shares to MM. They stand to lose money amount M_s in the event of E. This amount is equal to the number of shares they have sold MM (in total) minus the sum of the prices of all those previous sales (in the event of E each share is worth $V = 1$ and hence its seller loses one minus the price received on its sale). In effect, the traders who have sold shares to MM have wagered a total of M_s on outcome E at probability zero (see the numerical example above where trader $i = 3$ does the same thing). The assumption here is that amount M_s is deposited with MM as security so that sellers' positions (i.e., short positions) are covered for the event they lose.

Under this regime the aggregate betting pool at any moment in the life of the security is $(B + M_b + M_s)$. The implied market probability of E is

$$\begin{aligned} q &= \frac{B(p) + M_b(1) + M_s(0)}{B + M_b + M_s} \\ &= \frac{B(p) + M_b}{B + M_b + M_s}. \end{aligned} \tag{28.3}$$

The corresponding market probability of event *not-E* is

$$\begin{aligned} 1 - q &= 1 - \frac{B(p) + M_b}{B + M_b + M_s} \\ &= \frac{B(1 - p) + M_s}{B + M_b + M_s}. \end{aligned}$$

It is important to note that no mention has been made so far in this derivation of what prices MM buys shares at or sells shares for. Rather the quantities M_b and M_s are defined as merely the aggregate money amounts collected from buyers and sellers (respectively). Within these amounts it is typical that no two traders buy at the same price, or sell at the same price, since the prices quoted by MM change with the level and direction of preceding trade. The pricing mechanism by which this occurs is explained below.

3.1 Ask Price for n Shares

If the Ask price for n shares is Ask_n, then the trader's return in the case of $E(V = 1)$ from buying n shares is factor $n = Ask_n$. That is, by buying n shares for total price Ask_n the trader ends up with share value n in the event of E (each share is worth $V = 1$) and, therefore, a return of $n = Ask_n$. Note that returns are expressed as factors and that the trader's ending cash is the amount wagered or risked (here Ask_n) multiplied by that factor.

The MM offers buyers a return of $1 = q$ in the event of E, where q is the market probability of E after their trade. If the price paid for n shares is Ask_n, then the market probability at that instant is

$$q = \frac{Bp + (M_b + Ask_n)}{B + M_b + M_s + Ask_n}.$$

Note that the price paid by the buyer Ask_n increases the total money amount received by MM from selling shares.

Now, setting $n/Ask_n = 1/q$, gives

$$\frac{n}{Ask_n} = \frac{B + M_b + M_s + Ask_n}{Bp + (M_b + Ask_n)}.$$

Solving this quadratic equation for Ask_n gives just one sensible solution,

$$\frac{1}{2}\left[n - B - M_b - M_s + \sqrt{(n - B - M_b - M_s)^2 + 4(Bnp + nM_b)}\right]. \qquad (28.4)$$

For example, suppose that there have been no trades to date (implying that $M_b = M_s = 0$) and that $B = 100$ and $p = 0.5$, then the Ask price for $n = 1$ [2] {10}($n = 100$) [[$n = 1{,}000$]] shares is from (28.4). $Ask_n = 0.5025$ [1.01] {5.249}(707.107) [[9524.94]] or 0.503 [0.505] {0.525}(0.707) [[0.952]] per share.

The problem with these prices is that the trader will notice that to buy a parcel of n shares it is cheaper to buy them one at a time or better still in tiny fractions of a share. This can be seen as follows. Suppose the order is for $n = 2$. The first share costs 0.5025. The price of the second share is found by substituting $M_b = 0.5025$ in equation (28.4), since this is how much MM collected from the first share sold, and letting $n = 1$, giving an Ask price for the second single share of 0.5049 (note that $M_s = 0$ in this calculation, since no money has been received from short sellers). The sum of these two prices is 1.007, whereas if they were bought in one parcel of $n = 2$, the price would have been the higher amount of 1.01.

Suppose that a trader sets out to buy n shares. For each dollar the trader pays to MM he or she drives up q a little bit more and hence the return factor that she will earn in the event of E, that is, $1/q$, gets smaller for each dollar spent. To encourage a rational trader to take a larger n, and drive the instantaneous market probability q closer to that trader's personal probability of E, the MM allows traders to buy any order of $n > 0$ shares in arbitrarily small fractions of a share, thus minimizing the price paid for each infinitesimal fraction of a share and, therefore, maximizing overall expected return.

The MM Ask price for $n = N$ shares is then

$$\int_{n=0}^{N} \frac{\left[n - B - M_b - M_s + \sqrt{(n - B - M_b - M_s)^2 + 4(Bnp + nM_b)}\right]}{2n} \delta n \qquad (28.5)$$

Hence, for the example immediately above, the theoretical Ask prices for $n = 1\{10\}(n = 100)$ $[n = 1000]$ shares, calculated by numerical integration, are 0.501 {0.512}(0.613) {0.867}per share.

3.2 Bid Price for n Shares

If the Bid price for n shares is Bid_n, then the trader's return in the event of *not-E* ($V = 0$) from selling n shares is $n/(nBi - d_n)$. By selling n shares (to the MM) for total price Bid_n, the trader risks a loss of $(n - Bid_n)$, which occurs in the event of E. To commit to this transaction the trader deposits the full amount of the potential loss $(n - Bid_n)$ with MM. In the event of *not-E*, the MM pays this deposit back plus amount n. The trader is thus left with net gain of $n - (n - Bid_n) = Bid_n$, which is the agreed price of the shares sold by the trader to MM.

The MM offers sellers a return of $1/(1 - q)$ in the event of *not-E*, where q is the market probability of E given that trade

$$q = \frac{Bp + M_b}{B + M_b + M_s + (n - Bid_n)}.$$

Now, setting $n/(n - Bid_n) = 1/(1 - q)$, gives

$$\begin{aligned}\frac{n - Bid_n}{n} &= 1 - q \\ &= \frac{B(1 - p) + M_s + (n - Bid_n)}{B + M_b + M_s + (n - Bid_n)}.\end{aligned}$$

Solving this quadratic equation for Bid_n again gives just one sensible solution,

$$Bid_n = \frac{1}{2}\left[B + M_b + M_s + n - \sqrt{(n + B + M_b + M_s)^2 - 4(Bnp + nM_b)}\right].$$

The MM Bid price for $n = N$ shares is then

$$\int_{n=0}^{N} \frac{\left[B + M_b + M_s + n - \sqrt{(n + B + M_b + M_s)^2 - 4(Bnp + nM_b)}\right]}{2n} \delta n. \qquad (28.6)$$

Taking the same example as above, the theoretical Bid prices for $n = 1$ $\{10\}(n = 100)$ $[n = 1{,}000]$ shares, calculated from (28.6) by numerical integration are 0.499 {4:875}(38.701) [132.970] or 0.499 {0.488}(0.387) [0.133] per share.

4 Alternative Method of Computation

Another way to calculate the prices found above proceeds as follows. First consider the Ask price. The trader who buys shares is charged a "price" per infinitesimal "unit" of stock

$$\frac{Bp + M_b + M}{B + M_b + M_s + M}$$

that increases continuously with the money amount M spent to acquire those shares (M_b and M_s are as defined above). Note that this "price" function corresponds to (28.3) and reacts to M in the same way as a conventional pari-mutuel market probability.

The unit average price paid over a total expenditure of amount $M = m$ is then

$$\begin{aligned}\bar{a} &= \frac{1}{m}\int_{0}^{m} \frac{Bp + M_b + M}{B + M_b + M_s + M}\delta M \\ &= \frac{1}{m}\left\{m - (B(1-p) + M_s)\left[\log(m + B + M_b + M_s) - \log(B + M_b + M_s)\right]\right\}.\end{aligned}$$

The number n of shares obtained by the trader who wagers (pays) total amount m is therefore $m/\bar{a}$. Thus

$$n = \frac{m^2}{m - (B(1-p) + M_s)\left[\log(m + B + M_b + M_s) - \log(B + M_b + M_s)\right]}. \qquad (28.7)$$

The Ask price m at which the trader can buy some chosen number n of shares is then found by substituting that number n into (28.7) and then solving this equation numerically to find m. Note that m is the amount that the buyer risks losing by buying n shares or, in other words, the full amount of his wager.

In Mathematica, the best method to solve (28.7) is to use the *FindRoot* function, starting the search at $m = n/2$ (which is a good guess since n shares are always worth somewhere between 0 and n). This computation method leads to the same Ask prices as found from (28.5). An advantage of (28.7) over (28.5) is that although both equations require numerical solution, (28.7) does not involve numerical integration and is solved extremely quickly and accurately. When calculated from (28.7), the Ask prices for an order of $n = 1\{10\}(n = 100)$ $[n = 1{,}000]$ shares (with $B = 100$, $p = 0.5$, and $M_b = M_s = 0$ as assumed above) are 0.501 {0:512}(0:610) {0:869} per share. Note the small discrepancies between the last two of these results and those found by solving (28.5).

Now consider the matching approach to calculating the Bid price. The trader who sells shares to MM obtains a "price" per infinitesimal "unit" of stock

$$\frac{Bp + M_b}{B + M_b + M_s + M},$$

which decreases continuously with the money amount M wagered (risked) by selling those shares (the terms M_b and M_s are again as defined above).

The unit average price obtained over a total wager of amount $M = m$ is then

$$\begin{aligned} \overline{b} &= \frac{1}{m}\int_0^m \frac{Bp + M_b}{B + M_b + M_s + M}\delta M \\ &= \frac{1}{m}\left\{(Bp + M_b)\left[\log(m + B + M_b + M_s) - \log(B + M_b + M_s)\right]\right\}. \end{aligned}$$

The amount m risked by the trader who sells shares to MM equals $n(1 - \overline{b})$, that is $m = n(1 - \overline{b})$. Hence

$$m = n\left[(1 - \frac{1}{m}\left\{(Bp + M_b)\left[\log(m + B + M_b + M_s) - \log(B + M_b + M_s)\right]\right\}\right]. \tag{28.8}$$

The Bid price $(n - m)$ for some arbitrary number n shares is then found by substituting that number n into (28.8) and solving this equation numerically for m. In Mathematica the best method again is to use the *FindRoot* function initiating the search at $m = n/2$. This computation method leads to the same Bid prices as found from (28.6), but as for (28.7) the computation does not require numerical integration and is quicker and more accurate. The Bid prices for an order of $n = 1\{10\}(n = 100)$ $[n = 1{,}000]$ shares (with $B = 100$, $p = 0.5$, and $M_b = M_s = 0$ as assumed above), when calculated from (28.8), are 0:499 {0.488}(0:390) {0.131} per share. Note again the small discrepancies between the last two of these results and those found using (28.6).

5 Illustrated Example

Suppose that MM initiates the market with $B = 5$ and a probability estimate $p = 0:3$. Suppose also that the money amounts received so far from traders are $M_b = 25$ and $M_s = 5$. The weight of money received from buyers rather than sellers suggests that the market believes that the probability of E is higher than $p = 0.3$. Indeed the current instantaneous probability of event E is

$$\begin{aligned} q &= \frac{Bp + M_b}{B + M_b + M_s} \\ &= \frac{5(0.3) + 25}{5 + 25 + 5} = 0.757. \end{aligned}$$

The current bid and ask price schedules offered by the MM for any order size up to $n = 200$ shares are as shown in figure 28.1 (cf. figure 28.2). These prices are found by solving (28.5) and (28.6) or, alternatively, (28.7) and (28.8). Note that prices behave consistently with the Glosten and Milgrom (1985) and Easley and O'Hara (1987) explanation of the bid-ask spread. Specifically, a buy (sell) order increases (decreases) the market probability of E, and this price change is more pronounced the larger the order size. Note also that as the accumulated value $M_b + M_s$ of preceding trade increases the spread becomes narrower, promoting still further volume and further narrowing of the spread. For example, in the case of $B = 1$ and $p = 0.3$, suppose now that $M_b = 250$ and $M_s = 50$.

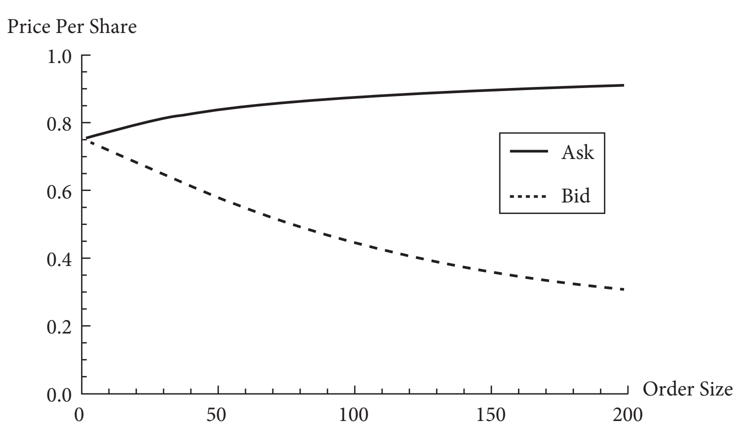


FIGURE 28.1 Bid and ask prices for orders of given size with $B = 5$, $p = 0:3$, $M_b = 25$ and $M_s = 5$

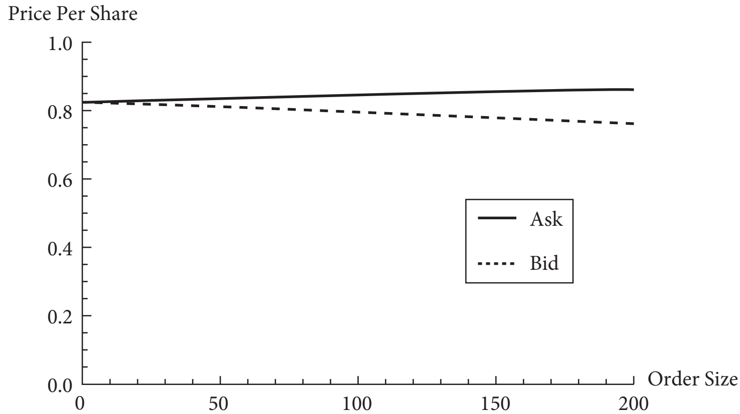


FIGURE 28.2 Bid and ask prices for orders of given size with $B = 1$, $p = 0.3$, $M_b = 250$ and $M_s = 50$

The current market probability then is

$$q = \frac{Bp + M_b}{B + M_b + M_s}$$
$$= \frac{5(0.3) + 250}{5 + 250 + 50} = 0.825.$$

The corresponding bid and ask price schedules are as shown in figure 28.2.

6 Market Properties

The MM described here has one main advantage over the Hanson market maker and one clear disadvantage. Its advantage is that the MM can set the opening share price at any probability p rather than at 0.5, which is the Hanson requirement for binary assets. Its most obvious disadvantage is that the potential or worst-case MM loss, accruing over any future sequence of trade, is not limited to a mathematically known quantity. By comparison, the Hanson market maker can lose at most, over any arbitrary trade sequence, a known factor of B, specifically $B\log(2)$ in the case of a binary security.

MM allows traders to buy and sell shares at given prices, which in effect allows them to make fixed-odds bets. For example, if a trader buys $n = 10$ shares for 5.12, then the implicit odds are 5.12 to 4.88 in favor of $V = 1$, since the bet of amount 5.12 will return a total of 10 if it wins (if $V = 1$).

Unlike each of the individual traders, MM does not know how much money will be won or lost by the end of trading when all bets have been made and settled. The amount of possible profit or loss to MM can be calculated at any time during trading based on what bets have been taken already, but not until all trades have taken place can the MM calculate how much will be won or lost conditional on V. Maximum losses depend, that is, on the trades that are yet to occur and which cannot be predicted.

Potential losses are generally tempered, however, by the way that prices react to order size. For example, if there is a strong wave of buying the automated price per share quickly goes toward one, and hence the possible losses on the recently sold shares approach zero per share (meaning that the possible gains approach one). The biggest losses occur on those shares that are bought from MM at prices near zero or sold to MM for prices near one. These prices can exist only under conditions that in and of themselves tend to constrain losses. Specifically, the Ask price might be near zero only if there is little trade from buyers or if there is much trade with sellers, implying the possibility of small losses (in the event of $V = 1$) at worst. Similarly, the Bid price can be near one only when there is little trade with sellers or when there is much trade with buyers, implying again only small losses in the event of $V = 0$, at worst. The other important limiting factor is that trade volume may not be so disproportionate in either direction that prices ever get near zero or one, particularly in applications where the common perception is that E has probability nowhere near zero or one (i.e., there is no near certainty about whether E will occur or not).

The ultimate MM safeguard is of course to close off trading in one direction if trading has been too heavily in that direction and potential losses have climbed beyond a tolerable level. This happens sometimes in betting markets but is far from desirable, as it defeats the purpose of an automated market maker, namely, to provide liquidity and allow a trader to either buy or sell at all times. It also prevents full "price discovery," since the market probability is not allowed to move to the point that traders want to take it.

7 Markets with Known Maximum Loss

The MM described next is a variant MM* of the one above, modified to have predictable maximum possible loss without contemplating any need for a market shut down. To meet this requirement the modified MM prices trades so that all trades preceding the current trade are covered and the market makers' only exposure is to the last trader.

The method is explained in the following example. Let $B = 100$ and $p = 0.5$. Suppose that the first trade executed by MM* was for x shares at total price C, where negative x represents a sale of shares by MM giving positive C and positive x represents a purchase by MM* giving negative C. The modified MM* requires each new trader to inherit MM*'s position (x, C) existing after the last trade and then to make whatever incremental trade is required to reach the desired net inventory purchase or sale.

Assume that the first trade is a sale of 30 units, and hence MM*'s position after this trade is $(x,\ C) = (-30,\ 16.0912)$. The sale price of 16.0912 (0.536 per unit) is found from the ask price equation (28.7) with $M_b = M_s = 0$. The second trader assumes a trading account of −30 shares and cash of 16.0912 and hence takes over MM*'s position resulting from the previous trade. Suppose that the second trader wants to sell 50 shares. The trader must now sell just 20 shares to MM since he or she starts with −30 (inherited from MM*). The price for these shares is found by substituting $n = 20$ in the bid price equation (28.8). It is important to note that M_b and M_s are again (always) set at zero, since the net effect of previous trade by MM* is passed on to the current trader. The total price received by the second trader is therefore 9.50949 (calculated from (28.8) with $n = 20$) plus the cash balance of 16.0912 inherited from MM*. Thus the net price received from MM for 50 units is 25:6007 (0.512 per unit).

The MM*'s position after the second trade is $(x, C) = (20, -9.50949)$. Suppose now that the third trader arrives and wants to buy 100 units. This trader must first take on the MM*'s position of $(x, C) = (20, -9.50949)$. To achieve a net inventory position of 100, this trader must buy just 80 more units. The price of these is 47.2402, found by solving (28.7) with $n = 80$ and $M_b = M_s = 0$. In effect, the trader ends up paying a total of $47.2402 + 9.50949 = 56.7497$ for a net purchase of 100 units (0.5675 per unit).

In general, the modified MM logic is as follows. A trader who wants to take a position of y units (where y can be negative) inherits x units from MM* and then buys another $y - x$ units from MM*. Implicitly, if $y - x$ is negative, then the trader in fact sells $-(y - x)$ units to MM*. Upon completion of this trade, the change in the trader's cash position is $-A + C$, where A is the price of the units bought from MM (when units were in fact sold to MM, A is negative).

The results of the example calculations outlined above are summarized in table 28.2 along with results for a further arbitrary sequence of trades. Table 28.2 gives the unit price for each trade alongside the corresponding price found using the Hanson market maker (see Pennock 2006 for details of how to calculate the Hanson prices). The two automated market maker prices are comparable only because we have assumed that MM* sets $p = 0.5$, as is implicit for Hanson. The two unit prices are quite similar in all trades, with the proviso that MM* reacts more slowly than the Hanson price to volume (albeit in the same direction of course). This presents no clear advantage either way. By reacting less sharply to orders, MM is "more liquid" than Hanson (with the same B), and by reacting more quickly the Hanson market maker puts a tighter constraint on maximum possible losses.

To visualize the difference between the MM* and Hanson prices, consider the plots in figures 28.3 and 28.4 of the bid and ask prices existing at the instant that trade 5 (in table 28.2) takes place. Note that the instantaneous prices (probabilities) at that moment are slightly different, as a result of different "interpretations" of the preceding trades, and that the Hanson prices approach zero and one a little more sharply with order size.

By approaching probability limits less quickly than Hanson, MM* risks exacerbating losses, since traders can acquire higher volumes before prices become prohibitive. MM* is designed so that all trades executed before the last trade are self-financing.

Table 28.2 Example Unit Prices for Trades in Sequence Assuming $B = 100$ and $p = 0.5$

	Trader	MM Price	Hanson Price	(x, C) after Trade
1	Buys 30	0.536	0.537	(−30, 16.0912)
2	Sells 50	0.512	0.512	(20, −9.509)
3	Buys 100	0.568	0.572	(−80, 47.2402)
4	Buys 50	0.708	0.740	(−130, 82.6199)
5	Buys 50	0.768	0.824	(−180, 121.02)
6	Sells 100	0.738	0.782	(−80, 47.2402)
7	Buys 20	0.686	0.711	(−100, 60.9591)
8	Buys 400	0.846	0.923	(−500, 399.325)
9	Buys 1000	0.952	0.999	(−1500, 1351.55)
10	Sells 1500	0.901	0.954	(0, 0)

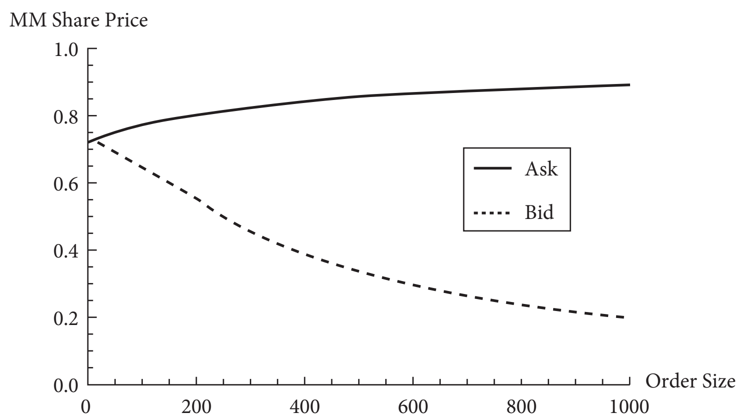


FIGURE 28.3 MM* bid and ask prices for orders of given size with $B = 100,\ p = 0.5$

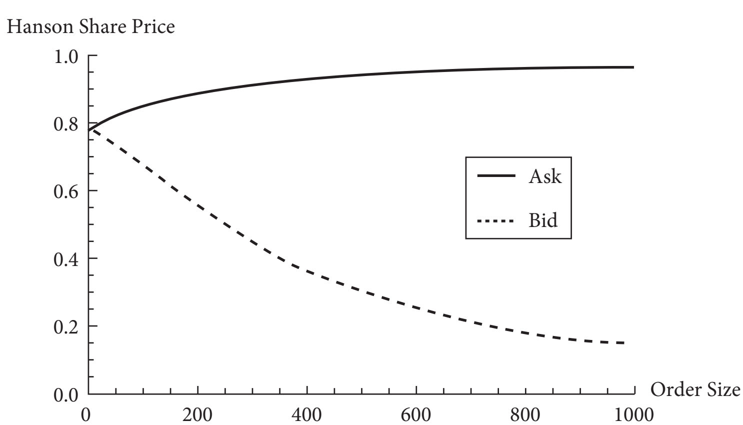


FIGURE 28.4 Hanson bid and ask prices for orders of given size with $B = 100,\ p = 0.5$

Specifically, each new trader takes the position held by MM* after the preceding trade and then makes an incremental trade against MM* to achieve the intended position. That incremental trade is the only source of possible loss to MM*.

The incremental trade is priced using (28.7) or (28.8), depending on whether the trader buys or sells (with $M_b = M_s = 0$). This final trade gives rise to a maximum possible loss that increases with order size, albeit at a decreasing rate as the instantaneous market probability is pushed toward one or zero. Figure 28.5 shows the maximum possible loss for MM* under the assumptions of $B = 100$ and $p = 0.3$. Mathematically the possible MM* loss when the last trader's incremental trade is a purchase of n shares at average price $\bar{a}$ for total amount $M = m$ equals

$$\begin{aligned} n(1-\bar{a}) &= \frac{m}{\bar{a}}(1-\bar{a}) \\ &= m\left(1/\bar{a}-1\right), \text{where } a = \frac{1}{m}\int_0^m \frac{Bp+M}{B+M}\delta M \\ &= m^2/\left\{m-(1-p)\left[\log(B+m)-\log(B)\right]\right\}-1. \end{aligned} \tag{28.9}$$

The instantaneous probability implied by this incremental trade is

$$q = \frac{Bp+m}{B+m}. \tag{28.10}$$

Solving (28.9) and (28.10) so as to eliminate m yields a unique solution for the maximum possible loss as a function of B, p and q.

$$B\left(\frac{(p-q)(1-p)\log[(1-q)/(1-p)]}{(1-q)(1-p\log[(1-q)/(1-p)]-(p-q)}\right). \tag{28.11}$$

Now consider the maximum possible MM* loss when the last trader's incremental trade is a sale of n shares to MM* at average price $\bar{b}$. In this case the amount $M = m$ wagered (risked) by the trader equals $n(1-\bar{b})$, implying that $n = m/(1-\bar{b})$. Hence the amount that MM* can possibly lose is

$$\begin{aligned} n\bar{b} &= m\bar{b}/(1-\bar{b}), \text{where } \bar{b} = \frac{1}{m}\int_0^m \frac{Bp+M}{B+M}\delta M \\ &= \frac{mBp\left[\log(B+m)-\log(B)\right]}{m-mBp\left[\log(B+m)-\log(B)\right]}. \end{aligned} \tag{28.12}$$

The associated implicit instantaneous probability is

$$q = \frac{Bp}{B+m}. \tag{28.13}$$

Solving (28.12) and (28.13) simultaneously to eliminate m gives the following unique result for the maximum possible MM* loss as a function of q:

$$B\left(\frac{p(q-p)\log\left[p/q\right]}{q-p+pq\log\left[p/q\right]}\right). \tag{28.14}$$

Figure 28.5 shows a plot of the maximum possible MM* loss as a function of q, with $B = 100$ and $p = 0:3$. The two segments of this function are (28.11) for $q > 0:3$ (when the trader buys) and (28.14) for $q < 0:3$ (when the trader sells). Here q is the instantaneous probability implied by the incremental trade made by the last trader.

Note that it is quite possible that MM* can lose a multiple of B, especially if the final trader's assessment of the probability of E is very high, encouraging the trader to buy shares up to a price somewhere near one. Strictly, the feasible MM* loss is infinite because there is no limit to how many shares might be bought at prices approaching (but never reaching) one or sold at prices approaching zero. In most practical applications it is unlikely if not unthinkable that the incremental trade will be so large that the implied q is higher than say 0.99 or lower than 0.01 This level of certainty will rarely arise naturally in any commercial context, and subjective probability assessments close to the MM's prior p will often be far more realistic. Losses are limited in such circumstances to a small factor of B. Such losses are possibly warranted economically on the grounds that traders bring forward opinions and information to the market in exchange for the MM risking such a loss. In real-money experimental markets the experimenter can x parameter B so that the maximum plausible loss as a factor of B is not beyond the funds available for the experiment.

Note that although the possible loss under MM* is a larger factor of B than under Hanson, B can be set at a lower amount under MM because the market price under MM* is slower in reacting to trades than under Hanson (for a given B). By setting

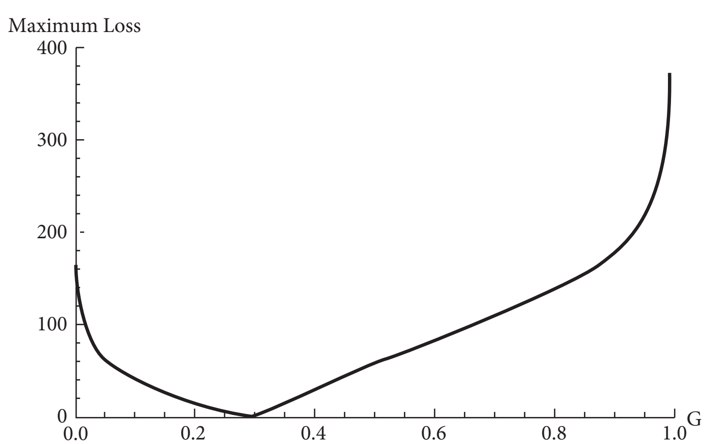


FIGURE 28.5 Maximum market maker loss as a function of q assuming $B = 100$, $p = 0.3$

B lower, this price reaction can be made to closely approximate the Hanson prices and lead to typically very similar market maker losses (or gains) after a given set of trades. The problem of how to choose an appropriate value of *B* has been made very clear by Chen and Pennock (2010) and is virtually the same problem for MM* as for Hanson.

8 Conclusion

In some market applications the MM may not be prepared to risk accruing losses. There are two ways to think about this. The first is that in most prediction markets contexts the potential for the MM to lose is justified by the information obtained from traders through operating the market. Based on this view, losses are good because they are repaid with information and tend to attract traders (especially inside traders) to play and, thus, to impart greater information. The second outlook, applicable in contexts where a MM wants to profit in aggregate or on average from all the trade attracted, is that a commission can be charged to all trades, or even just to winning trades, so as to tip the balance such that the MM is generally (if not always) profitable. Much thought has been directed toward resolving this issue by those involved in prediction market design. See, for example, the recent discussions by Othman et al. (2010) and Chen and Pennock (2010).

References

Agrawal, Shipra, Erick Delage, Mark Peters, Zizhuo Wang, and Yinyu Ye. 2011. A Unified framework for prediction market design. *Operations Research* 59(3):550–568.

Berg, Joyce E., Forrest D. Nelson, and Thomas A. Rietz. 2008. Prediction market accuracy in the long run. *International Journal of Forecasting* 24(2):285–300.

Chen, Yiling, and David M. Pennock. 2010. Designing markets for prediction. *AI Magazine* 31(4):42–52.

Chen, Yiling, Chao-Hsien Chu, Tracy Mullen, and David M. Pennock. 2005. Information markets vs. opinion pools: An empirical comparison. *Proceedings of the 6th ACM Conference on Electronic Commerce.* New York: ACM, 58–67.

Easley, David, and Maureen O'Hara. 1987. Price, trade size, and information in securities markets. *Journal of Financial Economics* 19(1):69–90.

Glosten, Lawrence R., and Paul R. Milgrom. 1985. Bid, ask and transaction prices in a specialist market with heterogeneously informed traders. *Journal of Financial Economics* 14(1):71–100.

Hanson, Robin. 2003. Combinatorial information market design. *Information Systems Frontiers* 5(1):107–119.

——. 2007. Logarithmic market scoring rules for modular combinatorial information aggregation. *Journal of Prediction Markets* 1(1):3–15.

Othman, Abraham, Tuomas Sandholm, David M. Pennock, and Daniel M. Reeves. 2010. A practical liquidity-sensitive automated market maker. In *Proceedings of the 11th ACM Conference on Electronic Commerce*. New York: ACM, 377–386.

Pennock, David M. 2006. Implementing Hanson's market maker. Oddhead blog; http://blog.oddhead.com/2006/10/30/implementing-hansons-market-maker.

Pennock, David M., Steve Lawrence, C. Lee Giles, and Finn Årup Nielsen. 2001. The real power of artificial markets. *Science* 291(5506):987–988.

Wolfers, Justin, and Eric Zitzewitz. 2004. Prediction markets. *Journal of Economic Perspectives* 18(2):107–126.

CHAPTER 29

THE LONG HISTORY OF POLITICAL BETTING MARKETS: AN INTERNATIONAL PERSPECTIVE

PAUL W. RHODE AND KOLEMAN STRUMPF

Election betting markets have been growing in popularity. These markets are chiefly an Internet phenomenon, leveraging the ability of a large number of participants to quickly and cheaply place wagers on the outcome of upcoming elections. The first such market, the Iowa Political Stock Market, was founded in 1988 and involved a few hundred traders playing for modest stakes. More recent incarnations, most prominently Intrade and Betfair, have thousands of traders making millions of dollars in wagers. There is strong evidence that prices in these markets provide accurate forecasts of election outcomes.[1]

The prominence of Internet election markets often obscures the long history of such markets. While it is often claimed that election markets are a recent phenomenon, we have previously documented that wagering on presidential elections has occurred in the United States for over a century.[2] In this chapter we demonstrate that such markets are even older and that betting on elections has occurred for hundreds of years in many Western countries. The twentieth century is distinguished less by the creation of betting markets than by their absence in the middle of the century.

This chapter discusses the historical evolution of the legality and microstructure of political betting markets in several countries. The structure, operation, and public prominence of these markets reflect the prevailing culture and electoral institutions. Betting focused on the most important political outcomes of the time: the choice of government officials in Italian city-states during the sixteenth and seventeenth centuries, papal selection in sixteenth-century Italy, the timing and winning party of parliamentary elections in eighteenth- and nineteenth-century Britain, the outcomes of local and national elections in nineteenth-century Canada, and presidential and congressional winners in the nineteenth- and early-twentieth-century United States. There were also markets on other political events, such as the outcome of no-confidence votes, the tenure of leaders and their successors, or the outcome of foreign/military ventures.[3]

While there were important differences across countries, several similarities emerge. First, pivotal elections energized these markets. Not only would important contests lead to greater betting activity (such as the 1916 U.S. election or the 1948 Italian race), they could even lead the markets to reemerge from extended dormancies (as with the 1964 contest in Britain). Alternatively, there was far less betting in periods of political apathy or one-party rule (such as in the 1930s and 1940s when the Fianna Fáil party dominated the government of the Irish Free State and the Republic of Ireland). Second, while newspapers were often uncomfortable reporting on domestic markets, they were less averse to printing stories on political markets abroad. This likely reflects the moral uncertainty surrounding election betting. Third, there was a general parallel between Britain and the United States in terms of a rapidly changing but generally unfavorable legal environment, and in both countries public election betting virtually disappeared around the start of World War II. The conclusion draws additional parallels and suggests how further research can build on the histories presented here.

It is important to note a potential limitation. Largely due to language issues our analysis is centered on Anglophone countries. (It might also be useful to point out that our analysis is restricted also to countries that have real elections (i.e., unlike the sham elections in communist/totalitarian countries). This might not be significant problem since political betting markets require at a minimum some form of popular vote and typically an independent media source to report the resulting prices. In the pre-twentieth-century period we focus on, these conditions were primarily found in the English-speaking world. Still we return to this issue in the conclusion.

1 Early Markets: Italian City States and the Vatican

In Italy there were historical markets on both civic elections and the papacy. Betting was common in the Italian city-states in the early modern period, 1500–1700.[4] In addition to voting, selection to public office often included intentional randomization, for example, drawing lots to name the nominators or candidates. In Venice and Genoa gambling on the outcome of such contests was popular. D. R. Bellhouse has suggested that the Genoese lottery, one of the first modern numbers games, originated with betting on the drawing of lots—pulling balls associated with specific candidates from an urn.[5] Political betting continued into Italy's recent history, including, at times, as part of its national lotto (established in 1863). As an example, in the pivotal 1948 election the state-run lottery experimented with a betting pool on the composition of parliament.

Gamblers have also long wagered on the selection to offices in the Catholic Church. Quotes of betting odds on papal succession appear as early as 1503, when such wagering was already considered "an old practice."[6] During the troubled papal conclave of 1549 the Venetian ambassador Matteo Dandolo observed that the Roman "merchants are very well informed about the state of the poll, and ... the cardinals' attendants in Conclave go partners with them in wagers, which thus causes many tens of thousands of crowns to change hands."[7] Odds were offered not only on which candidate among the "papabile" would win but also on when the conclave would end. About two months into this long and conflict-filled process, the market odds were 10 to 1 (implying a probability of approximately 9%) that this conclave would never elect a pope. Aversion to such activities eventually led Pope Gregory XIV, in March 1591, to ban on pain of excommunication all betting on the outcome of papal elections, the length of the papal reign, or the creation of cardinals.

Gregory XIV's threat pushed wagering over papal succession underground, but at times it resurfaced. As a 1878 *New York Times* article noted, "The deaths and advents of the Popes has always given rise to an excessive amount of gambling in the lottery, and today the people of Italy are in a state of excitement that is indescribable. Figures are picked out which have some relation with the life or death of Pius IX. Every day large sums are paid for tickets in the lottery about to be drawn."[8] Betting over the successor to Leo XIII in 1903 and to Benedict XV in 1922 attracted considerable press attention.[9] With the recent rise of Internet betting markets, betting on the new pope could again occur in public on a large scale.

2 Election Betting in Britain

2.1 Eighteenth and Nineteenth Centuries

Political betting also has a long history in Great Britain. As one prominent example, Charles James Fox, the late-eighteenth-century Whig statesman, was known as an inveterate gambler. His biographer, George Otto Trevelyan, noted that "(f)or ten years, from 1771 onwards, Charles Fox betted frequently, largely, and judiciously, on the social and political occurrences of the time."[10] His wagers recorded in the betting book of the Brooks' Club included whether the Tea Act would be repealed, how long Lord North's minister would last, or on other events related to the coming of the American Revolution. Newspapers in the 1760s, 1770s, and 1780s are filled with brief notes about public betting in London over events in the life of John Wilkes, the fate of the Stamp Act, and the other political outcomes.[11] Wagering took place at gentleman's clubs—such as Almack's, Boodle's, Brooks', and White's—and in the colleges of leading Universities—such as All Souls and Magdalen Colleges at Oxford and Gonville and Caius College at Cambridge—as well as in less elite public coffeehouses—including Lloyds. Such activity was considered in keeping with national tradition: "As far back as the reign of William the Third, foreigners had observed that, on matters great and small, the only sure test of English opinions was the state of the odds."[12] A common phrase was "Bet or be silent."[13]

Wagering was generally legal under British common law so long as it did not to lead to immortality or impolity.[14] Bets about the outcome of events in war, over the death of political leaders, over court cases, or between voters over election results were illegal on these grounds.[15] In the Victorian and Edwardian periods the British government increasingly attempted to limit gambling, especially among the working classes. The Gaming Act of 1845 made gambling contracts and debts unenforceable in court (but otherwise liberalized what amounts could be wagered); the Betting Houses Act of 1853 outlawed the operation of betting establishments other than private clubs; the Betting Houses Act of 1874 cracked down of the advertisement of wagering; and the Street Betting Act of 1906 made acceptance of wagers in streets and public places illegal.[16] Despite legal uncertainty in the late nineteenth and early twentieth centuries, the Fleet Street press reported on election wagering at the London Stock Exchange and at Lloyd's in markets for parliamentary "majorities."[17]

2.2 Early Twentieth Century

Election betting grew in popularity with the adoption of spread betting. In this system bets are based not simply on the winner of the election but the size of the margin (spread betting is common in political, sports, and financial markets in twenty-first-century England).

Laura Beers provides a fascinating account of the evolution of the parliamentary "majorities" market in the British Stock Exchange between 1910 and 1940 (a small spread market also existed in 1906).[18] The market differed from the American examples (see below) because wagers were placed not chiefly on which party would win but on the size of their parliamentary majority. That is, the buyer and seller agreed on a threshold (or "seat price") for the number of seats won and an amount to be paid for each seat difference between this threshold and the actual majority. For example, if the threshold is 15, the actual majority is 20, and the amount per seat is £5, then seller pays the buyer (20–15)*£5 = £25. The focus on "majorities" reflected the standard vocabulary of British political analysis.[19] No cash was initially fronted, and the "debts of honor" were settled after election day. Newspapers would report the buying and selling prices—the gap was commonly 10 seats—but not the names of the participants. Lloyd's of London also offered insurance on the election outcome.[20]

Table 29.1 summarizes the election markets and the actual outcomes between 1910 and 1935. The first market to gain substantial attention off the trading floor occurred in the run-up to the December 1910 election. Price quotes appeared in the financial press on nearly a daily basis. The starting and ending values of the prices were very close to the actual outcome, though there was substantial divergence in the middle of the contest.[21] There is little information about the operation of "market for majorities"

Table 29.1 Spread Bets on pre-WW2 British Parliamentary Elections

	Dec.1910	1922	1923	1924	1929	1931	1935
A. Final Prices							
Conservative				272			
Liberal				97			
Labor		100		182.5	245		
Coalition Majority	130						
Conservative Majority			34				
(Nat. Gov't) Majority						204	169
B. Election Outcomes							
Conservative	271	344	258	412	290	473	386
(Nat. Gov't)						-556	-**430**
Liberal	272	62	158	40	59	33	21
Labor	42	142	191	151	287	53	154
Other	85	65	8	12	9	59	55
Actual Majority	122		−99			503	243

Prices: listed values are mid-points in the bid-ask spread; values correspond to seat totals except in rows where majority is indicated

Election Outcome: Actual Majority corresponds to the party or coalition for which there was a Majority price listed in the top of the table

for the elections of December 1918 and November 1922,[22] but the market on the December 1923 election was apparently the largest to date, with over £100,000 changing hands. (This is the equivalent of $6.1 million in 2010 purchasing power as measured by consumer prices.)[23] The market price indicated that the Conservatives would hold a small majority. But the vote yielded a hung parliament with the Conservatives winning more seats than any other single party but effectively a minority relative to the whole. This outcome resulted in large losses for bettors taking the Conservative side and considerable squabbling over the nature of the betting contract. Specifically, some who bought the Tory side of a Conservative majority bet argued that their liability was limited when the majority reached zero. But they were made to cover the entire deficit. Over the next two elections brokers shifted to bet on the number of party seats won and not on the size of majorities.[24] Since bettors continue to be able to set the size of their per-seat wager, seat totals are also a version of spread bets.

The political situation remained unstable, and a new election was called for October 1924. Labour, which was in power, was initially expected to expand its majority. But the campaign featuring "Red Scare" tactics by the right-wing press led to the shift against Labour—a decline in support far beyond what the market anticipated—and a Conservative landslide. The next contest did not occur until May 1929. And for the first time the popular press covered the election market intensively. Both the *Daily Express* and the *Daily Mail* regularly published stock market spreads. The betting market generally favored the Conservatives, though price fluctuated significantly. One source of uncertainty was the extension of suffrage to women under the age of 30. In the popular voting the Conservatives narrowly outpolled Labour; the Liberals finished in a strong third position. However, Labour won the most seats in the hung Parliament, and its leader, Ramsay McDonald, emerged as prime minister.[25]

Labour's victory was again short-lived, as splits within the ruling coalition over responses to the Great Depression led to the call for a new election for October 27, 1931. (Recall that Britain left the gold standard in late September 1931.) The fragmentation of the Labour and Liberal parties and the creation of the National Government coalition with the Conservative party at its core led to the reemergence of a market in majorities. The market highly favored the prospects of National Government, but it did not go far enough, as Labour shed over 200 seats. The rise in the price for the National Government majority was said to help revive British financial markets. As examples, in late October 1931 the financial section of the *Daily Express* carried such headlines as "Markets More Confident on Majorities Rise" and "Foreign Money Comes Back as Majorities Rise."[26] The election betting market was very active, with "Over a Million (Pounds Sterling) Won and Lost in the City" in 1931 (this is equivalent to $72.9 million in 2010 dollars). "Nothing like it has been known before."[27]

The market's shortcoming in the 1931 race created significant problems. Because the market significantly underestimated the number of seats that the National Government would win,[28] the losses to those who bet against them were great. One prominent broker, W. A. Bignell, refused to honor his bet with another, Gower W. Elias. This led to a lawsuit, wherein Justice McCardle voided the contract under the Gaming Act of 1845.[29] In response to growing concerns that the now highly visible majorities market tainted it as a gambling institution, the Stock Exchange formally cracked down on election betting.[30]

Betting activity on the next election (November 1935) centered on the large London bookmakers, such as Ladbrokes and Seaham, rather than on the Stock Exchange. The prices on majorities continued to appear in the daily press but off the front page. Again the market favored the National Government but by too little. This did not totally end such betting in the city. There were still reports of action on the "black bourse,"[31] and in the autumn of 1940, during the battle for Britain, London brokers among others ran organized betting sweepstakes regarding how many German planes would be shot down each night. The winnings were used to fund the construction of Spitfire fighters.[32]

As table 29.1 indicates, spread betting was quite accurate in forecasting early elections but became increasingly less accurate. Beers suggests this has to do with a new set of factors shaping the vote outcome. While the 1910 contest largely involved only the Conservatives and Liberals, in the 1920s and 1930s the Fourth and Fifth Reform Acts substantially expanded suffrage, the Labour Party rose to prominence, and the Liberal Party began to splinter. The wealthy, male London-based investors who bet on the Stock Exchange lost touch with an electorate comprised of women and working-class voters.

In 1929 the *Daily Express* mused that "London has never been famous for knowing much about British politics, and the Stock Exchange has been rather notorious for knowing even less than the rest of London."[33] While absolute accuracy declined, this seemed largely due to increasingly difficult political contests in which to forecast. It is important to note that the markets still outperformed the other available forecasts from pundits, big bettors, and straw polls (scientific polls did not yet exist, making the market forecasts that much more impressive). As Beers writes, the Stock Exchange predictions "appear to be no worse, and usually slightly better, indicators" than forecasts based on polling or expert opinion. Spread bets also must forecast a more challenging outcome than traditional binary wagers: while the seat totals were faulty, the markets still managed to correctly predict the winning party or coalition in all but one election. In Beers's words, "predicting electoral outcomes in three-party first-past-the-post political systems is a notoriously tricky business."[34]

2.3 Postwar Twentieth Century: Decline and Rise

In the immediate post–World War II period public election betting in Britain appears to have slowed to a trickle. Newspapers offer only a handful of quotes regarding the 1945 and 1950 contests.[35] And in 1950 the *Economist* observed: "It is curious that in a nation devoted to gambling as the British, so little opportunity should nowadays be taken of a general election, the most sporting of all events."[36] This situation changed with time.

The modern era of open, large-scale political betting in Britain began in October 1963.[37] Following Harold Macmillan's surprise resignation as prime minister after the Profumo affair, the gambling house Ladbroke's overcame the "long-standing reluctance to make book on political events" by taking bets on his successor as leader of the Conservative party.[38] Prior to 1963 Ladbroke's had handled the political betting demands of its more gentlemanly clientele in a private election book.[39] In 1964 William Hill, the country's largest bookmaker, also "quickly reversed its earlier policy not to handle election betting."[40] By the end of that year, political betting totaled an estimated £1,000,000 (the equivalent of about $23 million in 2010 dollars.) About nine-tenths of this sum was placed on British contests, including the Wilson–Heath general election, and about one-tenth was placed on the 1964 American presidential race. Political markets represented less than 2 percent of national gambling turnover.

Several features of the modern political markets' microstructure were notable: the stakes were anonymously wagered; much of the activities focused on party odds rather than the "majorities" common in the Stock Exchange period;[41] house profit rates initially averaged about 7 percent (taking in £107 for every £100 it paid out);[42] and professional bookmakers set the fixed lines rather than accept bets in the form of pools. This last feature mattered at times when, for example, Mr. Hill set a line too favorable to a candidate he supported.[43] Odds makers such as Ron Pollard of Ladbrokes became celebrities, providing color analysis on election night television news.[44] In 1965 London bookmakers began offering odds on the German election contests. And in early 1966, with new general elections in Britain, they handled over £2,100,000 (about $44.5 million in 2010 dollars). This was purportedly the largest total ever taken on a single event.[45] British markets also opened on American elections, a good 30 years prior to the return of a legal election market in the United States. It was estimated that $100 million exchanged hands ($1,540 million in 2010 dollars) following the 1972 presidential election.[46] Despite complaints about the immorality of such wagering, the British betting public never looked back.[47]

3 Election Betting in Former British Possessions and Colonies

Similar bouts of political betting occurred in many of the British offshoots with parliamentary forms of government throughout the late nineteenth and early twentieth centuries. In countries including Australia, New Zealand, Canada, Singapore, South Africa, and the Republic of Ireland, local bookmakers and members of the stock exchanges periodically wagered over the outcome of no confidence votes, the timing of the elections, and the composition of the new majority. In the remainder of the section the betting markets in several of these countries are discussed in more detail.[48]

Ireland has had political betting markets as long as the United Kingdom. In the eighteenth century these were primarily person-to-person bets and formal markets did not exist. Prior to the Union of Great Britain and Ireland in 1801, the wagers tended to focus on political events outside of Ireland, such as the odds on the American Revolution ending, whether peace would be declared in the Anglo-Dutch War, or the election of the king of Poland.[49] There also were several bets reported on elections for the British Parliament.[50] Similar person-to-person wagers continued in the 30 years following the Union, with one addition that there were also bets on acts of Parliament related to Ireland.[51] There are no reports of betting markets starting in 1830 and continuing for the next one hundred years. It is unclear whether this is due to an absence of such bets or a censoring of newspaper articles due to the conservative mores of the Victorian era. One exception is that there was some coverage of Canadian elections in the 1890s and 1910s and of U.S. election markets in the late 1890s and early 1900s.[52] There were also nonmonetary wagers at this time, though apparently not at the scale or intensity of those in the United States, which are discussed later.[53] Betting seems to have returned during the 1920s and 1930s, with both bookmaker and person-to-person wagers on Irish elections as well as coverage of U.K. parliamentary elections in which insurance companies played a role in setting odds and offering policies.[54] Following the creation of the State of Ireland in 1937, there was some mention of election betting for both the ceremonial president as well as Parliament.[55] Political bets continued to grow in prominence with one Member of Parliament even serving as a bookmaker.[56] Wagers on both domestic and international elections were definitely present at the time of their revival in the United Kingdom in the 1960s.[57] The markets have continued to grow, up to the present day.

Election betting in Australia has existed at least since the 1940s (there were also occasional mentions of informal person-to-person betting dating back to before after the establishment of the Commonwealth of Australia in 1901).[58] The greatest activity appeared to be in such major cities as Canberra, Melbourne, and Sydney, where bookmakers as well as sporting clubs posted odds on both state and federal elections.[59] Such betting was reduced, as was newspaper coverage of them, since election wagers were illegal, with fines set by the Federal Electoral Act.[60] Despite this law, newspaper articles listed bet amounts and broad descriptions of individual bettors (including an unnamed senior cabinet minister).[61] While most bets involved modest stakes, some bettors in the 1949 federal election had stakes of £4,000 or $151,000 in 2010 dollars (The 1949 election marked the departure of the Labor Party, which would not return to office for more than 20 years). Election betting odds from other countries also were reported as a means of handicapping their races. There were reports on U.S. presidential betting odds starting in the late 1890s and continuing through the 1940s and also on U.K. parliamentary betting odds during the late 1940s and early 1950s.[62] This reporting on international odds was common in other commonwealth countries as the discussion below shows.

Election betting was also prevalent in New Zealand during the late nineteenth and early twentieth centuries. The island's newspapers did not publish the local betting odds—this was apparently illegal—but rather ran frequent admonitions against betting.[63] Freak bets, nonmonetary wagers, were common and considered harmless. There was a celebrated case involving a former New Zealand premier and future chief justice, Robert Stout, where his enemies accused him of corruption for using an agent to buy votes through election bets. That is, the agent agreed to bet with a voter who received the stake if the principal won the race.[64] The newspapers also reported about election betting in the United States (with odds), the United Kingdom (circa 1910, including bets over when the next election would be called), and Ireland in the 1930s.[65]

Several English-language newspapers in colonial Africa and Asia carried articles about election odds, chiefly recapping U.S. presidential races based on wire stories from Reuters and United Press International.[66] Singapore presents one of the more interesting cases. Under the period of colonial rule, the English-speaking expatriates used the betting markets to keep track of political events in the Western world. For example, throughout the 1900–1940 period the Singapore *Straits Times* reported odds from Western markets on papal elections, "majorities," the calling of elections in the British Parliament, the first elections in the Republic of Ireland, presidential and state elections in the United States, elections in Canada, and the Saar plebiscite.[67] Newspaper stories continued to be published on the subdued U.K. betting markets during the 1950s.[68] Following decolonization (Singapore became self-governing in 1959 and declared independence in 1963), local political betting markets arose that focused on both Singapore and Malaysian elections.[69] The members of the expatriate Chinese community participated actively in these markets. As was the case in other countries, there remained some social distaste for gambling on politics. Politicians warned that election bets in Malaysia just following independence could "pervert the electoral process and dishonestly influence the results of democratic elections."[70] Such complaints continued through the 1970s.[71]

In Canada there were many reports of betting over results in both national and local elections during the late nineteenth century. For example, the *Toronto World* had several reports on betting markets covering the 1882 and 1887 parliamentary elections, the 1886 West Quebec provincial election, and the 1885 and 1887 Toronto mayoral elections.[72] There was additional coverage of gambling on many of the parliamentary elections through 1930, with a half a million U.S. dollars bet at Montreal's markets in 1911 alone ($12 million in 2010 dollars). In addition, there were occasionally active markets on local elections.[73] While many of the bets were one-shot affairs involving prominent individuals, there were more traditional markets associated with the stock exchanges in Toronto and Montreal.[74] The *Toronto Star* provided extensive coverage of election betting in the United States, reporting New York City odds right before election day to bring its readers up to date.[75]

4 Election Betting in the United States

In this section we trace the development of American betting markets in the nineteenth through twentieth centuries. A more formal analysis of the forecasting accuracy and financial efficiency of postbellum markets is described in two companion papers.[76]

4.1 Pre-Civil War

Betting on political events was commonplace in the United States ever since the early national period.[77] Advocates of a candidate frequently offered public bets on his behalf as a standard part of the election campaign. This became an expected sign of support, even for races of lesser offices. As an example, William Cooper of Cooperstown, New York, enjoyed the strong betting backing of his friends during his race for Congress in 1796.[78] Political wagering became especially intense during the partisan conflicts of the Jacksonian era.[79] The practice, with its torch-lit parades, chanting partisans, hard cider, and captive newspapers, fit right into the campaigning spirit of this period, as most press outlets were closely tied to the political machines of either the Democrats or the Whigs. Newspapers were at the heart of much of the early betting activity.[80] Many of the election betting articles that appeared in the press were boasts or challenges rather than reports of actual wagers transacted. As one instance, "to test the sincerity" of local supporters of Gen. Jackson who "express their entire confidence in the success of their favorite candidate," John Leach issued a slate of a dozen bets in his local newspaper during the 1828 contest.[81] The *Albany Argus*, voice of the New York regency, published its own list of challenges in 1832 and 1836.[82] Similar advertisements to wager appear during most other major elections of the period.[83]

We know that it was not all bluster; real money was wagered. For example, archival records show that in late October 1832 John Nevitt of Natchez, Mississippi, placed a $960 bet on Andrew Jackson's reelection. This sum was worth the equivalent of $25,000 in 2010 dollars and was more than double what Nevitt annually paid the manager of his Clermont plantation.[84]

Such big-stakes wagering was not limited to private citizens. Politicians were often involved. In 1816 future president James Buchanan lost three tracts of land in northwest Pennsylvania on an election wager. (Oil was later discovered under these lands.)[85] As candidate for governor of New York in 1828, future president Martin Van Buren wrote to a follow politico: "Bet on Kentucky, Indiana and Illinois jointly if you can, or any two of them; don't forget to bet all you can."[86] Battles between the Jackson forces and the "Bankites" raged during the 1832 contest.[87] And in 1834 Van Buren's son, John, and friend, Jesse Hoyt, recorded making over one hundred election bets, amounting to $12,000 to $15,000 ($315,000–$394,000 in 2010 money). At this time John Van Buren was New York attorney general and Martin Van Buren was the nation's vice president.[88] As another indication of the involvement of elected officials, the Washington D.C. correspondent for the *North American* reported in early 1840: "Some heavy bets were made between members of the House, to-day, on the approaching Presidential Election."[89] Election betting in 1840 was carried on as never before.[90] The 1844 contest between Henry Clay and James K. Polk witnessed an even greater flurry of betting.[91] Press reports indicate that more than $6 million ($180 million in 2010 dollars) changed hands in New York in the 1844 contest between Clay and Polk.[92]

A debate over the information value of polls versus betting odds arose during the antebellum era. The 1824 election was an open race, with no party nomination process, several potential candidates, and more democratic electorate. Politicians and journalists were eager to gauge support for leading candidates, including John Quincy Adams, Henry Clay, and Andrew Jackson, among others. They explored different measures, such as the number of endorsements, favorable editorials, and toasts at Fourth of July celebrations. Using the magnitude and direction of betting on elections was also explicitly considered. But such wagering was judged immoral and too closely tied to electioneering propaganda to be a reliable source of information. Instead conducting and reporting on (unscientific) straw polls of potential votes became common.[93]

4.2 The Ebb and Flow of Election Betting in the Pre-Civil War Period

To provide a better sense of the ebb and flow of election betting in the antebellum period, we surveyed the historical newspapers and periodicals available in the leading online sources—African American newspapers of the nineteenth century, the

Cengage–Gale nineteenth-century U.S. newspapers, PaperofRecord.com, the Proquest American Periodical Survey and Historical Newspapers, and the Readex Early American Newspapers—for relevant articles over the 1800 to 1860 period. Our tabulation excluded articles concerning legislative action to outlaw election betting as well as those discussing nonfinancial bets and included roughly 150 articles. The cumulative distribution of this sample is displayed in figure 29.1. The sample contains a small number of articles in the first decade of the nineteenth century, but observations drop off during the so-called Era of Good Feelings (1815–1823) period. The number of articles picks up in the mid-1820s with the beginning of the Jacksonian movement and Whig reaction. The peak of activity occurs in 1840 and 1844 and then falls off again. Activity falls in the 1850s before rising during the 1860 election season.[94]

Wagering on elections became highly controversial. In 1840 Van Buren supporters charged that British gold was being invested in "bragging bets" and "buying votes" in favor of Harrison.[95] In turn, in the aftermath of the 1844 contest, the Whigs protested that a combination of gamblers favoring Polk had committed voting fraud using the winnings from election bets to defray their expenses.[96] New York governor Silas Wright complained vigorously in his 1845 message to the state legislature of "the extensive and rapidly increasing practice of betting upon elections, and the interested and selfish, and corrupting tendencies which it exerts upon the election itself." Wright urged the legislature to make election betting a criminal offense.[97] The evangelical reform movements associated with the Second Great Awakening also preached long and hard against election betting.[98] And the Illinois Supreme Court did invalidate one bet as "against public policy and the best interests of the whole country."[99] Election betting was commonly considered a form of vote buying.[100]

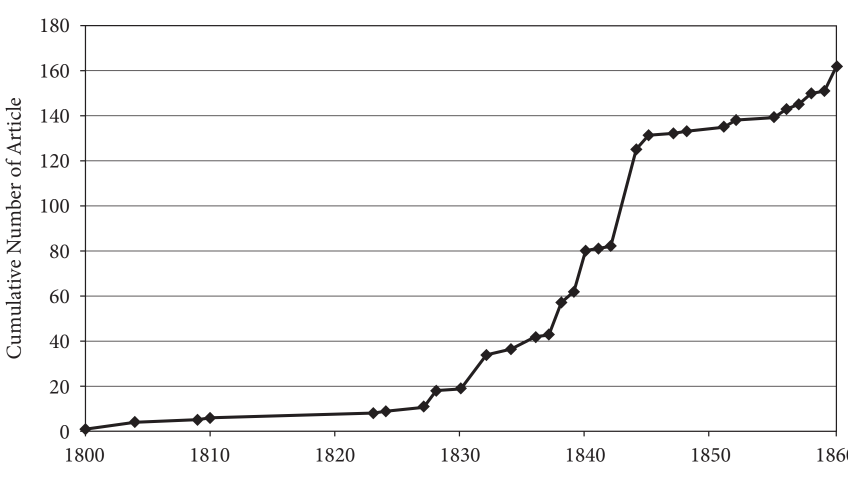


FIGURE 29.1 Cumulative distribution of articles on election betting, 1800–1860

With the collapse of the Second Party system and the ongoing Democratic–Whig rivalry, election betting appears to have slowed.[101] We can only speculate why. By the late 1840s a large number of states had made election betting illegal. The reorientation of the parties and the development of intense sectional conflicts may have reduced the sphere of personal contact leading to wagering as well as trust that the losing stake would actually be paid. Political wagering did not disappear, however, as the career of Abraham Lincoln makes clear. In 1857 his law firm handled a case involving a bet over the 1856 presidential election.[102] During the 1864 election Lincoln also apparently employed agents to entice Democrats in swing states into wagering on the election in order to disqualify their votes come election day.[103]

Much of the activity in the period surrounding the Civil War took the form of public challenges for propaganda purposes. In 1864 August Belmont, a wealthy New York Democrat and representative of the Rothschilds' interests in America, boasted that he would "bet heavily" on George B. McClellan being elected president. Belmont's terms, however, represented a conditional wager, stating that a victory for Lincoln's former general would bring peace while Lincoln's reelection would result in continued war and eventual disunion.[104] Other proposals were offered for bragging rights and were not serious wagers. An extreme example of this purportedly occurred in 1868 when New York drugstore owner H. T. Helmbold offered to bet $1 million cash at even odds to take the Democratic side on a slate of election propositions. J. Kinsey Taylor of Philadelphia, meanwhile, offered to take the Republican side headed by Ulysses S. Grant.[105] It is unclear whether both sides actually staked this wager. Such even-money boasts do not provide a meaningful set of odds concerning which candidate would win the election. But markets generating such odds would soon come.

4.3 Post–Civil War Wall Street Betting Market

Election betting involving real financial stakes occurred in almost every city, but increasingly over the postbellum period such wagering became organized in markets centralized in New York City. In the late 1860s and early 1870s activity was focused in pool halls such as Johnson's and Morrissey's. Betting in this period took the recently developed pari-mutuel form. That is, participants would buy fixed-dollar shares in the final pot and the odds would be determined at the end of all betting (a candidate's final odds of winning were determined by the proportion of the total bet volume wagered on him). The New York dailies reported substantial activity in the national and state contests of the 1870s, but the form of betting made the odds difficult to translate into subjective probabilities. In addition, problems arose with the 1876 Rutherford B. Hayes–Samuel J. Tilden presidential contest. This election was essentially a draw, with the political parties charging each other with fraudulently manufacturing votes. The House of Representative eventually decided this highly contested election. The acrimony spilled over into the betting market, where $4 million was wagered ($84 million in 2010 terms).[106] John Morrissey, the leading New York pool seller and an active

Democrat, opted to cancel the pools, returning the stakes minus his commission. This solution left many unsatisfied, a situation contributing to the push during the next session of the New York legislature to outlaw pool selling.

After a brief lull in the late 1870s and early 1880s, election betting revived in the mid-1880s and began to flourish in the 1890s. Activity moved out of pool rooms and onto the Curb Exchange in the financial district and to the major Broadway hotels. The politically connected hotels included the Republican-oriented Fifth Avenue Hotel and the neighboring Democratic/Tammany–oriented Hoffman House.[107] The Metropol and Waldorf Astoria also were locations for betting on elections. The leading bet commissioner, or stakeholder, in the public eye was Charles Mahoney, who held sway at the Hoffman House until 1910.[108] Over most of this period the standard betting and commission structure was for the betting commissioner to hold the stakes of both parties and charge a 5 percent commission on the winnings. If the commissioner trusted the creditworthiness of the bettors, it was not necessary to actually place the stakes, and instead the signed memorandum or letter of obligation sufficed.[109]

Figure 29.2 graphs the cumulative number of articles returned from online searches for "election bet" in the *New York Times* from 1851 to 1950 and the *Washington Post* from 1880 to 1950. It is clear from this figure that the heyday of election betting extended from the 1890s through the mid-1910s.[110] During the late 1890s and early 1900s the names and four-figure stakes of bettors filled the pages of New York's daily newspapers.

The environment for election bets became less favorable starting around 1910. The key developments were changes in tax laws, New York state antigambling legislation, and public attitudes toward organized financial markets. The Hart–Agnew Act outlawing professional bookmaking that employed written bets was passed by the New York

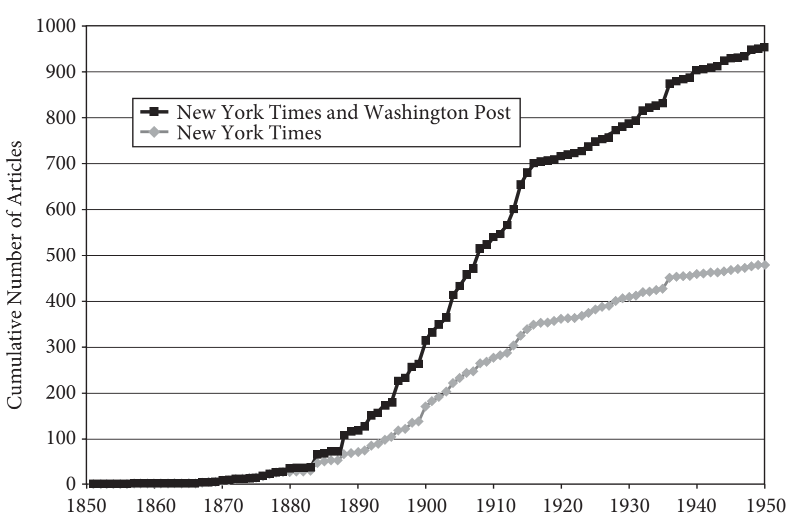


FIGURE 29.2 Cumulative "Election Bet" articles in the *New York Times and Washington Post*, 1851–1950

legislature in 1908 (and was extended to cover oral bets in 1910). The prohibition was directed primarily against horse racing and the Tammany-linked Metropolitan Turf Association, but the law's passage also reduced betting on elections for several years.

In 1912 the New York Curb Association publicly reminded its members that placing bets was contrary to New York laws. "Any member found betting, placing bets, or reporting alleged bets to the press will be charged with action detrimental to the interest of the association, which may lead to his suspension."[111] The betting commissioners in the financial district initially responded by revising their contract form—creating a memorandum between "friends" to transfer money conditional on the election outcome—and by raising the commission rates to reflect their increased legal exposure. There was some talk of moving operations to New Jersey, and many commissioners reduced or stopped keeping book.[112] When the heat was reduced after a few years, election betting revived. Ironically, in the 1916 contest between President Woodrow Wilson and Charles Evans Hughes, who as New York governor had signed the Hart–Agnew act into law, election betting on Wall Street reached its peak: $10 million (or $205 million in 2010 dollars) was wagered on the national election.

By the late 1910s newspapers more commonly published stories centering on bet commissioners and bucket shops within the financial district. (Bucketing was the practice of a broker accepting an order to buy a stock without actually executing it. The broker was essentially betting with the client about the changes in the stock's price, a bet catered to low-stakes investors.) In the early 1920s three so-called brokerages dominated election betting in the Wall Street financial district: W. L. Darnell & Co., 44 Broad Street; J. S. Fried & Co., 20 Broad Street; and G. B. de Chadenedes & Co., also of 20 Broad Street.[113] Other prominent New York bookmakers of the period included John Doyle, owner of a Broadway billiard academy, who principally handled wagers on sporting events, such as prize fights and the World Series, and Fred Schumm, a politically connected café owner in Brooklyn, who dealt in both election and sports bets.

The organized financial markets continued to attempt to limit involvement of their members. For example, in May 1924 both the New York Stock Exchange and the Curb Market passed rules/resolutions against election gambling. The exchanges liked to distinguish between their risk-sharing and risk-taking functions, which were deemed socially productive, and gambling on sporting events, such as horse races or prize fights, which were viewed as zero-sum entertainment activities with outcomes that did not affect the broader world. But unlike with sporting events, betting on elections potentially belonged in the risk-insurance category, and the information it provided had real-world value. One could readily imagine a risk averse owner of an investment project betting for a candidate unfavorable to the project in order to hedge against a "bad" election outcome. However, in practice it appears that bets were partisan in the sense that bettors took the side of their preferred candidates. Reflecting their growing marginalization, election bets became anonymous. In contrast to the earlier period, newspapers in the 1920s and 1930s no longer reported the names of those making wagers. Instead, bets were reported to involve six-figure amounts advanced by unnamed leaders in the business or entertainment worlds.

4.4 Demise of the Wall Street Election Betting Markets

The formal political betting markets appear to have largely disappeared by 1944, though informal bets continued to take place right up to the current period of Internet-based markets. There are several explanations for the demise of the Wall Street markets: (1) the rise of scientific polling, (2) the passing of several of the leading election betting commissioners, (3) the active suppression of the New York illegal gambling scene, (4) the contraction, during the early 1940s, of key sources of betting dollars, and (5) the legalization of horse race betting.

The press attention devoted to the Wall Street betting odds was due in part to the absence of creditable alternatives. In the early years of the twentieth century the only other information available concerning future election outcomes came from the results from early-season barometer contests (such as the mid-September contest in Maine), overtly partisan canvasses, and unrepresentative straw polls.[114] Over the 1894–1918 period the *New York Herald* published the results of its massive straw polls in the weeks leading up to election day. In November 1916, for example, it reported its tabulations of nearly one-quarter of a million straw ballots collected from across the country.[115] In the 1920s and 1930s *Literary Digest* issued the best-known nonrepresentative poll based on mass-mailing postcard ballots to millions of names listed in telephone directories and automobile registries. After predicting every presidential elections correctly from 1916 to 1932, the *Digest* famously called the 1936 contest for Alfred Landon, the Republican candidate, in the election that Franklin Roosevelt won by the largest Electoral College landslide ever.

The early polls based on scientific samples correctly predicted Roosevelt's victory. George Gallup, who had left academia and the advertising industry and in 1935 formed the American Institute of Public Opinion, was often credited with a singular gift of prophesy.[116] However, the polls of the other pioneers of public opinion research, including Elmo Roper, who began the Fortune Survey in 1935, and Archibald Crossley, also called the 1936 race correctly (as did the Wall Street betting odds). The numbers from scientific polls were available on a relatively frequent basis and were not subject to the moral objections against election betting. Newspapers, including the *Washington Post*, began to subscribe to the Gallup polling service and to tout its weekly results in its pages. At the same time, the paper reduced its coverage of betting markets. Such trends are displayed in figure 29.3, which reports the cumulative number of articles published in presidential election years in the *New York Times* and *Washington Post* returned from an online search of selected "poll" and "election betting" terms from 1916 to 1944.

Other factors also contributed to the demise of the Wall Street betting market. Several of the preeminent betting commissioners active in election wagering left the trade either due to death by natural causes (John Doyle) or to gang-land slayings (Sam Boston).[117] New York mayor Fiorello La Guardia's general crackdown on illegal gambling, including "raids on brokers' offices," also made it "difficult to find betting commissioners in the financial district" by 1944.[118] Tammany Hall, which had often taken the Democratic side of wagers during the heyday of New York election betting, also fell on hard times.

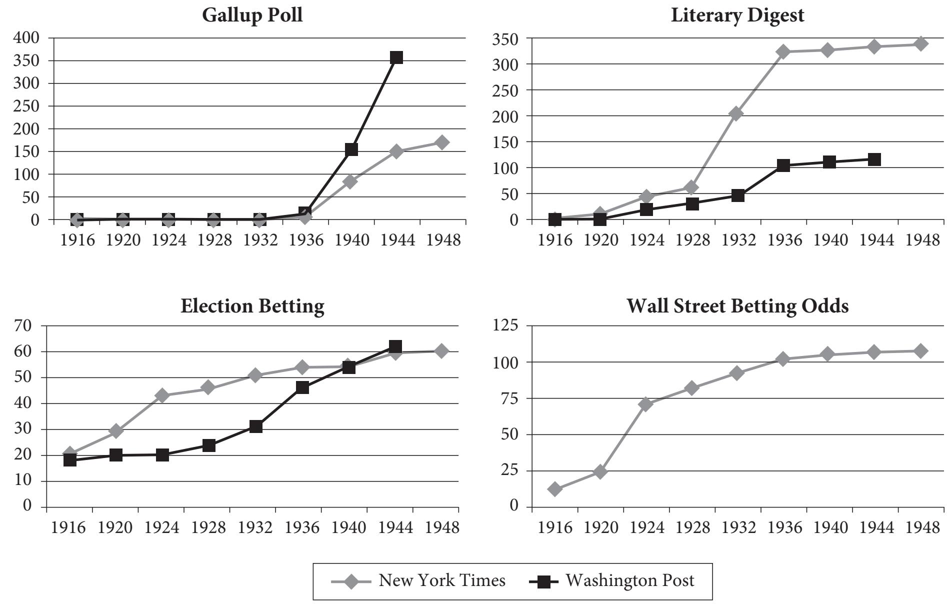


FIGURE 29.3 Cumulative number of articles returned from selected "Poll" and "Election Betting" search terms, 1916–1948

La Guardia's repeated reelection as mayor cut off much of Tammany's patronage, driving the organization to declare bankruptcy in 1943. In addition, wartime taxes were purportedly crimping the pockets on Wall Street.[119] A final factor was the legalization of horse race betting in New York in 1939. The possibility of betting several times each day at the track, rather than once or twice a year on elections, siphoned the dollars of bettors and bookmakers.

5 Conclusion

Election betting has a long history that is often characterized by higher stakes and greater emotion than what is exhibited in the Internet markets of today. Wagering was such a central cultural feature of the premodern era that even those who lacked the money to place a wager got involved. In the United States during the eighteenth and nineteenth centuries, nonfinancial bets—where losers had to roll peanuts with a toothpick down a street, climb up a greased pole, shave their hair or make other public gestures—were wildly popular. In 1900 at least half a million such "freak bets" were made.[120] Although it is sometimes claimed that political betting markets are a recent invention, our research shows that clearly they are not. Rather it is the absence of such markets during the mid and late twentieth century which is the exception.

A further comparison of the experience in different countries during this period may shed light on the factors that promoted or suppressed betting markets. For example, one could look at variations in terms of when scientific polls were introduced in different countries to see if this was a key factor in the displacement of the markets. An alternative approach would be to explain why the Internet was needed to spawn modern markets in the United States, whereas far more low-tech markets emerged a quarter century earlier in Britain. Finally, study of the rapid creation of political betting markets in previously colonized countries such as Singapore might uncover evidence of the role played by social norms inherited from the period of British rule. By gaining a better understanding of the historical dynamics of political betting markets we can begin to analyze how current developments are likely to shape and alter their current incarnations.

We are confident that future research will build on this chapter in terms of both depth (greater precision on the genesis of the markets described here) and breadth (adding discussion of other countries). One reason is technological. This work has benefited from the relatively new creation of online newspaper archives that contain coverage of historical political betting markets. As more newspaper corpora become available, a more refined and broader timeline will be possible. A second reason is the possibility of crowdsourcing. This work has been hindered by the authors' limited language proficiency. For example, we know from English-language sources that election betting periodically occurred during French elections but are unable to track its prevalence in French sources.[121] Future research involving researchers with a variety of linguistic backgrounds can expand our perspective on when and where bets have been placed on elections.

As a postscript, we note that in March 2013 the highly visible online political betting market, Intrade, closed due to unspecified "financial irregularities." The most active Intrade market at the time involved forecasting the next Pope. (FN. John Cassidy, "What Killed Intrade?" New Yorker, 11 March 2013.) Among the problems facing the site were the recent death of its founder and a crackdown by the US Commodity Futures Trading Commission on participation by American citizens. Despite this, several other online sites such as Betfair continue to provide platforms which have thick markets on political elections and other topics. Given the long history reviewed above, we anticipate that betting markets on the US Presidential race will be as active as ever in 2016. Policymakers can stand in the way of their efficient operation; they cannot push them out of existence.

Notes

1. Justin Wolfers and Eric Zitzewitz, "Prediction Markets," *Journal of Economic Perspectives* 18 (2004): 107–126; Paul W. Rhode and Koleman Strumpf, "Historical Presidential Betting Markets," *Journal of Economic Perspectives* 18, no. 2 (2004): 127–142; Paul W. Rhode and Koleman Strumpf, "Manipulating Political Stock Markets: A Field Experiment and a Century of Observational Data," working paper, June 2008; http://people.ku.edu/~cigar.

2. Rhode and Strumpf, "Historical Presidential Betting Markets."
3. There also were markets that indirectly captured election outcomes. Insurance premia, exchange rates, and security prices of politically connected assets often reflect (or span) the same fundamentals that would drive political stock market prices.
4. Jonathan Walker, "Gambling and the Venetian Noblemen, *c*. 1500–1700," *Past & Present* 162, no. 1 (1999): 28–69, esp. 31 on the practices of *scommetter*, betting on elections.
5. D. R. Bellhouse, "The Genoese Lottery," *Statistical Science* 6, no. 2 (1991): 141–148; Nicole Martinelli, "Online Gaming, Italian Style," *Wired*, Dec. 18, 2006.
6. Frederic J. Baumgartner, *Behind Locked Doors: A History of Papal Elections* (New York: Palgrave, 2003), 88, 250. See also Renaud Villard, "Le Conclave des Parieurs: Paris, Opinion Publique et Continuité du Pouvoir Pontifical à Rome au XVI^{e} Siècle,"*Annales* 64, no 2 (2009): 375–403.
7. Frederic J. Baumgartner, "Henry II and the Papal Conclave of 1549," *Sixteenth Century Journal* 16 no. 3 (1985): 301–314; quote on p. 305.
8. *New York Times*, 2 March 1878, 2.
9. *New York Times*, 11 July 1903, 2; *Atlanta Constitution*, 11 July 1903, 3; *Chicago Tribune*, 27 July 1903, 4; *Los Angeles Times*, 18 Aug. 1903, 5; *Scotsman*, 24 Jan. 1922, 4, and 7 Feb. 1922, 5. See also *Manchester Guardian*, 9 Aug. 1978, 2, which notes the role of clergymen placing bets.
10. George Otto Trevelyan, *The Early History of Charles James Fox* (New York: Harper & Brothers, 1880), 416; and *New York Times*, Nov. 7, 1880, 4.
11. *Freeman's Journal*, 3 Jan. 1763, 3; 31 Jan. 1763, 2; 11 Feb. 1763, 3; 25 Oct. 1763, 2; 23 July 1765, 2; 21 Dec. 1765, 2; 15 Feb. 1766, 2; 8 March 1766, 2; 4 Aug. 1767, 2; 19 March 1768, 2; 28 May 1768, 2; 20 Sept. 1768, 3; 18 Feb. 1769, 2; 19 Feb. 1769, 2; 21 March 1769, 2; 4 July 1769, 2; 18 Dec. 1781, 4; 20 June 1789, 4; *Finns Leinster Journal*, 14 Oct. 1772, 2; 10 April 1776, 2. Translated into percentage terms, the odds reported in *Freeman's Journal*, 8 March 1766, 2, in favor of the repeal of the Stamp Act varied between 59–64 percent in March 1766. *Freeman's Journal*, 20 Sept. 1768, 3, also reports betting in the "Court End" of Dublin.
12. Trevelyan, *Fox*, 414. For an account of partisan betting behavior in the 1837 parliamentary contest see Charles Greville and Henry Reeve, *The Greville Memoirs: A Journal of the Reigns of King George IV and King William IV*, vol. 2 (New York: Appleton, 1883), 510. See also Algernon Bourke, *The History of White's*, 2 vols. (London: Waterlow & Sons, 1892). Bourke (1:101) noted that among the British elite during the eighteenth and nineteenth centuries the "custom of deciding everything by wager is so universal" that polite conversation is fill "with little more than bet after bet, or now and then a calculation of the odds."
13. John Robinson, *The Last Earls of Barrymore, 1769–1824* (London: S. Low-Marston, 1894), 113–114.
14. Among the court cases decided on gaming were *Foster v. Thackery*, 1 Term Reports 57 (1781) regarding the outbreak of war between England and France; *Allen v. Maur*, 1 Term Reports 56 (1786), regarding an election wager between two voters; *Atherfold v. Beard*, 2 Term Reports 610 (1788), regarding the level of the duty on hops; *Lacaussade v. White*, 7 Term. Reports 535 (1798), regarding the date when England and France would sign articles of peace. British Parliament, House of Lords, *The Three Reports from the Select Committee on the Lords Appointed to Inquire into the Laws Respecting Gaming* (London, 1844), 41–42. For the legal standing of wagers see T. Starkie, "Appendix I: Substance of the Common and Statute Law Relating to Gaming," 223–231, in House of Commons,

Report from the Select Committee on Gaming; Together with the Minutes of Evidence, appendix and index (London, 1844).

15. Joseph Chitty, *A Treatise on the Laws of Commerce and Manufactures and the Contracts Related Thereto: With an Appendix of Precedents*, vol. 3 (London: A. Strahan, 1824), 82–83. In a widely publicized article, Sir Frederick Milner charged that he lost a recent parliamentary race due to effects of election betting. "Betting as a Force in Politics," *Pall Mall Gazette* (London), 6 Aug. 1886, 1.
16. David Dixon, *From Prohibition to Regulation: Bookmaking, Anti-Gambling, and the Law* (Oxford: Clarendon, 1991), 38–81; Mark Clapson, *A Bit of a Flutter: Popular Gambling and English Society* (Manchester: Manchester University Press, 1992), 18–38; Jim Orford, Kerry Sproston, Bob Erens, Clarissa White, and Laura Mitchell, *Gambling and Problem Gambling in Britain* (Hove, U.K.: Brunner-Routledge, 2003), 3.
17. "Latest Parliamentary Betting," *Punch*, 21 July 1894 (which may be meant ironically); *Times of London*, 5 Dec. 1910, 6, and 7 Dec. 1910, 12 (which are not). *Annual Register: A Review of Public Events at Home and Abroad for the Year 1910* (London: Longsmans-Green, 1911), 256–257. Wire stories about election betting were also carried in papers throughout the British Empire. As one example, Liverpool betting odds on Gladstone's 1892 prospects appear in New Zealand papers including the *Manawatu Herald*, 7 July 1892, 2; *Feilding Star*, 5 July 1892, 2; *Bush Advocate*, 5 July 1892, 3, *Poverty Bay Herald*, 5 July 1892, 2, *Hawke's Bay Herald*, 6 July 1892, 3, and others.
18. Laura D. Beers, "Punting on the Thames: Electoral Betting in Interwar Britain," *Journal of Contemporary History* 45, no. 2 (2010): 282–314.
19. Pundit commentary and newspaper election contests were commonly framed in terms of majorities. Examples of newspaper prediction contests based on majorities include London *Daily Mirror*, 11 Jan. 1906, 6; 4 Dec. 1923, 2; 9 Oct. 1924, 2; and London *Daily Express*, 8 Nov. 1922, 10; and 29 Nov. 1923, 8.
20. Beers, "Punting," 282; London *Daily Express*, 28 May 1929, 1. The London *Times*, 3 Dec. 1910, 12; London *Daily Mirror*, 2 Dec. 1910, 17, and *Irish Times*, 22 Dec. 1910, 9, provide good firsthand descriptions of the early market; the *Irish Times*, 18 Oct. 1924, 3; London *Daily Express*, 13 March 1929, 13; and *Manchester Guardian*, 22 April 1929, 14, cover the later period.
21. Beers, "Punting," 285. Examples of prices in 1910 appear in the London *Daily Mirror*, 3 Dec. 1910, 17; 5 Dec. 17; 6 Dec. 13; 8 Dec., 13; 9 Dec., 13; London *Times*, 5 Dec. 1910, 6. The activity attracted attention in the United States; see *New York Times*, 11 Dec. 1910, C2.
22. Dublin's *Sunday Independent*, 25 Sept. 1921, 5, provides quotes from London insurance firms over losses resulting from the dissolution of the British Parliament in 1921.
23. Exchange values calculated on MeasuringWorth.com, www.measuringworth.com/exchange.
24. Beers, "Punting," 287–288, 291.
25. Beers, "Punting," 296.
26. London *Daily Express*, 17 Oct. 1931, 12, and 20 Oct. 1931, 12. Similar claims appear in the next contest: "Election 'Majorities' Set Market Tone," London *Daily Express*, 6 Nov. 1935, 14.
27. London *Daily Express*, 29 Oct, 1931, 1. London *Daily Mirror*, 21 Oct. 1932, 18, estimated that "three-quarters of million changed hands" in the 1931 contest.
28. *New York Times*, 27 Oct. 1931, 1.
29. London *Daily Express*, 21 Jan. 1932, 3; London *Daily Mirror*, 21 Oct. 1932, 18.

30. London *Daily Mirror*, 19 Oct. 1933, 7, 21; Beers, "Punting," 301. Summarizing the prewar situation, "Election Gambling," *Economist*, 4 Feb. 1950, 252 noted: "There was a much publicized lawsuit when a trader, unable to honour his debts, pleaded the provisions of the Gaming Act. The sequel was a ban by the Council of the Stock Exchange on all such dealings, which has been reaffirmed at each subsequent election."
31. London *Daily Express*, 31 Oct. 1935, 10; 5 Nov. 1935, 2.
32. *Toronto Star*, 26 Oct. 1940. Mike Smithson, *The Political Punter: How to Make Money Betting on Politics* (London: Harriman House, 2007), 4–5, writes of the record in the betting book of Magdalen College, Oxford, of the gentlemen's wagers between physicists James Griffiths (one of the developers of radar) and (Bernard) Rollin in August 1940 regarding the number of German planes downed each evening. Brian Howard Harrison, in "College Life, 1918–1939," writes: "All Souls SCR (Senior Common Room) regularly conducted sweepstakes on general elections between the wars"; in Brian Howard Harrison, ed., *The History of the University of Oxford*, vol. 8, *The Twentieth Century* (Oxford: Oxford University Press, 1994), 87.
33. London *Daily Express*, 19 March 1929, 12.
34. Beers, "Punting," 307, 309.
35. Quotes for the 1945 election appear on the front page of the Sydney, Australia, *Morning Herald*, 23 July 1945; the odds come from the insurance brokers or Lloyds and heavily favor the Conservatives. Odds for the 1950 contest are given in London *Daily Express*, 11 Jan. 1950, 1, with the provisos that "Only a few members in the Stock Exchange are doing business on the Election, Very unofficially" and that such election betting was banned.
36. "Election Gambling," *Economist*, 4 Feb. 1950, 252.
37. Two years earlier the 1961 Betting and Gaming Act substantially liberalized wagering on sporting events in Britain. Graham Rock, "Gambling a-gogo," London *Observer*, 29 April 2001.
38. "Odds-On Politics," *Economist*, 21 Aug. 1965, 715–716.
39. *New York Times*, 10 May 1964.
40. "Whirl in the Pools," *Economist*, 17 Oct. 1964, 273; *Manchester Guardian*, 2 Oct. 1964; *Observer*, 10 Oct. 1964; *Financial Times*, 8 Oct. 1964, 14; Smithson, *Political Punter*, 9–11.
41. In 1966, as an example, 72.5 percent of the total parliamentary betting at Ladbrokes was on general election results; 9.0 percent was on "majorities"; and 18.5 percent was on the results in specific constituencies. *Irish Times*, 22 March 1966, 8.
42. Through taxes on wagering the British Chancellor of the Exchequer earned about 6 percent on the volume of activity. *Irish Times*, 2 March 1974, 8.
43. "Odds-On Politics," *Economist*, 21 Aug. 1965, 715–716.
44. See *Manchester Guardian*, 4 Nov. 1969, for Pollard's election night appearance. The *Guardian*, 15 March 1966, 10, noted Pollard saying that "the weight of money is more accurate an indication of public opinion than anything the polls can produce. He casts a cold eye on some of the wilder swings forecast by other methods."
45. Martin Rosenbaum, "Betting and the 1997 British General Election," *Politics* 19, no. 1 (1999): 9–14.
46. "Betting on the White House race a big business," The Singapore *Straits Times*, 30 July 1972.
47. "Election Betting Scored in Britain; Wager Affect Voters and Results, Opponents Say," *New York Times*, 5 April 1966, 5; *Manchester Guardian*, 5 April.1966, 3.

48. There was some betting on local elections in South Asia in the pre-1947 era. See *Ceylon Observer* (Colombo, Sri Lanka), 14 Dec. 1911, 15. We have seen little evidence of election betting in India before the mid-1980s, but we know that "poll betting" has become commonplace. See *Times of India*, 7 Jan. 1985, 3; 23 Nov. 1989, 3; 21 May 1991, 5; 8 Nov. 1993, 9; 7 May 1996, A1; 3 Oct. 1999, 8; and "Betting: Fluctuating Fortunes," *India Today*, 31 May 1991, 58–59.
49. *Finns Leinster Journal*, 10 April 1776, 2; *Freemans Journal*, 18 Dec. 1781, 4; 5 Oct. 1763, 2.
50. *Freemans Journal*, 19 March 1768, 2; 14 Oct. 1772, 2; 25 March 1784, 4.
51. There were bets on the timing of elections in the French Chamber of Peers (*Freemans Journal*, 5 March 1819, 2), the military success of Admiral Horatio Nelson (*Freemans Journal*, 23 April 1801, 2), and Parliamentary acts that would grant greater rights for Catholics (*Freemans Journal*, 26 April 1828, 2).
52. The Canadian market covered included bets on both the winner and the number of seats in the Parliament of Canada: *Irish Times*, 6 March 1891, 5; 5 Sept. 1911, 7. The U.S. markets are reported in *Irish Times*, 22 Dec. 1865, 3; 30 July 1904, 10; 8 Nov. 1904, 5; 7 Nov. 1905, 1; 3 Nov. 1906, 8; 3 Nov. 1908, 8; 6 Nov. 1916, 5; 4 Nov. 1924, 5; *Irish Independent*, 7 Nov 1905, 5.
53. *Irish Times*, 27 Jan. 1906, 12.
54. Irish election bets on the Dáil Éireann, the lower house of the Irish Parliament, are discussed in *Scotsman*, 5 Jan. 1933, 9; Irish elections odds from Irish bookies are in *Irish Independent*, 4 Sept. 1925, 12; Irish person-to-person bets are in *Connacht Sentinel*, 26 Jan. 1932, 2. Coverage of U.K. parliamentary elections include those on the winning party (*Irish Independent*, 10 Oct. 1924, 12; *Sunday Independent*, 12 Oct. 1924, 7) and majorities bets (*Freeman Journal*, 11 Dec. 1923, 2). The role of insurance companies on the U.K. bets is in *Sunday Independent*, 25 Sept. 1921, 5.
55. *Irish Times*, 15 July 1945, 1; *Southern Star*, 14 Feb. 1948, 3.
56. *Irish Times*, 15 Feb. 1964, 10.
57. Irish election coverage includes odds on individual seats, *Irish Independent*, 7 March 1966, 10; the governing party in Dáil Éireann, *Irish Independent*, 4 Feb. 1968, 6, and, 17 June 1969, 14; and local elections in Northern Ireland, *Irish Independent*, 24 Feb. 1969, 6. Coverage of the U.K. markets included odds from U.K. books like Ladbrokes, Corals, and William Hill (*Sunday Independent*, 16 Aug. 1964, 6; *Irish Independent*, 4 Oct. 1974, 1) as well as Irish books (*Irish Independent*, 5 Feb. 1974, 1). In addition to early coverage of U.S. markets in the 1940s (*Irish Times*, 20 Sept. 1940, 6), Irish books also set odds for American elections (*Irish Independent*, Apr. 1967). There was even coverage of British books' odds on Australian elections (*Sunday Independent*, 16 Nov. 1975, 1).
58. The Melbourne *Argus*, 17 Jan. 1868, 5; 13 Sept. 1917, 8.
59. The Hobart, Tasmania, *Mercury*, 7 Aug. 1948, 8; the *Canberra Times*, 26 May 1954, 1.
60. The aim of the law is to avoid voting based on the wager rather than preferences. The Hobart, *Mercury*, 29 May 1913, 6.
61. The Melbourne *Argus*, 12 Dec. 1949, 6; 13 Dec. 1949, 8.
62. For the United States seethe Hobart *Mercury*, 29 Jan. 1897, 3; the Sydney *Morning Herald*, 13 Nov. 1916, 8; the Hobart *Mercury*, 1 Nov. 1944, 2. For the United Kingdom see the Sydney *Morning Herald*, 23 July 1945, 1; the Melbourne *Argus*, 27 Sept. 1951, 3.
63. Wanganui *Chronicle*, 22 June 1875, 2; Ashburton *Guardian*, 31 Oct. 1893, 2; see the Wellington *Evening Post*, 4 Nov. 1873, 2 and 10 May 1884, 2; *Poverty Bay Herald*,

14 Dec. 1889, 2, Nelson *Evening Mail*, 30 Nov. 1899, 3; and the *Grey River Argus*, 5 Aug. 1920, 3.

64. Nelson *Evening Mail*, 28 Nov. 1893, 1; 4 Jan. 1894, 1; Marlborough *Express*, 3 Jan. 1894, 2
65. Wellington *Evening Post*, 17 Jan. 1933, 8.
66. See the Mombasa, Kenya, *East African Standard*, 12 Nov. 1910, 2; Rhodesia *Herald*, 7 Jan. 1910, 10; Buluwayo (Zimbabwe) *Chronicle*, 12 March 1914, 12. For examples of local election betting in South Africa see *Scotsman*, 13 June 1929, 9; *Tribune* (Lahore, Pakistan), 11 Nov. 1910, 4.
67. The Singapore *Straits Times*, 8 Sept. 1903; 25 March 1929 and 26 Feb. 1929; 10 Jan. 1933; 4 Nov. 1920 and 6 Nov. 1906; 29 July 1930; 15 Jan. 1935.
68. The Singapore *Straits Times*, 28 Jan. 1950, 24 Feb. 1950, 19 Oct. 1951, 20 May 1955, and 8 Oct. 1959.
69. The Singapore *Straits Times*, 18 Aug. 1959 and 10 May 1969.
70. The Singapore *Straits Times*, 27 June 1963. Other newspaper articles in which election betting is criticized appear on 20 March 1961 and 28 June 1963.
71. The Singapore *Straits Times*, 1 July 1978.
72. The issues of the *Toronto World* are: 8 June 1882 for the 1882 parliamentary election; 22 and 25 Feb. 1887 for the 1887 parliamentary elections; 1 Oct. 1886 for the West Quebec election; 6 Jan. 1885 for the 1885 Toronto mayoral election; 3, 4, and 5 Jan. 1887 for the 1887 Toronto mayoral election.
73. Canadian Parliamentary elections are discussed in *Manitoba Daily Free Press*, 5 March 1891; *New York Times*, 30 Oct. 1904, and *Toronto Star*, 1 Nov. 1904; *New York Times*, 22 Sept. 1911; *Scotsman*, 28 July 1930. There were also markets on a by-election in London Ontario; see *Toronto World*, 20 Nov. 1920; for a market on Quebec provincial election see *Winnipeg Free Press*, 19 Oct. 1939.
74. The Montreal markets are discussed in *New York Times*, 22 Sept. 1911.
75. As examples see *Toronto Star*, 2 Nov. 1896, 2 Nov. 1908, 5 Nov. 1912, 4 Nov. 1916, 4 Nov. 1924.
76. Paul W. Rhode and Koleman Strumpf, "Historical Presidential Betting Markets," *Journal of Economic Perspectives* 18, no. 2 (2004): 127–142, and "Historical Prediction Markets: Wagering on Presidential Elections," working paper, November 2003; www.unc.edu/~cigar/papers/BettingPaper_10Nov2003_long2.pdf.
77. As early examples see *Connecticut Gazette*, 17 Dec. 1800, 2; and the *Democrat*, 10 Nov. 1804, 2.
78. Alan Taylor, "'The Art of Hook & Snivey': Political Culture in Upstate New York during the 1790s," *Journal of American History* 79, no. 4 (1993): 1371–1396, 1386. Taylor considered "bets between the friends of candidates" one of four main instruments in early electioneering (1380). He noted that "rival interests strove to intimidate one another and to impress voters with bets. A bet between the supported of rival interests was an exercise in competitive self-assertion. A public bet on a candidate was an investment of reputation and honor as well as of money."
79. As examples at the beginning of this era see *Baltimore Patriot*, 9 Nov. 1824, 2; and *New-Hampshire Patriot & State Gazette* 1, Dec. 1828, 2. Glenn C. Altschuler and Stuart M. Blumin, *Rude Republic: Americans and Their Politics in the Nineteenth Century* (Princeton, N.J.: Princeton University Press, 2000), 73, hold that while political parties were not directly responsible for most election betting in the period, they did encourage the practice among their partisans.

Several noteworthy recent surveys of political history in the early national and Jacksonian period, including Sean Wilentz, *The Rise of American Democracy: Jefferson to Lincoln* (New York: W. W. Norton, 2005), and Michael F. Holt, *The Rise and Fall of the American Whig Party: Jacksonian Politics and the Onset of the Civil War* (New York: Oxford University Press, 1999), are remarkably silent about election betting.

80. Regarding the political scene in New York in the 1790s, Taylor, "The Art of Hook & Snivey," 1386, writes that "the newspaper office became a kind of brokerage house for wagers. There a gentleman could leave a note or bond indicating what he would bet on a candidate; there a rival gentleman could agree to take up that note or bond or leave one of their own. The curious could call to inquire about who had bet and how the wagers stood. Like the accumulation of nomination notices in the papers, reports of the ebb and flow of bets served as public opinion polls."
81. *New-Hampshire Statesman and Concord Register*, 20 Sept. 1828.
82. *Essex Gazette*, 25 Oct. 1828, 2; *New-Hampshire Patriot & State Gazette*, 10 Sept. 1832; and *Connecticut Courant*, 29 Aug. 1836
83. As examples see the *Spirits of the Times*, 8 Sept. and 20 Oct. 1832; the *Globe* (Washington, D.C.), 6 Oct. 1836; *Barre Gazette*, 30 Oct. 1840, 2.
84. Entries dated 27 Oct. and 12 Dec. 1832 in the John Nevitt Diary #543, Southern Historical Collection, Wilson Library, University of North Carolina at Chapel Hill.
85. Philip S. Klein, *President James Buchanan: A Biography* (University Park: Pennsylvania State University Press, 1962), 29, 434. For Buchanan's activities see also the Salisbury (N.C.) *Carolina Watchman*, 2 Nov. 1848.
86. Edward M. Shepard, *American Statesman: Martin Van Buren* (Boston: Houghton-Mifflin, 1900), 453. Calendar of the papers of Martin Van Buren, prepared from the original manuscripts in the Library of Congress (Washington, D.C.: GPO, 1910), includes references to election bets in his correspondence in 1813 (p. 21), 1826 (p. 78), 1828 (p. 93), 1834 (p. 220), 1835 (p. 245), 1836 (pp. 272, 274).
87. Augustus C. Buell, *History of Andrew Jackson: Pioneer, Patriot, Soldier, Politician, President.* 2 vols. (New York: Charles Scribner, 1904), 2:270–272.
88. William L. MacKenzie, *The Life and Times of Martin Van Buren* (Boston: Cooke, 1846), 255–256; *New-Hampshire Patriot and State Gazette*, 10 Sept. 1832, 3; *Vermont Gazette*, 6 Oct. 1832, 2; *Eastern Argus Semi-Weekly*, 22 Oct. 1832, 2; and the *Pittsfield Sun*, 25 Oct. 1832, 3.
89. *North American and Daily Advertiser* (Philadelphia), 26 Feb. 1840.
90. *The Farmers' Cabinet* (Philadelphia), 13 Nov. 1840, 2.
91. *New York Herald*, 15 Sept. 1844; *Daily National Intelligencer* (Washington, D.C.), 5 Sept. 1844; *Boston Daily Atlas*, 25 Sept. 1844; *Scioto Gazette* (Chillicothe, Ohio), 31 Oct. 1844.
92. *New-Hampshire Patriot* (Concord, N.H.), 5 Dec. 1844, [4].
93. *Daily National Intelligencer*, 6 Aug. 1824, 3; James W. Tankard Jr., "Public Opinion Polling by Newspapers in the Presidential Election Campaign of 1824," *Journalism Quarterly* 49, no. 2 (1972): 361–365; Tom W. Smith, "The First Straw? A Study of the Origins of Election Polls," *Public Opinion Quarterly* 54, no. 1 (1990): 21–36.
94. The number of newspapers covered in the online sources generally expands over time. This makes interpreting these trends somewhat problematic. The decline in the number of observations on election betting articles after 1844 is even more significant once the expansion in overall coverage is taken into consideration.

95. See *New-Hampshire Patriot*, 7 Sept. 1840, for these specific charges and 14 Sept. 1840 for a more general criticism of corrupting influences of election betting from the Democratic side. See *Pittsfield Sun*, 22 Oct. 1840, 2, for the slate of bets allegedly offered by agents of the "British Whigs."
96. David Bacon, "The Mystery of Iniquity: A Passage in the Secret History of American Politics, Illustrated by a View of Metropolitan Society," pt. 2, *American Whig Review*, July 1845; Calvin Colton, *The Life and Times of Henry Clay*, 2 vols. (New York: A. S. Barnes, 1846), 2:443. The alleged frauds against Clay echoed in Republican charges against the Tilden campaign in 1876. See *Republican Campaign Textbook for 1880* (Washington, D.C.: Republican Congressional Committee, 1880), 57.
97. MacKenzie, *Life and Times*, 205. Such messages were often mixed. In December 1838 Pennsylvania governor Joseph Ritner (Anti-Masonic party) had railed against the vogue for election betting, "the very worst and most pernicious species of gambling . . . a people is preparing for despotism when it turns the elective franchise of its highest offices into a mere subject of pecuniary speculation." *Atkinson's Saturday Evening Post*, 5 Jan. 1839, 2. But earlier in the election season the pro-Ritner Philadelphia newspaper the *Pennsylvania Inquirer and Daily Courier*, 20 Aug. 1838, offered to bet $10,000 in his favor in the race for governor against David Porter. (For a counteroffer see *Harrisburg Reporter and State Journal*, 21 Sept. 1838.) In this hotly contested election, which ended in the so-called Buckshot War, the sum wagered purportedly totaled more than half a million dollars. *Colored American*, 28 Nov. 1838.
98. *New York Evangelist*, 6 July 1839, 106; *Christian Register and Boston Observer*, 12 Dec. 1840, 200; *Christian Reflector*, 8 Aug. 1844, 125; *Christian Inquirer*, 27 Nov. 1858, 2.
99. The case involved a wager on the 1864 presidential contest. *Cleveland Morning Herald*, 8 Aug. 1871.
100. Altschuler and Blumin, *Rude Republic*, 71–72; *North American and Daily Advertiser* (Philadelphia), 23 June 1840. In criticizing the "ridiculous, immoral, and pernicious custom" of betting on elections, the *Middlesex Gazette*, 8 Oct. 1828, 2, noted that the practice was "quite prevalent in many States, but it is unfashionable in New England," adding "long may it remain so."
101. For examples of election betting in the 1850s see *Mississippian and State Gazette* (Jackson), 25 June 1852; *Vermont Watchman and State Journal* (Montpelier), 21 Oct. 1852; the *Pittsfield Sun*, 24 July 1856, and *Bangor Daily Whig and Courier*, 10 Nov. 1856.
102. Jesse W. Weik, *The Real Lincoln: A Portrait* (Boston, Houghton-Mifflin, 1923), 174–176.
103. *Irish Times*, 10 Oct. 1864, 4.
104. *New York Times*, 26 Oct. 1864, 4, 31 Oct. 1864, 43 Nov. 1864, 4, 5 Nov.1864, 4, 7 Nov. 1864, 4.
105. *Charleston Tri-Weekly Courier*, 31 Oct. 1868; *North American and United States Gazette* (Philadelphia), 30 Oct. 1868.
106. *The Teller* (Lewiston, Idaho), 2 Dec. 1876.
107. Downtown hotels, including the Fifth Avenue Hotel on Fifth Avenue at Twenty-third Street and the Windsor on Forty-six Street near where Jay Gould lived, were secondary locations for trading stocks and bonds in the mid-1890s. New York Curb Market, Committee of Publicity, 1929, 9.

108. *New York Times*, 26 March 1910, 16.
109. *New York Times*, 10 Nov. 1906, 1; 29 May 1924, 21; 4 Nov. 1924, 2; *Wall Street Journal*, 29 Sept. 1924, 13. *New York Times*, 9 Nov. 1916, 3. For the long tradition of election betting see *New York Herald Tribune*, 2 Nov. 1940, 23.
110. One indication of the predominance of election betting in the United States during this period comes from across the Atlantic. In response to the appearance of a news item about the betting odds on a 1908 by-election involving Winston Churchill, the *Manchester Guardian* (8 May 1908, 8) complained "we have definitely adopted the bad American custom of studying the betting as a guide to a forecast of election results."
111. *Wall Street Journal*, 8 June 1912, 5. In May 1924 both the New York Stock Exchange and the Curb Market passed resolutions barring their members from engaging in election gambling. Again, in late 1927, both exchanges blocked the use of "when issued" contracts to discourage gambling. *Wall Street Journal*, 23 Dec. 1927, 11.
112. *New York Tribune*, 30 Oct. 1908, 1. See also *New York Times*, 22 Oct. 1909, 1; 11 July 1912, 10; 18 July 1912, 1. Regarding changes in commission rates see the *New York Tribune*, 30 Oct. 1908, 1.
113. Two of the three (Fried and Darnell) in fact were owned jointly by Samuel Solomon (aka Sam Boston) and the Silinsky brothers (Abraham, Frank, William). Although the newspapers often referred to the odds as quotations from the Curb, the links with the New York Curb Exchange were informal at best. Frank Silinsky did have a seat on the Exchange, and Richard C. Fabb, an early publicist for the market, also worked for the Fried firm over the mid-1920s. "Bets to Exceed $5,000,000," *New York Times*, 31 Aug. 1924, 3.
114. Claude Everett Robinson, *Straw Votes: A Study of Political Prediction* (New York: Columbia University Press, 1932); Louis Bean, *How to Predict Elections* (New York: Knopf, 1948); Susan Herbst, *Numbered Voices: How Opinion Polling Shaped American Politics* (Chicago: University of Chicago Press, 1993), 69–88; Thomas B. Littlewood, *Calling Elections: The History of Horse-Race Journalism* (Notre Dame, Ind.; University of Notre Dame Press, 1998), esp. 42–45, 85–10, and 113–119.
115. *New York Herald*, 5 Nov. 1916, 1.
116. Polling the Nations, "A Brief History of Polling," presents a "potted" history of these events; http://poll.orspub.com/static.php?type=about&page=briefhistory.
117. *New York Times*, 8 Nov. 1940 14; 4 Aug. 1942 1; *New York World-Telegraph*, 11 Oct. 1944. Doyle retired and then died; Boston left the business after a close associate was killed as a result of a double-cross.
118. *New York World-Telegraph*, 11 Oct. 1944, which includes an analysis of why wagering in New York City on the 1944 Dewey–Roosevelt election contest "was extremely quiet." For coverage of La Guardia's intensified wartime campaign against gambling and vice see *New York Times*, 18 Jan. 1943, 17; 21 June 1943, 1; 4 Dec. 1943, 16; 18 Dec. 1943, 17.
119. Warren Moscow, *The Last of the Big-Time Bosses: The Life and Times of Carmine De Sapio and the Rise and Fall of Tammany Hall* (New York: Stein & Day, 1971), 24; *New York Journal American*, 18 Aug. 1944; *New York News*, 1 Nov. 1944.
120. E. Leslie Gilliams, "Election Bets in America," *Strand Magazine*, February 1901, 185–191. For an examination of freak bets as rituals see Mark Brewin, "The Freak Bet and

the Performance of the Democratic Paradox," *Communication Review* 9, no. 1 (2006): 37–62. *Irish Times*, 9 Feb. 1901, 9.

121. London *Daily News*, 5 July 1871. See also *New York Times*, 3 Dec. 1887, 1; 28 June 1894, 1; 18 Jan. 1895, 1. The 1887 article highlighted "the large number of betting agencies started in the streets near the Chamber" of Deputies in Paris.

SECTION VII

LOTTERIES AND GAMBLING MACHINES

CHAPTER 30

THE EFFICIENCY OF LOTTERY MARKETS

DAVID FORREST AND O. DAVID GULLEY

Introduction

Each day millions of people around the world spend the equivalent of millions of dollars on various lottery games. In many jurisdictions, a clear majority of adults buy tickets (59% in the United Kingdom, according to Wardle et al. 2011). So many people purchase such tickets that, in 2010, worldwide sales of lottery games reached about $245 billion.[1]

Why do so many people play lottery games? At first glance it does not seem rational to buy a ticket to a game which nearly always features an expected value of much less than the price of the ticket, where the odds of winning the large prizes are extremely long, and where the vast majority of players in a given drawing do not win any type of prize. Other forms of gambling, such as horse racing and casino games, offer a much higher expected value, relative to the price of playing, far higher odds of winning prizes, and a larger proportion of bettors winning at least a small prize. Above and beyond these advantages, other forms of gambling usually offer obvious non-pecuniary benefits to the players, such as the fun and excitement of watching a race or competing against other players or the attraction of a convivial venue.

So why do so many people play? A possible reason is for investment purposes. Lottery games offer generally poor payouts in that takeout rates are around 50% as compared to casino-type games, which offer payouts of around 90 percent or even higher. Moreover, the vast majority of bettors win nothing. The investment motive for play thus does not appear very attractive at first glance. Yet one type of lottery game—lotto—offers a potential return of millions of dollars on an investment of pocket change and is the only route to vastly higher wealth for most people. Paying such high grand prizes generally implies a high degree of skewness in the returns. As discussed below, skewness is often a desired feature of a variety of gambling opportunities.

Another reason people play lottery games is for the fun and entertainment involved. While it is not as obvious as with other forms of gambling (consider the exciting atmosphere of a racetrack, for example), lottery games do offer their own form of non-pecuniary benefits. As noted by Jonathan Simon (1998b), when players buy lotto tickets to a game with a massive jackpot, they are buying a dream. Thus for up to a few days players can daydream about how they would spend the winnings, what they would tell their boss at work, and so forth. Lotto tickets are also quite cheap and convenient to buy. In the United States, for example, there are around 150,000 sales outlets that sell state lottery tickets (Matheson and Grote 2005). These outlets are in grocery stores, convenience stores, gasoline stations, and other similar outlets. Further, proceeds of lottery games are often designated for spending on such things as education, recreation, arts, and so on Players may value supporting these endeavors.

A third explanation for lottery play, irrationality on the part of bettors, will be discussed below. Appealing to player irrationality to explain the existence of a $245 billion industry is, well, unappealing. We contend there is sufficient evidence for the behavior of lottery players being consistent with rationality that there is no necessity to resort to the story that they must just be "stupid." In short, there must be some good reasons why people spend so much money on lottery games.

Assuming that players are rational utility maximizers, the combined utility from the investment and entertainment components are enough to induce many people to play lottery games. This chapter is concerned with how players use information relevant to lottery games. As detailed below, such information includes the odds of winning a particular game, the value of prizes, how prizes are distributed across winning tickets, how other players behave, and the alternative gambling options that compete for players' money.

Why is it important to understand bettor behavior with regard to lotteries? First, state-sponsored lottery games continue to expand both in terms of the number of games offered by each state and by the number of states offering games. Expansion of lottery games is driven by the revenue generated from lottery games. Second, lotteries have been heavily criticized on moral, ethical, and policy grounds for taking advantage of ill-informed consumers who do not understand the true odds of winning and how little money (relative to other forms of gambling) is returned to players as prizes. Finally, the nature of lottery games offers economists an excellent opportunity to study decision-making under uncertainty and how consumers process relevant information in making playing decisions. Our focus will be on the extent to which lotto games are efficient because, as discussed below, lotto games offer several particularly interesting features.

An Example of a Typical Lotto Game

To win the grand prize in a typical lotto game a player buys a $1 ticket (or the local equivalent of a modest, round sum of money). To win the grand prize jackpot players must correctly match 6 numbers drawn randomly without replacement from, say, 49 numbers. This is called a 6/49 game. Smaller prizes are awarded for matching fewer than six numbers and, sometimes, a bonus number. Lottery operators retain around fifty percent of each dollar bet (give or take 10%), some of which covers operating costs and the rest of which is turned over to the state.[2]

Lotto games are a pari-mutuel game, which means that there can be multiple winners. Winning ticket holders share equally in the grand prize. If the jackpot is not won on a given draw, the jackpot is rolled over into the jackpot of the next drawing. Multiple rollovers over several draws can create very large jackpots. Smaller prizes are also (usually) pari-mutuel, but there is almost never a rollover for smaller prizes because the odds are low enough so that smaller prizes are nearly always paid out.

In deciding whether to purchase a lotto ticket for a given drawing, a potential bettor is confronted with a complex probability problem. To evaluate whether or not the bet is utility maximizing, the bettor must evaluate the characteristics of the ticket. These include the cost and convenience of buying the ticket along with the expected value, risk, and skewness of the returns to the gamble.

Although betters in general may not know how to calculate the expected value of a ticket, they are aware that, as the jackpot grows, a larger payout is available with no change in the relevant probabilities. Thus they perceive that the value for money (i.e., the expected value) of a wager has been increased. In fact, demand modeling exercises track drawing-by-drawing sales closely when they model sales as a function of expected value, so it seems reasonable to assume that potential players behave as if they make decisions based on expected value where expected value is serving as a proxy for buyer perception of value for money.

The expected value of a $1 lotto ticket depends on several factors: the structure of the game, the amount of the previous jackpots (if any) rolled over into the current jackpot, and the number of tickets purchased in the current drawing. Generally, the expected value is:

$$EV = [p]^*[JACKPOT]^*[SHARE] + EV_s, \tag{30.1}$$

where p is the probability of winning the jackpot, *JACKPOT* is the value of the jackpot,[3] *SHARE* is the expected proportion of the jackpot a winner will keep, and EV_s is the expected value of smaller prizes.

As can be seen from the equation above, to calculate the expected return on a ticket bettors must understand (or at least act as if they do) the probabilities of winning the various prizes, how sharing the jackpot might affect the expected value, and how the smaller prizes contribute to the expected value.

From equation (30.1), we can see why lotto games are conducive to testing for efficiency. First, the structure of the game is such that bettors have a lot of the information required to compute the expected value. The probabilities of winning the various prizes are known. The size of the jackpot is not known perfectly because it depends on how many tickets are purchased. However, bettors do know whether the previous jackpot has been won or not. If it has not been won, then they know the value of the rollover amount. Players have previous draws from which to infer the implications of a given rollover for the size of the jackpot, and in any case, many lottery operators also provide an estimate of the size of the jackpot.[4] Second, rollovers can cause jackpots to skyrocket into the hundreds of millions of dollars. As shown below, large jackpots raise the expected value of the typical lotto ticket, but the entry fee remains the same. The variation in expected value caused by rollovers allows economists to examine how players change their behavior in response to new information. Third, lotto games are pari-mutuel in that players share the grand prize and most of the smaller prizes. This feature requires bettors to use information about, and forecast the behavior of, other players. Fourth, the value of the asset (the ticket) has a known value once the drawing takes place—many financial assets do not have known terminal values, which makes it more difficult to assess the final outcome of participants' decisions. Finally, data from many lotto games around the world are available so that researchers can examine the behavior of players in many settings.

Victor Matheson (2001), among others, has developed a more formal equation for the expected value of a single ticket for a given drawing:

$$EV = \left[\sum^{i} w_i V_i + \frac{\left(AV_j\right)\left(1 - e^{-Bw_j}\right)}{B} \right] (1 - \theta) + \left(\sum^{i} w_i + w_j \right) \theta \tau \qquad (30.2)$$

where w_i and w_j are the probabilities of winning non-jackpot prizes and the jackpot prize, respectively; V is the various values of the non-jackpot prizes; AV is the advertised jackpot prize; B is the number of other ticket buyers for the drawing; θ is the marginal tax rate on any winnings; and τ is the price of the ticket.

If players are risk neutral the expected value would be enough to inform the playing decision. Risk neutrality for a population of gamblers does not seem a plausible assumption, however. Players also will likely consider the risk of the gamble, loosely defined as the chance of winning nothing. If lotto players were strictly risk averse and cared only about the expected value and risk, then explaining why otherwise rational people play lotto games would be rather difficult. Players could have Friedman–Savage utility functions and be risk loving for increases in wealth offered by lotto games. This would imply a preference for positive skewness, such as would be associated with high jackpots. The attraction of skewness, referring to how the expected value is distributed across various levels of prizes, potentially outweighs the negatives of low expected value and high risk. Lotto games feature a very high level of skewness compared to other lottery games and to other gambling opportunities in that a disproportionate share of the prizes is very large and paid out to a very few grand prize winners. While we will not write them down here, the equations for these expressions are also complex. Do players use appropriate information and act as if they understand how the expected value, risk, and skewness of lotto games are determined?

Overall, lotto games offer one of the cleanest available avenues to test for efficiency in the processing of information by economic agents in the market for a financial asset.

If lotto players can process relevant information efficiently, then, since most other financial markets are conducted among investment specialists, there is at least a chance that participants there can do likewise.

Do Players Behave Rationally?

Efficient processing of information will result in efficient functioning of markets only if the behavior of economic agents, lottery players in this case, is underpinned by rationality.[5] By *rationality* we mean that players act in generally predictable ways that are consistent with economic theory and behavior in other markets. The demand for lottery games is examined later in this section (see "The Economics of Lotteries: A Survey of the Literature," by Kent Grote and Victor A. Matheson). Some of the results from lottery demand studies are relevant here, though, because they demonstrate rationality in various dimensions of behavior. First, demand curves for lotto games slope downward. That is, as the effective price of a ticket declines the quantity demanded increases. Effective price, defined as the price of a ticket less the expected value, declines when jackpots rise. Thus bettors bet more given a larger jackpot. While the estimates of price elasticity vary somewhat, the results are overwhelming consistent with a downward sloping demand curve. Levi Pérez (2011) has provided a thorough review of the literature. Second, demand is positively correlated with income—lotteries may be said to be weakly regressive from the viewpoint of assessing impact on income distribution, yet ceteris paribus measures of income elasticity tend to be positive. Third, the introduction of alternative gambling opportunities tends to affect betting on lottery games. Fourth, lottery players substitute away from other betting opportunities when the relative effective price of a lotto ticket falls.[6] Fifth, people buy more lottery tickets if transaction costs decrease.[7] Finally, bettors seem to have stable preferences in that they prefer less risk (i.e., are risk averse) and greater skewness (to a point, at least).[8]

In sum, a large literature shows that lottery players act in ways consistent with rational behavior and in ways that are consistent with other consumer goods. Such behavior indicates that players process available information and change their behavior in predictable ways when new information becomes known. We now turn to the main focus of the chapter: are lottery markets efficient? Efficiency is one step beyond rationality and requires even more sophisticated behavior on the part of players. For example, rational players will increase lotto play in response to a larger expected value. Efficiency requires bettors to act as if they understand equation (30.2) above.[9]

Do Players Use Information Efficiently?

It is crucial for properly functioning markets that available and relevant information be correctly evaluated and used by market participants when making buying and selling decisions. Markets usually contain incentives for participants to exploit information. Markets do not work well in allocating scarce resources with, for example, asymmetric information, little or no information available, or when participants do not act correctly on relevant information. Moreover, as pointed out by Ian Walker (1998) and David Forrest, O. David Gulley, and Robert Simmons (2000), studies of demand for lottery products implicitly assume that lottery markets are efficient. Financial markets are said to be efficient when relevant information is incorporated into the price of the asset. Not surprisingly, most of the academic literature on market efficiency uses financial markets as testing grounds. Financial markets offer easily available data (trading volume, high-frequency bid/ask prices, news that may affect asset prices, etc.) and incentives for market participants to exploit available information (the ability to earn abnormal profits above and beyond a normal risk-adjusted rate of return). The general finding is that financial markets are weak form efficient (asset prices fully reflect all past price data), mostly semi-strong form efficient (asset prices fully reflect all publicly available information), and usually not strong form efficient (asset prices fully reflect all information, including inside information). Many anomalies to efficiency are found but often are difficult to exploit.[10]

How can these notions of financial market efficiency be applied to lotto markets? Richard Thaler and William Ziemba (1988) have provided a starting point for answering this question. Weak form efficiency is defined by the average ticket not having a positive net expected value (i.e., the price of the ticket is greater than the expected value from equation (30.2)). Strong form efficiency is defined as all bets having an expected value of $1-t$, where t is the takeout rate. Lotto games are overwhelmingly found to be weak form efficient. Grote and Matheson (2006) studied over 18,000 lotto drawings in the United States and found very few instances where the average ticket had a positive net expected value. In other words, lotto players, when faced with large jackpots due to rollovers, tend to increase betting to beyond the point at which the expected value of the average ticket is driven below the price of the ticket.[11] Equation (30.2) shows that the reason for this result is that, while the increase in sales will increase the size of the jackpot, the likelihood of more winners sharing the jackpot also increases. The latter effect becomes more dominant as sales increase.

However, lotto games are not strong form efficient by Thaler and Ziemba's definition. There are two reasons for this finding. First, as discussed in detail below, players often do not play randomly chosen combinations; some combinations are heavily played and others are lightly played or not played at all in a given drawing. Tickets for the former combinations will have relatively low expected values, whereas tickets for the latter combinations will have higher expected values. Thus different bets in the same draw have different expected returns, a violation of strong form efficiency. Second, across different drawings, by equation (30.2), the expected value of a lotto ticket approaches $1-t$ as sales rise. "Approaches" is the key word. As equation (30.2) shows, rollovers will increase the expected value at a normal level of sales, typically to higher than $1-t$, but then the resulting increase in sales will work to reduce the expected value back toward $1-t$. However, expected value will not fall all the way back to $1-t$ (i.e., sales will not increase enough) since, as expected value closes in on $1-t$, marginal players would decline to play because the lotto would offer a relatively unattractive bet.[12]

In one sense therefore, lotto markets cannot be fully "efficient" since some draws (rollover draws) offer higher expected rates of return than others (non-rollover draws). But Frank Scott and Gulley (1995) developed a method to test for a form of efficiency using the concept of a rational expectations equilibrium. In markets where participants must forecast the future (and/or some other unknown(s)) the market is said to be in rational expectations equilibrium when the expectations of market participants match, on average, the actual outcomes. Rational expectations equilibrium of course does not imply perfect forecasts but, rather, that the forecasts are unbiased and that any forecast errors are not serially correlated.

Scott and Gulley examined four lotto games in three states (Kentucky, Massachusetts, and Ohio) and proceeded in two stages. In the first stage they calculated the expected value of tickets in each drawing using the equivalent of equation (30.2) and then regressed this expected value on all data available to bettors at the time bets are placed. The data included the value of the rollover (and in the case of the two Massachusetts lotto games the estimated value of the jackpots), a time trend, and assorted control variables.

In the second stage the difference between the actual and fitted expected values from stage one were regressed against actual ticket sales. If bettors correctly forecast sales using available information, then ticket sales should be uncorrelated with the forecast errors. The authors found that bettors, on average, correctly forecast sales and that a rational expectations equilibrium exists.

In studies of financial market efficiency trading volume (of shares, contracts, currency units, etc.) is often used as a proxy for information flow in the market. A higher volume of trading indicates more information flow, as indicated by buying and selling orders. Each of these orders represents a buyer's or seller's views about whether or not the asset is appropriately priced. In lotto markets, trading volume would be represented by ticket sales. More sales would of course indicate the view that the particular drawing was a good deal in terms of an investment and/or entertainment. In financial markets noise traders are those who trade on little or no information. A sufficient number of noise traders can cause financial asset prices to deviate from efficient values. Is there an equivalent problem with rollover drawings? Are those who buy extra tickets sold in a rollover noise traders, caught up in the excitement, who act without really understanding the nature of the bet and the expected value of the ticket? Thus we may see deviations from efficiency in rollover drawings. Scott and Gulley's findings suggest that this is not the case.

It is also possible that inefficiencies could be observed early in the history of a new lotto game, as players have not yet had much experience with the game. Forrest, Gulley, and Simmons (2000) replicated Scott and Gulley's (1995) findings for the U.K. National Lottery and showed that the U.K. National Lottery market is efficient in the sense of conformity with rational expectations. They also examined how quickly bettors learn the rules of the game, finding that efficiency is achieved with the first 30 weekly draws of the lotto game. Thus they found little evidence that inexperienced players cause the lotto market to be inefficient.

Consider one last test of the efficiency of lotto markets. U.S. state lotteries pay out jackpot winnings in annuities and advertise the undiscounted sum of the annuity payments. Advertised jackpots are therefore larger than the present value of the jackpot.[13] Are buyers fooled by this behavior? That is, by manipulating the nature of the annuity to increase its nominal value do players react by increasing their purchases of tickets? Matheson and Grote (2003) found that players are not fooled. They analyzed five states (California, Florida, Georgia, Virginia, and Washington) that lengthened the annuities associated with jackpot. This of course allows the lottery operators to advertise larger jackpots. The authors used Scott and Gulley's concept of rational expectations equilibrium in lotto markets to examine the periods before and after the annuity change. By doing so, Matheson and Grote tested whether or not bettors adjust their forecasts of expected value to account for the lengthened annuity period. They found that in all five states the residuals from the expected value equation were uncorrelated with actual sales after the annuity period change, indicating that players acted as if they were able to adjust their expected value forecasts. Such behavior is strong evidence of efficient use of information on the part of players.

Efficiency in Other Lotto Games and Other Forms of Gambling

For the reasons discussed above, this chapter focuses on lotto games. Other lottery games can also provide insight into player rationality and market efficiency. Daily numbers games offer prizes for correctly matching a randomly drawn four-digit (usually) number. Prizes may either be fixed or pari-mutuel. For fixed-prize games, daily lottery game players have no incentive to choose unpopular numbers, and as a consequence they tend to concentrate their number choices as shown by Herman Chernoff (1981) and Charles Clotfelter and Philip Cook (1989). Clotfelter and Cook compared the distributions of numbers played in the Maryland and New Jersey daily numbers games. Maryland's game features fixed prizes while New Jersey's game is pari-mutuel. They found that the distribution of numbers played by New Jersey players is much more uniform than that of Maryland players. This is strong evidence that at least some players in New Jersey understand the costs of playing popular numbers. Players of daily games also demonstrate the gambler's fallacy. Clotfelter and Cook (1989) noted that in the Maryland daily numbers game the winning number for the most recent draw sees a steep drop-off in the number of players choosing it. Play of that number does not return to normal levels for several months. However, the behavior of Maryland players is not irrational because avoiding recent winning numbers does not cost them anything because the prizes are fixed, not pari-mutuel.

Instant games offer payouts for correctly matching a certain number of images that are revealed after a player scrapes away the opaque covering on a playing card (thus they are also called scratch-off games). Lottery operators offer a variety of instant games at any one time. In these games a fixed number of tickets are printed and distributed to lottery retailers. Prizes are randomly distributed throughout the printed tickets. Grand prizes may be quite large—one U.K. instants game offered a top prize of £1 million a year for the rest of the winner's life. Given the structure of the game, the odds of winning the grand prize can change over the life of the game as winning tickets are purchased. If the large prizes are won quickly, then the odds of winning those prizes of course fall to zero. But if the large prizes are not won quickly, the odds of winning improve as more tickets are sold. Some states do not publish how many winning tickets have been claimed. This behavior has led to a least one lawsuit.[14] In a technical sense, efficiency is certain to be violated in this case because the expected value of a ticket varies according to when in the life of the game the ticket is purchased.

We end this section with a brief discussion of the efficiency of non-lottery gambling markets. There is a large literature on the efficiency of other gambling markets. See Vaughan Williams (2005) and Hausch and Ziemba (2008) and the citations therein. While it is somewhat difficult to draw conclusions from the large and disparate literature, the general findings seem to be that participants in non-lottery gambling markets do tend to process information rationally, and often efficiently, even though the nature of the bets is more complex than for lotto games. There are, though, a number of exceptions to the general conclusion of efficiency. For example, one of the best documented anomalies is the favorite-long shot bias, where bettors overbet long shots and underbet favorites. This behavior could of course be interpreted as evidence that at least some bettors are acting irrationally. However, consistent with our discussion above, the act of betting on a long shot (even when bettors know perfectly well it almost surely will not pay off) could bring sufficient utility to bettors to warrant being described as rational behavior. Consider the years of stories that a successful long shot bet would allow the winner to tell friends. Love of skewness might also make such bets attractive for financially oriented bettors. Further, once transaction costs are accounted for, many of the reported inefficiencies and the implied trading rules offer little or no potential profit.

These findings are important in our context because they are broadly in line with the results for lottery games, indicating consistent behavior across various types of gambling opportunities.

Deviations from Efficiency

The literature on financial market efficiency documents deviations from efficient outcomes. The common feature of these deviations is the potential to earn abnormal profits (profits above those expected for a given level of risk) using a trading rule. For example, the trading rule for the January effect, which holds that stock prices have a net tendency to increase in January, would be to buy stocks in December and then sell at the end of January. Repetition of this rule could conceivably generate abnormal returns over time. For our purposes, are there trading rules (i.e., consistent actions on the part of players) that could be employed to earn abnormal profits in lotto games?

First, consider the findings above related to positive expected values of lotto tickets. The point of the discussion is that such occurrences are very rare, hard to predict based on available information, and becoming less frequent. Moreover, even if positive expected value drawings could be identified, it would take a large initial stake and perhaps millennia to have a realistic hope of making a profit. See Haigh (2008) for a detailed example. As we pointed out above, the risk and skewness of returns also are important. Efficiency conditions are not violated if players decline a bet with a positive net expected value if they dislike risk and the skewness is too extreme. Consider the case of the Lotto Extra game offered by the U.K. National Lottery from 2000 to 2006. It was a standard 6/49 game that offered only one prize—the jackpot. Moreover, to enter the game, players had to first buy a ticket to the main lotto game. Despite the fact that the Lotto Extra game held many positive net expected value drawings (several times offering expected returns of more than 50%), sales dwindled as week after week no one won anything. But this is not evidence of inefficiency since players looked at risk as well. (See Forrest and Alagic 2007 for a discussion.)

Second, consider lotto mania as examined by Michael Beenstock and Yoel Haitovsky (2001). Lotto mania is said to occur when rollovers have their own impact on lotto sales, even when controlling for the size of the announced jackpot. Using the Israeli lotto game (the structure of which has undergone a number of changes), the authors found that lotto mania tends to occur after three consecutive rollovers. Such behavior would seem inconsistent with efficiency because the authors control for the size of the jackpot such that the rollover should contain no new information for players to exploit. The trading rule here would be to avoid drawings that seemed manic because the potential is for the mania to drive down the expected value of a ticket. Yet by buying a ticket a player has secured a very cheap entry pass into "part of the excitement." In such circumstances the value of the entertainment (dream) component of the lotto ticket may be enhanced, making ticket purchases more likely. Matheson and Grote (2004) also investigated the related idea of "Lotto fever" whereby the increase in sales caused by a rollover actually reduces the expected value of a ticket (i.e., players overbet). The authors examined nearly 18,000 lotto drawings in the United States and found only 11 examples of such behavior, all involving relatively large jackpots. Matheson and Grote also found that lotto fever examples became less common over time. So while lotto mania may exist, players generally do not seem to overreact to the point that the resulting bet is a relatively poor one in terms of expected value.

Third, "jackpot fatigue," also discussed by Beenstock and Haitovsky (2001). Jackpot fatigue is thought to be a possible consequence of lotto mania: over time it takes ever larger jackpots to induce a given level of sales. Such behavior may be due to bettors becoming accustomed to particular levels of prizes, and so it takes ever larger prizes for them to increase purchases. Beenstock and Haitovsky found evidence for jackpot fatigue but also showed that it wears off (i.e., bettors return to their old levels of responsiveness) after several months. Matheson and Grote (2004) also found evidence of jackpot fatigue for U.S. state lotteries.

Fourth, the halo effect, as examined in Matheson and Grote (2007), occurs when sales rise for the drawing *after* a large rollover-induced jackpot is won. Other things being equal, the expectation would be that after a large jackpot is won sales will return to pre-rollover levels. One explanation for the halo effect is that publicity surrounding the large jackpot make people more aware of the lotto game and thus generates an increase in sales even though the expected value is now relatively low. Another explanation is that new players, drawn in by the large jackpot, continue to play once the jackpot is won. The authors used the U.S. Powerball lotto and found evidence of a halo effect. They found that past jackpots did not offer any improvement to forecasts of sales, which rebuts the publicity explanation. They did find that past sales help forecasts of sales but that the effect fades quite rapidly. This suggests that new players have not become addicted to the lotto game. The authors offered a third, and more likely, explanation of the halo effect. Smaller prizes can be claimed from lottery retailers. Since sales of a rollover-induced jackpot are relatively high, when the jackpot is won many more than the usual number of small prizes also are won. These players cash in their winning tickets and reinvest a portion of their winnings in tickets for the next lotto drawing. Such behavior is rational because transaction costs are low (players are already at the retail outlet). Also, as long as players do not overbet as previously described, the behavior is consistent with efficient behavior.

Fifth, do players really care most about the expected value of a ticket when making playing decisions? Indeed, we have pointed out the difficulties presented for the average player to work out equation (30.2). Do players really use expected value when making decisions, or do they use some other, simpler method to make playing decisions? Cook and Clotfelter (1993) and Forrest, Simmons and Neil Chesters (2002) argued that expected value does not drive player decision as much as the size of the jackpot. This is an appealing idea for several reasons. First, the jackpot is a well-known value that does not require players to even attempt to estimate the expected value (or risk or skewness) of a ticket. Second, very few people are motivated to play a lotto game just because of the smaller prizes—everyone plays for the grand prize.[15] Forrest, Simmons, and Chesters (2002) estimated a two-stage demand model for the U.K. National Lottery game in the spirit of Gulley and Scott (1993), and others, and found that using the jackpot rather than expected value did a better job of fitting the sales data. These results suggest that it is the entertainment (dream) component of the utility function that is relatively more important to players. If the investment component was very important, then the expected value of smaller prizes would help explain lotto game sales. Beenstock and Haitovsky (2001) discuss, but do not show, findings that suggest that Israeli lotto players are not affected by changes in the odds of winning the grand prize. These findings are consistent with Cook and Clotfelter's (1993) and Forrest, Simmons, and Chesters's (2002) results. This begs the question: if jackpots matter and odds do not, why do lottery operators around the world pay out smaller prizes—and lots of smaller prizes at that? In the U.K. National Lottery, for example, the bottom prize (£10 for matching three of six numbers) payout is higher than the jackpot payout. The Spanish lotto pays out 10 percent of sales in the form of the smallest prizes, which are equal to the entry fee; 10 percent of tickets are in fact randomly assigned this refund.

Finally, do lottery players choose their own numbers, or are they more apt to let the computer choose random numbers for them? The answer to this question is not clear. Lotto games universally allow players to let the lottery computer generate numbers for them. But many players select their own numbers that may have personal significance and can easily be remembered and played over and over. Because players' choices of numbers are positively correlated with each other, this behavior can dramatically reduce the expected value of a ticket with a popular combination relative to other tickets. If at least some bettors choose numbers nonrandomly, then other bettors can potentially improve their expected returns by playing unpopular combinations. Thus perhaps the most interesting, and potentially compelling, deviation from efficiency is the well-documented phenomenon of "conscious selection" whereby the selection of numbers becomes correlated across players. Clotfelter and Cook (1989) provided evidence that players of daily numbers games and lotto games do not pick their numbers (and combinations) at random. Hal Stern and Thomas Cover (1989) examined the Lotto 6/49 game in Canada, where the most popular numbers are 3, 7, 9, 11, 25, and 27 while the least popular are 20, 30, 38, 39, 40, 41, 42, 46, 48, and 49. George Papachristou and Dimitri Karamanis (1998) studied number choice in the Greek 6/49 lotto, where 24, 13, 25, 9, 16, 36, and 3 are the most popular numbers and 47, 1, 43, 31, 21, 42, and 39 are the least popular. However, they concluded that playing even these numbers does not offer a realistic hope of earning an abnormal profit when such behavior leads to positive net expected value wagers. Patrick Roger and Hélène Broihanne (2007) found that the most popular numbers for the French 6/49 lotto are 7, 12, 9, 13, 11, 10, 8, 24, and 25 and that the least popular numbers are 32, 41, 39, 40, 38, and 43.

Why do people pick numbers (and, by implication, combinations) that many other people also play? Clotfelter and Cook (1989) offers a helpful starting point. The authors cite "events and dates," "personal lucky numbers," and "numbers of convenience" as methods people use to choose numbers that a lot of other people also end up choosing. "Numbers of convenience" refers to choosing numbers that form a pattern on the betting slip. They note the fact that the most popular combination in one of the Massachusetts lotto games is formed by the left-to-right diagonal of the playing slip. Another reason for nonrandom choice may be that players are unaware that others are also choosing numbers nonrandomly, thus making it more likely that many people may end up choosing the same combinations. Lottery players surely know that others are also using birthdays and other such numbers as part of their combinations, though they may well underestimate the actual number of people playing a particular combination. Players may also think they are picking numbers at random but actually are not. For example, Simon (1998a) noted the case when 133 people won the U.K. National Lottery jackpot. He mapped the winning combination on to the playing slip for the game and found that there was actually a pattern of play on the slip. People also may pick combinations that for them provide some level of utility—a combination of relatives' birthdays or, as Clotfelter and Cook (1989) put it, their own "personal lucky numbers." These may be easy-to-remember, rather than lucky, numbers—PINs used to access cash machines are probably correlated across individuals even among the nonsuperstitious. Finally, there could be underlying psychological reasons why people pick particular combinations. This last reason opens the door for a brief discussion of the psychology literature related to gambling and lotteries in particular. The reason for delving into the psychological motivations for lottery gambling is that it is crucial to understand people's motives for playing the lottery, as these motives are in turn helpful in understanding how people might react to available information that may influence the decision to participate and also their level of play.

Simon (1998a) provides a detailed discussion of nonrandom number choice by lotto players. He analyzed the combinations chosen for a single draw of the U.K. lotto game. The U.K. National Lottery provided him with summary information about the combinations played for the drawing held on October 19, 1996. Thus he has data on the number of times each combination was played but not the numbers that make up the combinations. In the lotto drawing that Simon exploits, about 87 percent of tickets sold were to players choosing their own numbers. In contrast, Grote and Matheson (2006) pointed out that about 70 percent of U.S. lotto players opt for random generation.[16] Expanding on Clotfelter and Cook (1989), Simon discusses Ellen Langer's "illusion of control,"[17] which, if accurate, would imply that bettors believe they have at least a little control over their chances of winning if they pick their own numbers. Note that even perceived control can generate player utility. Simon also discusses various heuristics that people may use to help them make decisions under uncertainty. These heuristics cause people to overestimate the likelihood of rare events, such as winning a lotto jackpot. This behavior is consistent with prospect theory, as Cook and Clotfelter (1993), among others, have noted.

As discussed above, we assume that people are rational, utility maximizers. Another possible explanation is that people who play lottery games are acting irrationally, in the sense that they don't understand the odds of winning and are throwing money away. Paul Rogers (1998) has provided a thorough examination of the psychological underpinnings of lottery play. He noted several reasons why people may choose combinations that many others choose as well. Indeed, Rogers discusses a large psychological literature which argues that gambling (and maybe especially lottery gambling) is the result of irrational beliefs on the part of players, citing in his argument many of the motivations discussed by Simon (1998a). Rogers (1998, 115) argues that "... normative accounts [such as utility from the gamble] provide a far from adequate explanation of gambling habits, and in particular, that of lottery gambling." He goes on to note that "cognitive theories of gambling assume that the core beliefs of the regular gambler are in some way flawed." An example of such behavior is the gambler's fallacy, in which events that are actually independent of one another are believed to be correlated. As applied to lottery games this fallacy argues that players think that numbers (or combinations) that have been recently drawn are not likely to be drawn again for some time. Conversely, numbers that have not been drawn in some time are "due."[18] Rogers also argues that people generally do not have a good grasp of the objective probabilities of winning various prizes, especially the jackpot. Reasonably, he points out that for all intents and purposes people have no experience with one-in-fourteen-million probability events and so have no real understanding of how truly unlikely it is to win the jackpot. However, many players now have years, or even decades, of experience playing lottery and lotto games, so it is highly likely that they have learned over time that it is really hard to win large lotto prizes even if they still do not understand what a one-in-fourteen-million event truly is. This learning is one possible reason why sales of most lotto games, controlling for other factors, tend to trend downward over time.[19]

In effect, Rogers argues that gamblers often, or even usually, do not process relevant information correctly. Rogers (and many others, of course) notes that lottery play is most heavily concentrated among groups with relatively little education and those employed in middle status occupations. Lottery play has a relatively low prevalence among professionals. His arguments are consistent with the idea that those who tend to play the lottery (or even gamble at all) are those unable to appreciate the nature of the wagers they are placing.[20] However, this conclusion is at odds with a great deal of the observed behavior of lottery players as detailed above. Indeed, in situations where most players choose numbers randomly, many of Rogers' arguments are weakened.

The general conclusion is that some players choose their own numbers, and thus combinations, that are not random but are somewhat predictable. Player choices seem to be guided by a variety of factors, ranging from the design of the playing slip to false beliefs about the nature of random numbers. If it was the case that such nonrandom selection of numbers had little or no impact on expected values of tickets or any other effect, then this behavior would not be very interesting. But it does in fact have several effects. First, the nonrandom selection of numbers reduces the probability that a jackpot is won and is instead rolled over to the next drawing.[21] More importantly, if players concentrate their bets on a relatively few combinations, those combinations will have lowered expected values, and other, lightly played combinations will have higher expected values. The magnitude of the impact on expected value can be very high. Consider Simon's (1998a) example that for the U.K. National Lottery "over 10,000" tickets play the combination of 1, 2, 3, 4, 5, 6. Assume for the sake of convenience that exactly 10,000 tickets choose this combination and, combining this figure with the actual jackpot of about £11 million, that each player would win only £1,100 if that combination is chosen. Given the probability of winning the jackpot of 1 in 13,983,816, the expected value of the jackpot portion of the ticket is nearly zero, which, given the prize structure of smaller prizes, leaves the expected value of the ticket at about 31 cents. This compares to an expected value of well over one pound for a combination that only one person (maybe at random) has chosen. In fact, 1, 2, 3, 4, 5, 6 is not even the most popular combination for the U.K. National Lottery. That is 7, 14, 21, 28, 35, 42, and 49. It is chosen "tens of thousands of times" per drawing (Simon 1998a).

Note that in the literature on financial markets deviations from market efficiency generally do not hold up once taxes and transaction costs are accounted for. Moreover, once information about the deviation comes to light, the potential for abnormal profit is generally eliminated, or at least greatly reduced, as market participants attempt to exploit the inefficiency. In lotto markets deviations from efficiency are extremely difficult, and thus costly, to exploit. As discussed above, the ability of players to exploit positive expected value situations is limited—players would need a very large initial stake and millennia to have a reasonable chance at a positive payoff from the strategy before suffering the fate of a gambler's ruin.

Despite the costs that players incur when choosing numbers nonrandomly, as long as the choice and/or the numbers themselves have sufficient utility, then such behavior is rational. It is not efficient because relevant information is not being exploited.

Conclusion

We examined the degree to which lotto markets are efficient in processing information available to market participants. While there are examples of inefficiencies to be found, the ability to exploit these to earn abnormal returns is either severely limited or nonexistent. Our conclusions are consistent with Matheson and Grote (2005, 2008) and Ziemba (2008), who also discuss the degree of rationality and efficiency in lottery markets. We conclude that lotto market participants act as if they understand the odds of winning, the risk of losing, the skewness of the returns, and how the behavior of other bettors influences the expected price of a ticket. We began this chapter by asking why and how people play lottery games and argued these questions are closely related: why people play affects how they play. Our findings are consistent with the notion that at least some players are motivated, at least in part, by the investment component. Other players are clearly motivated by the entertainment/dream component. Our findings have implications for players, lottery operators, and public policy.

What are the implications of our findings for players? First, players can materially improve the expected value of their tickets by choosing combinations that are not popular with other players. Simon (1998a) and many others have shown that it is possible for players to do better, in terms of expected value, by making sure that a randomly selected combination contains unpopular numbers. A randomly selected combination that happens to have popular numbers still may have a relatively large number of other players. So while they cannot affect the probability of a win, they can improve their payout if they do win. All the same, it is still quite difficult to earn abnormal profits. Second, players should be aware of the behaviors of lottery operators in terms of advertising and changing the structure of lotto games.

What are the implications of our findings for game operators? First, since lottery players seem to act as if they understand the rules of the game, operators may be insulated somewhat against the claim that they exploit unknowing and ill-informed customers. Second, therefore, lottery operators cannot fool players with maneuvers like changing the length of the annuity payment, so should not try to do so. Third, because lottery operators have the incentive to not publish the actual combinations that bettors play, and because some bettors themselves seem to have preferences for particular numbers/combinations, there is a material risk that at some point a very popular combination is going to win a major lotto game somewhere in the world. Having hundreds, or even thousands, of winners, each expecting a large prize, could be a public relations problem.[22] To minimize the chance of this happening, it might be worthwhile for lottery operators to consider publishing popular combinations. Lottery operators, not surprisingly, are likely to resist this idea because doing so may lead to fewer rollovers. In fact, Clotfelter and Cook (1989, 88–90) and Simon (1998a) have pointed out that lottery operators actively *encourage* their customers to choose numbers that are not random. Moreover, bettors may not be inclined to give up "their" combinations, leading to entrapment. Entrapment occurs over time as players spend more and more money on a particular combination fearing that if they stop playing it that particular combination will be a winner. See Simon (1998a). On the other hand, it is unlikely that bettors realize exactly how many other people play the most popular combinations and thus how little they will actually pocket in the event they do win the jackpot. Finally, our results suggest that the options for operators to increase sales are somewhat limited. From above, tricks like changing the annuity length will not work. Other, more feasible options include changing the face price of a ticket,[23] decreasing the takeout rate (increase the expected value), changing the odds of winning the jackpot, or changing the odds/payouts of smaller prizes (alter the skewness of the game). There are a variety of risks and limitations to pursuing these strategies. Walker and Juliet Young (2001) discuss lotto design issues with an eye to how to maximize revenue. One last option might be for a state to join a consortium of states in offering a long odds–large jackpot game. Cook and Clotfelter (1993) found a positive correlation between population and sales, implying room in the marketplace for such games. Indeed, these games (Powerball, Mega Millions, and EuroMillions, for example) feature at times massive jackpots that receive quite a bit of media attention. In relatively small market areas, multistate games outsell the local lotto game.

What are the implications for public policy? Lottery operators, at the behest of the state, usually act in ways that maximize revenue to the state. To what extent *should* lottery operators (i.e., the state) act to maximize net revenues accruing from the lottery? The focus on revenue maximization drives states to pursue behaviors that would not be acceptable in the private sector or other forms of state activity. The profit motive leads states to encourage nonrandom number selection, play down the true odds of winning large prizes, and aggressively advertise lottery games. This behavior can put states in an awkward ethical position. For example, some U.S. states run liquor monopolies, and they do not glamorize alcohol and exhort citizens to drink more, as is done in the marketing of lottery products. Clotfelter and Cook (1990, 103) suggest several alternatives to the revenue-maximizing lottery. One approach would "accommodate the widespread interest in betting on long shots without encouraging that interest." The other would "serve the interests of players as the players themselves define them." Either way removes the revenue imperative currently imposed on most lottery agencies. Pursuing these objectives, rather than revenue maximization, and providing more information to bettors in the form of greater disclosure of odds and combination choice would work to insulate lottery operators from much of the criticism leveled against lotteries.

Notes

1. See eLottery.com, www.elottery.com/markets.html.
2. Some lotteries are operated directly by the state, and others are run by private entities under contract from the state.
3. Historically U.S. state lotto games have paid out the jackpot in an annuity and the lottery operator has advertised the undiscounted value of the annuity as the jackpot. Thus in examining U.S. lotto games the present value of the annuity must be calculated.
4. There has been no study of how well lotto operators forecast jackpots and, by implication, sales. Operators would have a short-term incentive to overestimate jackpots in an attempt to increase sales. However, doing so can be damaging in the long run. See "Texas lottery considers changing jackpot calculation," LotteryPost.com; www.lotterypost.com/news/114498.
5. In most markets researchers are interested in how well both buyers and sellers process information. In lotto markets the seller (the operator) provides a perfectly elastic supply of tickets regardless of the size of the jackpot. Thus the behavior of buyers will be our focus. It is interesting to note, however, that lottery operators also seem to act in a rational manner. As a rule, operators are required to maximize revenue to the state. There are a number of game parameters that operators can choose (odds of winning various prizes, how payouts are distributed across different prizes structures, etc.). They can also choose the portfolio of lottery games offered (lotto games, daily numbers, instants, etc.). In general the findings in the literature indicate that lottery operators act in a way that is consistent with revenue maximization in that they structure the lotto game to maximize revenue to the state and that the various games offered are not cannibalizing one another, except for when multistate games are introduced into states with existing lotto games. Levi Pérez and Forrest (2010) showed that Spanish lottery game cross-price elasticities are fairly low, indicating limited cannibalization across games. Forrest, Gulley, and Simmons (2004) found similar results for the U.K. National Lottery. Grote and Matheson (2006) found that there is both a displacement effect (the introduction of a new game causes players to reduce play on existing games) and a substitution effect (the relative prices of games change, causing players to play games with the relatively higher expected value) when multistate lotto games are introduced in states with existing lotto games.
6. One strand of literature examines the displacement effects when a new form of gambling is introduced to compete with an existing form of gambling. Donald Elliott and John Navin (2002) and Stephen Fink and Jonathan Rork (2003) found that lottery revenue for U.S. states has been cannibalized by riverboat and commercial (non-tribal) casinos,

respectively. Mehmet Tosun and Mark Skidmore (2004) concluded that lottery sales in border counties typically fell substantially after neighboring states introduced large-scale slot machine gaming facilities at racetracks. Note that the reverse is also true: the introduction of lotteries can impact sales of other forms of gambling. Gulley and Scott (1989) and Richard Thalheimer and Mukhtar Ali (1995) showed that the introduction of state lotteries reduced betting on horse racing. Forrest (1999) reported a substantial decline in the football pools following the introduction of the U.K. National Lottery. Melissa Kearney (2005), though, found no effect of the introduction of state lotteries on participation rates in betting and bingo. Another strand of the literature examines substitution effects, when the price of one form or gambling changes relative to another. Using monthly data from the U.K. tax authorities, David Paton, Donald Siegel, and Leighton Vaughan Williams (2001) showed that aggregate betting tends to be lower in months when the lotto game features high prize levels. Forrest, Gulley, and Simmons (2010) found that large lotto jackpots had negative, albeit modest, impacts on horse and football betting. Collectively these findings strongly suggest that gamblers do not robotically participate in one or another form of gambling but rather actively and consciously make playing decisions that are affected by relevant information.

7. It has been documented that lottery players reinvest some of the proceeds from winning tickets. Small prizes can be claimed at lottery retailers so that transaction costs are reduced (the player is already at the retailer). See Matheson and Grote (2005) for evidence of reinvestment of winnings. However, Forrest and Gulley (2009) found no evidence of reinvestment of small prizes in the U.K. lotto game. This is also discussed further below.
8. See Walker and Young (2001). Bettor preference for skewness extends to other wagering markets. See Golec and Tamarkin (1998).
9. Throughout this chapter we abstract from the question of the playing behavior of problem and addicted lotto players. Lotteries "do not tend to be addictive for adults" (Griffiths 1999).
10. Over the past several decades many challenges have been made to the efficient markets hypothesis, most notably arising from the relatively new field of behavioral finance. See Malkiel (2003) and Shiller (2003) for a discussion. Behavioral finance uses cognitive theories in attempting to explain market inefficiencies.
11. When the expected value is greater than the price of a ticket there is incentive to "buy the pot," that is, for a person, or usually a syndicate of people, to buy tickets with all of the possible combinations so as to ensure at least a share of the jackpot. Small-scale syndicates are quite common among lotto players. The Internet allows for easy creation of larger syndicates (see, e.g., www.global-lottery-syndicates.com/category/youplayweplay). A syndicate attempted to buy up tickets to a May 1992 drawing of the Irish Lottery. Fortunately, the syndicate did have a winning ticket but had to share the jackpot with two other winners. Syndicates must buy each ticket separately—lottery operators do not allow anyone to simply write a check to cover the purchase of all combinations. For the July 12, 2011, drawing of the EuroMillions game, a rumor had a Russian syndicate working to buy a large number of tickets. Matheson (2001) Grote and Matheson (2006) discuss the possibility of buying a "trump ticket" (representing all combinations) in the situation where the expected value is greater than the cost of a ticket.
12. It is not clear that Thaler and Ziemba's concept of strong form efficiency is applicable here because the expected value of a rollover drawing is greater than the expected value of a non-rollover drawing for any finite level of sales. In a rollover drawing, expected

value approaches $1-t$ from above, while in a non-rollover drawing it approaches from below.

13. Winners do generally have the option to receive the present value of the jackpot in a lump sum. European lotto games, on the other hand, pay out the jackpot prize as a lump sum and therefore advertise this value as the jackpot.
14. See "Canada Lotto 6/49 faces lawsuit," www.worldlottery.net/news/canada-lotto-649-faces-lawsuit.asp. As information about the rate at which prizes in instants games are being won is potential useful to bettors, websites have sprung up to help players keep track of prizes won in instants games. See, for example, LottoCrawler.com.
15. This is not to imply that smaller prizes are irrelevant. See the discussion below.
16. Reasons for this difference in behavior are unclear. Perhaps relatively more U.S. players are casual players who are more likely to choose numbers randomly. It also may be the case that U.K. players have more gambling experience in general—the U.K. gambling market offers far more betting opportunities than are available in the United States. This experience yields a preference among players for choosing their own numbers. Finally, the opportunity to choose their own numbers was not introduced until March 1996, so most U.K. players had grown accustomed to choosing their own numbers.
17. See Langer (1982).
18. Jonathan Guryan and Kearney (2008) investigated a variant of the gambler's fallacy—the "lucky store effect." Using data from the Texas lottery they found that ticket outlets where winning tickets for large prizes have been sold experience subsequent increases in ticket sales that are far higher than for nearby stores. Guryan and Kearney speculated on why players expect negative serial correlation in winning numbers and positive correlation in the location of winning stores and came up with a number of possibilities, all involving choices made by humans (where to purchase a ticket, a store clerk "with good karma," and others). The point is that bettors are potentially incurring higher costs (transportation, etc.) without improving their chances of winning. Such behavior is at odds with rational behavior.
19. For an amusing discussion of probabilities surrounding the July 12, 2011, drawing of the EuroMillions game, which featured a £166 million jackpot with odds of 175 million to one, see www.williamhillmedia.com/index_template.asp?file=10859.
20. Consistent with Roger's views, the Internet is, to use the most appropriate description, littered with "how to win big on the lottery" sites. Many are in the business of selling software that purportedly helps people win lotto games. Others perpetuate various falsehoods regarding lotto games. See howtowinlotto.blogspot.com for examples of gambler's fallacy and other misconceptions regarding lotto games.
21. Farrell et al. (2000) found this to be true of the U.K. National Lottery game. They estimated a demand function that accounts for nonrandom play and found that the results with respect to the price elasticity of demand were not much different from studies that have not accounted for non-random play.
22. In a 1990 interview, the director of the Massachusetts State Lottery said that having a combination like 1, 2, 3, 4, 5, 6 win a lotto drawing would be a "nightmare."
23. Pérez and Forrest (2011) discuss how sales of the Spanish lotto game were affected by several changes in the nominal price of a ticket. It is quite infrequent, however, for lotto games to change the price of a ticket.

References

Beenstock, Michael, and Yoel Haitovsky. 2001. Lottomania and other anomalies in the market for lotto. *Journal of Economic Psychology* 22(6):721–744.

Chernoff, Herman. 1981. An analysis of the Massachusetts numbers game. *Mathematical Intelligencer* 3(4):166–172.

Clotfelter, Charles T., and Philip J. Cook.1989. *Selling hope: State lotteries in America*. Cambridge, Mass.: Harvard University Press.

———. 1990. Redefining "success" in the state lottery business. *Journal of Policy Analysis and Management* 9(1):99–104.

Cook, Philip J., and Charles T. Clotfelter. 1993. The peculiar scale economies of lotto. *American Economic Review* 83(3):634–643.

Elliott, Donald S., and John C. Navin. 2002. Has riverboat gambling reduced state lottery revenue? *Public Finance Review* 30(3):235–247.

Farrell, Lisa, Gauthier Lanot, Roger Hartley, and Ian Walker. 2000. The demand for lotto: The role of conscious selection. *Journal of Business and Economic Statistics* 18(2):228–241.

Fink, Stephen, and Jonathan Rork. 2003. The importance of self-selection in casino cannibalization of state lotteries. *Economics Bulletin* 8(10):1–8.

Forrest, David. 1999. The past and future of the British football pools. *Journal of Gambling Studies* 15(2):161–176.

Forrest, David, and Dika Alagic. 2007. The history of a lottery game that was seldom won. *Journal of Gambling Business and Economics* 1(1):57–68.

Forrest, David, and O. David Gulley. 2009. Participation and level of play in the UK National Lottery and correlation with spending on other modes of gambling. *International Gambling Studies* 9(2):165–178.

Forrest, David, O. David Gulley, and Robert Simmons. 2000. Testing for rational expectations in the UK National Lottery. *Applied Economics* 32:315–326

———. 2004. Substitution between games in the UK National Lottery. *Applied Economics* 36(7):645–651.

———. 2010. The relationship between betting and lottery play: A high frequency time-series analysis. *Economic Inquiry* 48(1):26–38.

Forrest, David, Robert Simmons, and Neil Chesters. 2002. Buying a dream: Alternative models of demand for lotto. *Economic Inquiry* 40(3): 485–496.

Golec, Joseph, and Maurry Tamarkin. 1998. Bettors love skewness, not risk, at the horse track. *Journal of Political Economy* 106(1):205–225.

Griffiths, Mark. 1999. Some comments on "health-related correlates of gambling on the British National Lottery" by Reid, et al. *Psychological Reports* 85:143–144.

Grote, Kent R., and Victor A. Matheson. 2006. In search of a fair bet in the lottery. *Eastern Economic Journal* 32(4):673–684.

Gulley, O. David, and Frank A. Scott Jr. 1989. Lottery effects on pari-mutuel tax revenues. *National Tax Journal* 42(1):89–93.

———. 1993. The demand for wagering on state-operated lotto games. *National Tax Journal* 46(1):13–22.

Guryan, Jonathan, and Melissa S. Kearney. 2008. Gambling at lucky stores: Empirical evidence from state lottery sales. *American Economic Review* 98(1):458–473.

Haigh, John. 2008. The statistics of lotteries. In *Handbook of sports and lottery markets*, edited by Donald B. Hausch and William T. Ziemba. Amsterdam: Elsevier, 481–502.

Hausch, Donald B., and William T. Ziemba, eds. 2008. *Handbook of sports and lottery markets*. Amsterdam: Elsevier.
Kearney, Melissa S. 2005. State lotteries and consumer behavior. *Journal of Public Economics* 89(11–12):2269–2299.
Langer, Ellen. 1982. The illusion of control. In *Judgement under uncertainty: Heuristics and biases*, edited by Daniel Kahneman, Paul Sovic, and Amos Tversky. Cambridge: Cambridge University Press.
Malkiel, Burton G. 2003. The efficient market hypothesis and its critics. *Journal of Economic Perspectives* 17(1):59–82.
Matheson, Victor A. 2001. When are state lotteries a good bet (revisited)? *Eastern Economic Journal* 27(1):55–70.
Matheson, Victor A., and Kent R. Grote. 2003. Jacking up the jackpot: Are lotto consumers fooled by annuity payments? *Public Finance Review* 31(5):550–567.
———. 2004. Lotto fever: Do lottery players act rationally around large jackpots? *Economics Letters* 83(2):233–237.
———. 2005. Rationality and efficiency in lotto games. In *Information efficiency in financial and betting markets*, edited by Leighton Vaughan Williams. Cambridge: Cambridge University Press, 313–329
———. 2007. Examining the "halo effect" in lotto games. *Applied Economics Letters* 14(4): 307–310.
———. 2008. U.S. lotto markets. In *Handbook of sports and lottery markets*, edited by Donald. B. Hausch and William. T. Ziemba. Amsterdam: Elsevier, 503–524.
Papachristou, George, and Dimitri Karamanis. 1998. Investigating efficiency in betting markets: Evidence from the Greek 6/49 lotto. *Journal of Banking & Finance* 22(12):1597–1615.
Paton, David, Donald S. Siegel, and Leighton Vaughan Williams. 2001. A time series analysis of the demand for gambling in the United Kingdom. Nottingham Business School discussion paper. Nottingham Business School, Nottingham Trent University.
Pérez, Levi. 2011. "The 'Who and Why' of the Demand for Lottery: Empirical Highlights from the Seminal Economic Literature." *Journal of Economic Surveys Working Papers* (2011): 1–30.
Pérez, Levi, and David Forrest. 2011. Own- and cross-price elasticities for games within a state lottery portfolio. *Contemporary Economic Policy* 29(4):536–549.
Roger, Patrick, and Marie-Hélène Broihanne. 2007. *Journal of Applied Statistics* 34(6):645–662.
Rogers, Paul. 1998. The cognitive psychology of lottery gambling: A theoretical review. *Journal of Gambling Studies* 14(2):111–134.
Scott, Frank A., and O. David Gulley. 1995. Testing for efficiency in lotto markets. *Economic Inquiry* 33(2):175–188.
Shiller, Robert J. 2003. From efficient markets theory to behavioral finance. *Journal of Economic Perspectives* 17(1):83–104.
Simon, Jonathan. 1998a. An analysis of the distribution of combinations chosen by UK National Lottery players. *Journal of Risk and Uncertainty* 17(3):243–276.
———. 1998b. Dreams and disillusionment: A dynamic model of lottery demand. In *Four essays and a note on the demand for lottery tickets and how lotto players choose their numbers*. Florence: European University Institute.
Stern, Hal, and Thomas M. Cover. 1989. Maximum entropy and the lottery. *Journal of the American Statistical Association* 84(408):980–985.

Thaler, Richard H., and William T. Ziemba. 1988. Parimutel betting markets: Racetracks and lotteries. *Journal of Economic Perspectives* 2(2):161–174.

Thalheimer,Richard, and Mukhtar M. Ali. 1995. The demand for parimutuel horse race wagering and attendance. *Management Science* 41(1):129–143.

Tosun, Mehmet S., and Mark Skidmore. 2004. Interstate competition and state lottery revenues. *National Tax Journal* 57(2):163–179.

Vaughan Williams, Leighton, ed. 2005. *Information efficiency in financial and betting markets.* New York City, New York: Cambridge University Press, 2005.

Walker, Ian. 1998. The economic analysis of lotteries. *Economic Policy* 13(27):357–401.

Walker, Ian, and Juliet Young. 2001. An economist's guide to lottery design. *Economic Journal* 111(475): F700–722.

Wardle, Heather, Alison Moody, Suzanne Spence, Jim Orford, Rachel Volberg, Dhriti Jotangia, Mark Griffiths, David Hussey, and Fiona Dobbie. 2011. British gambling prevalence survey 2010. London: National Centre for Social Research; http://www.gamblingcommission.gov.uk/PDF/British%20Gambling%20Prevalence%20Survey%202010.pdf.

Ziemba, William T. 2008. Efficiency of racing, sports, and lottery betting markets. In *Handbook of sports and lottery markets*, edited by Donald B. Hausch and William T. Ziemba. Amsterdam: Elsevier, 183–222.

CHAPTER 31

THE NATIONAL LOTTERY

JOHN LEPPER AND STEPHEN CREIGH-TYTE

Historical Background

The first National Lottery draw of the modern era took place on November 19, 1994. However, since the time of Queen Elizabeth I, running a state-sponsored lottery has been commonplace. Between 1566 and 1826, with a break between 1699 and 1710, state-sponsored lotteries were held in England either under a separate act of Parliament or by means of license from the sovereign.

The earliest state lottery in the United Kingdom was established in order to pay for improvements in facilities at the Cinque Ports or for "... reparation of the havens and strength of the Realme, and towards such other public good workes" (Ashton [1898] 2011, 223). This first Lotterie Generall consisted of 400,000 lots priced at 10 shillings each (the equivalent of £1,580 in 2009 currency) (Officer and Williamson n.d.).[1] The lottery was advertised as having no blanks; that is, every ticket was to be paid a sum of 2 shillings and 6 pence (2s 6d) and have the chance to win a prize. In addition, those purchasing 30 or more tickets would receive pensions in yearly amounts for the rest of their lives (Ewen [1932] 1972). However, only about one-twelfth of the lots were sold after more than two years. Thereupon, the conditions for the lottery were changed. Only one-twelfth of the tickets drawn had prizes on them, and those that did were one-twelfth of their originally designated value.[2] The draw of all 400,000 tickets lasted from January 11 until May 6, 1569. Prizes were generally of plate or precious objects. The Lotterie Generall was thus a method of borrowing by the state and differentially rewarding lenders by lot. The failure of the Lotterie Generall was such that the exercise was not repeated by the government until 1694.[3]

These early lotteries had a fixed number of tickets announced before the draw, and the prize draw did not take place until all tickets had been sold. Drawing of prizes took place in a public place, such as by the West Door of St. Paul's Cathedral. There were two variants on this general theme. In one type, called loan lotteries, there were no losing tickets (blanks) and the draw determined the return (whether high or low) that the subscriber to the debt obtained (Cohen 1953; Raven 1991). In the other type the draw acted like modern-day raffle in which there are winners and losers.

The demand for war financing led to the resurrection of the State Lottery in 1694 (Cohen 1953; Williams 1956; Woodhall 1964). In the Million Adventure game, subscribers (adventurers) bought a tontine annuity so that as each subscriber died his or her annuity was divided among the survivors. Annuities paid 10 percent over 16 years and were financed by duties on beer and salt. In addition, there were 2,500 prizes totaling £40,000. Large prizes were given for the first and last ticket drawn. £1 million was raised by selling 100,000 tickets at £10 each (the equivalent of £15,300 in 2009 terms) (Officer and Williamson n.d.).

Tickets were sold in books and consisted of three perforated sections. One was retained by the adventurer, one was kept in the book as a check, and the last was rolled up, secured with a silk thread and placed in a box labeled A. A second set of tickets (197,250 blank and 2,500 with prizes) was placed in another box marked B.[4] On the day of the draw the boxes were taken to Guildhall in London, and two "disinterested and fit" persons drew tickets from each box in turn (Williams 1956). In 1699, after the failure of the 1697 lottery in which only 1,763 of the 140,000 tickets were sold, Parliament made lotteries illegal. Queen Anne restarted the State Lottery, authorized by annual statute, in 1710. From then, until 1826 when it was abolished, the general features of the State Lottery remained in place. Nevertheless, during that period there many changes in the detail of the rules by which the lottery was conducted.

The State Lottery was run by the Commissioners of the Lottery appointed by the Treasury. As before, a loan of fixed amount was raised by selling annuities issued by the Treasury through the Lottery Office. After 1726 lottery tickets were called Joint Stock Lottery Annuities because the bearer of a ticket was entitled to cash the ticket for the original share of the total proceeds less the annuity interest plus any other prize (Ewen [1932] 1972, 144). This had the effect of removing the tontine element from the annuities. Jacob Cohen (1953) showed that the yield on lottery annuities varied between 1694 and 1784, when they were last issued, but for much of the eighteenth century yields were around 3 percent. Subscriptions were invited through lottery offices, the keepers of which were licensed after 1782. The prizes were of varying sizes and were in addition to the annuity payments. An adventurer was, therefore, gambling both on the size of the prize but also on his or her life expectancy (Cohen 1953). Sometime in the eighteenth century it became the custom to employ a roster of pupils from the Christ's Hospital Bluecoat School as the "disinterested and fit" persons who undertook the drawings. A number was drawn from one drum by one boy, and the associated prize or blank was drawn from the other drum by another boy. Results were announced and noted by clerks who sat between the drums and subsequently published.

In 1765 and 1769 and from 1785 lotteries reverted to the pure revenue-raising raffle type (Raven 1991). Tickets were no longer annuities, and prizes offered the only prospect of gain to adventurers. Moreover, after 1788 the agent system was abandoned and the Treasury sold all lottery tickets to contractors or stockbrokers (Woodhall 1964). The Bank of England acted as receiver of monies raised. Lotteries authorized by annual statute with preannounced quanta of tickets and prizes were put out to tender by the Commissioners of the Lottery. After 1804 even the amount to be raised was subject to the annual tender by contractors (Ewen [1932] 1972, 234). These contractors began the practice of selling smaller and smaller shares in the tickets for which they had tendered. This practice continued every year until 1826 with the exceptions of 1814 and 1819. It also seems that the contractors formed a cartel to bid for lottery business because the return to the government fell dramatically after the tender system was introduced in 1785 (Raven 1991, 372).

The cost of an individual subscription remained high throughout the eighteenth and early nineteenth centuries (Reith 1999). In 1710, for example, subscriptions to a loan of £1,500,000 were £10 each (the equivalent of £16,600 in 2009 terms) (Officer and Williamson n.d.). There were two consequences of such a high price. First, low-stake private lotteries (colloquially called Little Goes) proliferated despite their illegality and dubious honesty (Reith 1999). Second, subscriptions to the State Lottery were broken into many shares and those shares into further shares[5] by agents and subagents among the poor. Often these agents failed to distribute small prizes among small shareholders. In addition, a flourishing market in betting (or insuring) on the drawing of a particular lottery number either on a certain date or within a specified period became established.[6]

A bet is a contract between two parties wagering on the uncertain outcome of an event performance which can generally be monitored and enforced by backers and layers alike. Greater competition in betting markets can be expected to lead to higher standards of fairness. By contrast, a lottery is a pari-mutuel form of betting that cannot be monitored and enforced individually by bettors. Competition among a number of lottery operators is likely to lead to increasing market uncertainty and more widespread abuse of market power. The above arrangements for the State Lottery contain a number of perverse incentives which the Commissioners of the Lottery were apparently unable to eliminate. As a result, the State Lottery became increasingly notorious. Among these perversities were

- Lottery rules were complex and not understood.
- Prizes could be changed without notice.
- The State Lottery did not offer random chances, and people who bought more tickets had a greater chance of winning a prize.
- Subscription tickets could be stolen or borrowed, insured against, and subsequently selected.
- Agents who sold shares in successful subscriptions might abscond or not pay out minor prizes.
- Subscriptions might be received by agents but not paid over to the Commissioners.
- Subscription receipts might be forged so that prizes could be claimed.

The extent or frequency of these possible sources of unfairness or defalcation are not known with accuracy. Moreover, it is not clear whether they resulted from the nature of the State Lottery itself or the system of tendering introduced in 1785 by Pitt the Younger. Whatever the facts of the matter, by the early nineteenth century the State Lottery had become so notorious that it was finally stopped in 1826.[7]

The 1993 Act

By the time of the debate over the reintroduction of the State Lottery it was commonly held that the history of the State Lottery demonstrated the necessity of its strict regulation in order to ensure that it was fair to bettors. Furthermore, it was believed that abandonment of regulation in 1826 was a direct result of the neglect by successive administrations to provide that essential service. It is remarkable that when the National Lottery was re-commenced under the National Lottery etc. Act 1993 that many of the above perverse incentives were explicitly addressed.

The National Lottery etc. Act 1993 established three discrete and separable government interests covering the United Kingdom and the Isle of Man:

- *Regulation*: The National Lottery is to be run with all due propriety and players are to be protected.
- *Ownership*: Once the regulatory duties are satisfied, maximum net proceeds are to be obtained.
- *Compliance*: The disposal of monies raised and the transfer to distributing bodies is set out in the act.

The 1993 act, as amended, provides for an arm's-length body (ALB) called the National Lottery Commission (NLC) to be custodian of these interests on behalf of the government. The National Lottery is exempt from most provisions of the Gambling Act 2005, which applies only to Great Britain.

The ownership interest lies in the government's desire to finance good causes, in addition to grants from the Exchequer. This additionality requirement sets the National Lottery apart from many overseas lotteries, the proceeds of which either pass directly to government revenue or are used for the provision of state-provided education or health services. It is exercised by appointing an operator of the National Lottery. This appointment takes place after a competition for the right to hold the license to operate the National Lottery. As part of that competition bidders set out their plans for future sales and propose a payments structure by which returns going respectively to good causes and the operator are calculated. The NLC assesses the relative worth of bids, chooses the preferred bidder, and negotiates the detailed license to operate. By law, the NLC may issue a general license to operate to one operator or a specific license for a particular product or region to a number of operators. Only the first type has ever been issued.

Camelot Group is the only successful bidder for any license. The Third such license was commenced in February 2009 initially for 10 years since extended until January 2023.

Regulation of the National Lottery consists of ensuring that the various games are fair and safe. This is done by enforcing or manipulating the terms of the license to operate. Fairness involves ensuring that the outcomes of the games cannot be influenced by the operator or by forces other than chance. Checks ensure that they are random or pseudo-random and that prizes are appropriately distributed. Safety is ensured by making it difficult to play excessively (through low ticket prices, time and wallet load limits on interactive games and by not allowing risky games to be sold) or for children to play.[8] Much effort is expended in ensuring that it is impossible to acquire inside knowledge about the likely outcome of games. This might happen, for example, in the case of a shopkeeper who has activated a pack of scratchcards and, having sold most of the pack without a prize being claimed, reasons that a prize must lie among the remaining cards. As a consequence, where the National Lottery and betting companies compete for the supply of interactive games the gross gaming yield per National Lottery customer is approximately one-quarter of those visiting betting websites. Finally, it is illegal to offer bets on the outcome of the National Lottery in the United Kingdom.[9]

The compliance interest extends to ensuring that all operations are conducted in accordance with license conditions and that all monies are allocated appropriately between taxation, prizes, the operator, the National Lottery Distribution Fund (NLDF), and the Olympics Lottery Distribution Fund (OLDF). This involves undertaking checks on underage sales, randomness of games outcomes, and game security.

Nature of the National Lottery

Worldwide lotteries betray their past. Some, like the New York lottery, began as draw-based numbers games, such as those operated by organized crime. This same model continues to predominate, in the form of lotto. Other U.S. state lotteries have always been based around the sale of scratchcards. Still others, such as the New Zealand lottery, began selling scratchcards but are becoming more reliant on sales of draw-based games.

The U.K. National Lottery currently consists of draw-based (pari-mutuel) games, raffles, scratchcards, and interactive games played online. These products are described in table 31.1.

Only the draw-based games and raffles bear any relation to earlier state lotteries. However, even here there are important differences. First, there is no limitation on the number of tickets that can be sold. Second, the ticket price is in absolute (let alone real) terms much lower than state lotteries in the eighteenth and nineteenth centuries. Third, customers deal directly with the lottery operator via a national computer system and not with agents. Fourth, the government does not operate the National Lottery and cannot influence the outcomes. Fifth, side betting on the outcome of the National Lottery is illegal and of relatively minor proportions.

Table 31.1 National Lottery Products (Year ended March 2011)

	£ Million	Percentage of Total Sales[1]
Draw-Based Games		
Lotto	2,667	45.8
Lotto plus 5[2]	20	0.3
Thunderball	366	6.3
Hotpicks	206	3.5
Daily play[3]	46	0.7
Euromillions	1,056	18.1
Dream number[3]	39	0.7
Total draw games	4,400	75.5
Scratchcards[4]	**1,260**	**21.5**
IIWGs[4]	**165**	**2.8**

[1] Percentages may not sum to 100 due to rounding.
[2] Game introduced in year ended March 2011.
[3] Games no longer on offer at time of this writing.
[4] Data on sales of scratch cards are not published separately from IIWG sales. Estimated.

Source: National Lottery Commission (2011), appendix A, p 17.

All National Lottery products compete directly with products supplied by organizations regulated under the Gambling Act 2005 and, in the case of lotteries, are subject to regulations on maximum prizes. Draw-based games compete with society and local authority lotteries. National Lottery scratchcards compete with scratchcards sold by private companies. Interactive games compete with identical products available on betting websites. However, only National Lottery games attract lottery duty at 12 percent of the ticket price. Competing products are either not taxed or are charged general betting duty of 15 percent of gross gaming yield.

After the first draw on November 19, 1994, total sales rose to a peak in 1997–1998. This was predominantly due to a significant rise in lotto sales. After that, total sales fell steadily for the next five years. However, since 2002–2003 the total sales have risen despite relatively stagnant sales of draw games largely as a result of a resurgence of scratchcard sales, the success of EuroMillions, and the emergence of the interactive channel. The highest total sales in nominal terms in the history of the National Lottery were reached in 2010–2011 (see table 31.2 and figure 31.1).

Unlike many overseas lotteries, scratchcards have never represented more than 30 percent of total National Lottery sales. Hence the economics of the National Lottery is dominated by the economics of draw-based games in which the main prizes are determined in a pari-mutuel fashion and the main prize jackpot rolls over if not won.[10]

Table 31.2 National Lottery Sales, 1995–2011 (Years ended March)

Years	Draw Games	Scratchcards (1)	IIWGs (1)
1995	1,191		
1996	3,695	1,522	
1997	3,846	877	
1998	4,713	801	
1999	4,559	669	
2000	4,533	561	
2001	4,437	546	
2002	4,256	578	
2003	3,997	577	1
2004	3,974	635	6
2005	4,028	717	21
2006	4,148	802	63
2007	3,983	849	79
2008	3,857	1,001	108
2009	3,928	1,082	139
2010	4,138	1,184	155
2011	4,400	1,260	165

Source: National Lottery Commission.
Note: 1. From 2003 onwards scratchcards and IIWGs are not differentiated in the NLC's Annual Report. The scratchcard and IIWG data for 2003–2011 are estimated.

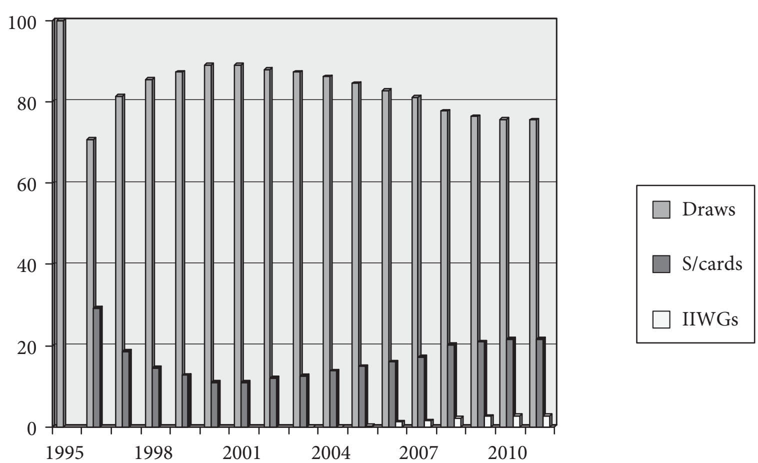


FIGURE 31.1 Pattern of National Lottery sales, 1995–2011 (years ended March; as percentage of total sales)

Source: National Lottery Commission.

Different games sold in different ways yield different levels of net proceeds for each unit of expenditure. There are two main reasons for this. First, typically scratchcards and IIWGs (interactive instant win games) offer higher (sometimes 25% higher) prize percentage payouts (PPP) than draw-based games. Second, retailers are paid a commission of 5 percent of face value for draw tickets and 6 percent for scratchcards, but no commission is payable for draw games sold online or for IIWGs. Hence as the National Lottery product structure and marketing strategies change the average PPP of the National Lottery also may change.

The general pattern of the flows of funds associated with the National Lottery is illustrated in figure 31.2. This figure does not aim to be complete. It excludes many of the transactions between a variety of trust and government accounts by which this general pattern is achieved. It also ignores the disposal of funds by lottery distributors on the one hand and the operator and its suppliers on the other. There are two lottery distribution funds. The National Lottery Distribution Fund (NLDF) provides a source of income for lottery distributors that operate independently of the government to distribute monies to arts, sports, heritage, and charitable causes. The Olympic Lottery Distribution Fund (OLDF) was founded in 2005 to help fund the London 2012 Olympic and Paralympic Games. In outline, revenue from ticket sales net of lottery duty is split according to the terms of the license to operate between retailers, prizes, the OLDF, the NLDF, and overheads disbursed by the operator. Overheads cover the costs of sales, including capital investment, operating costs, payments to suppliers, advertising and marketing, license fees, company tax, interest, amortization, and operator profit. The proceeds accruing to the NLDF and OLDF are supplemented by unclaimed prizes and investment income earned on accumulated balances. The gross income is available for paying government operating costs, for distribution, and for investment by the Commissioners for the Reduction of the National Debt (CRND) acting under instructions from the Department for Culture, Media and Sport (DCMS). In addition, since 2007 the NLDF has made a series of transfers to the OLDF.[11] For much of the past decade drawdown by distributors has been larger than gross income so the NLDF balance has generally fallen from year to year. In the year ended March 2011, the cost of raising funds amounted to 24 percent of NLDF and OLDF gross income and 6.9 percent of total ticket sales. Prizes paid out and NLDF and OLDF gross income, respectively, represented 49.6 percent and 28.8 percent of ticket sales.

The economics of the National Lottery is best considered under three headings,

- the economics of all National Lottery games,
- the economics of pari-mutuel games, and
- the economics of scratchcards and interactive games.

The economics of the National Lottery has largely been confined to the analysis of demand factors that explain the sales of draw games. So far as is known there has been no attempt to analyze the demand for scratchcards or the influence of supply-side factors, such as the behavior of the operator or the government.

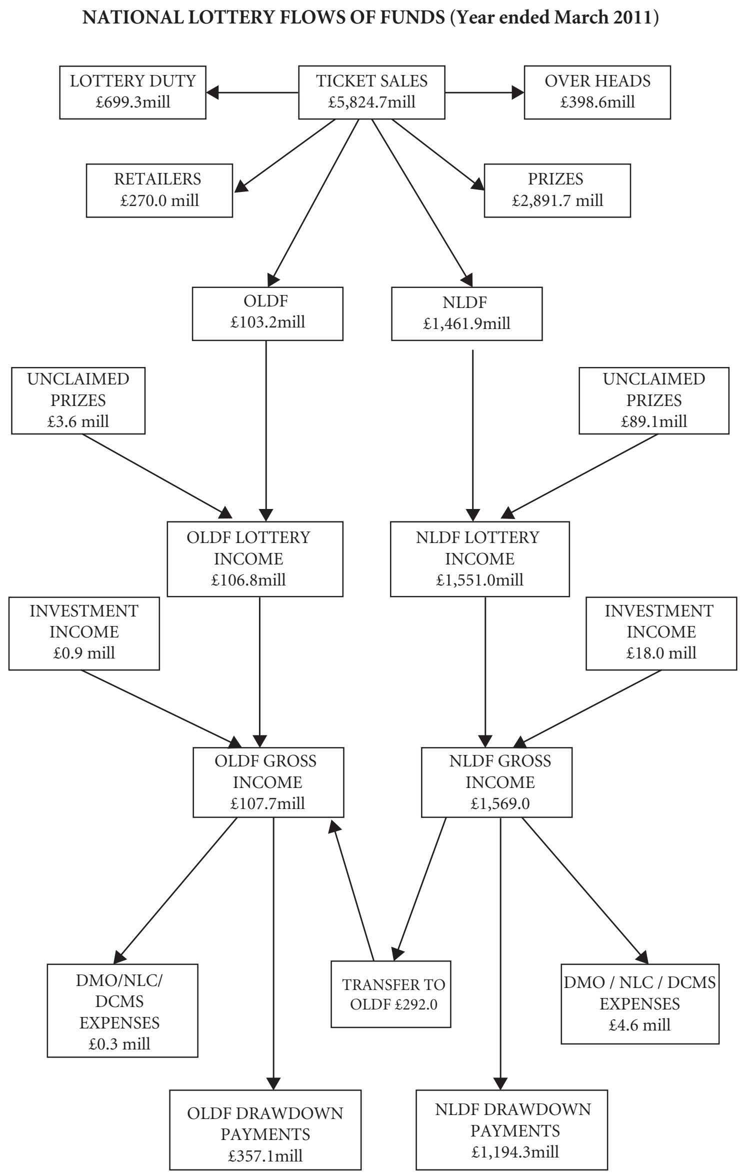


FIGURE 31.2 National Lottery flows of funds (year ended March 2011)

Sources: National Lottery Commission (2011); DCMS 2011.

U.S. Experience

There have been several attempts to explain variations in total lottery sales as a result of fluctuations in economic activity. For example, John Mikesell (1994) argued that total sales of U.S. state lotteries were part of household consumption and, hence, related to state personal income. He argued that changes in state personal income may be masked during recessions by transfer payments. Consequently, an indicator of short-term cyclical downturns like the unemployment rate was also part of the explanation. This was reinforced by the likelihood that the unemployment rate reflects the changed perception of the attractiveness of the lottery gamble which increases at times of high unemployment. Mikesell also conjectured that lottery sales had a natural tendency to reach a maximum and to decline thereafter and to be higher if the pari-mutuel game lotto was part of the game portfolio.

Mikesell estimated the following model:

$$L = f(Y, A, U, S),$$

where L is state lottery sales, Y is state lottery personal income, A is the age of the state lottery in quarters, U is the state unemployment rate, and S is the non-Lotto share of state lottery sales. The model was estimated using quarterly data for all U.S. states offering lotteries in 1991 with the exceptions of Rhode Island, for which quarterly data were not available, and Texas because of its newness. Lottery sales and personal income were expressed in per capita terms, all data were in logarithmic transformation, and income and sales data were adjusted to 1987 price levels. Dummies were included for each state except West Virginia to take account of state-by-state social, cultural, demographic, and economic differences beyond those specified in the model.

S is the outcome of purchase decisions by lottery customers, which is correlated with the error term of an OLS regression model. As a consequence the estimates of the influence of state economic activity on lottery sales were found to be unreliable. This effect was removed by regressing S on other independent variables together with state population and whether the state in question has a neighbor with no lottery. The new predicted value of S was then used to reestimate the lottery sales equation. The result of this TSLS estimation was as follows (see table 31.3):

Table 31.3 Log Per Capita Lottery Sales, Q4 1983 to Q4 1991

$N = 820$	Coefficient	Standard Error
Constant	−33.846	2.104
Logy	3.899	0.224
LogU	0.171	0.057
LogA	−0.561	0.078
Logs	−1.573	0.231
R^2	0.840	
AdjR^2	0.833	
F-statistic	114.240	

Source: Mikesell (1994), table 1, p. 168.

This shows that the income elasticity of lottery sales is 3.9. It also shows that for each 1 percent rise in unemployment there is a 0.171 percent rise in lottery sales. Thus as a recession begins lottery sales rise ceteris paribus, which partly offsets the fall in per capita income. Lottery sales decline in real terms with the age of the lottery and with a rise in the proportion of non-lotto sales.

U.K. Explorations

So far as is known, unlike in the United States, no study has been published on the social and economic factors underlying overall sales of the National Lottery (see figure 31.3). It is not, therefore, possible to determine whether or not the Mikesell approach could be employed in the United Kingdom. However, it is possible to remark that sales generally declined during much of the unprecedented economic boom of the 1990s and early 2000s (Creigh-Tyte and Farrell 2003). Moreover, the growth of National Lottery sales has continued unabated during the economic instability of the past three years. On the face of it, therefore, it appears that a search for robust relationships between sales and macroeconomic variables will prove fruitless.

The experience of the National Lottery appears to contradict the earlier findings from the United States. In part, the absence of a readily apparent relationship between National Lottery sales and socioeconomic variables may be the consequence of the peculiar features of the development since 1994 of lottery sales on the one hand and the economy on the other. It also may be because more subtle relationships lie hidden from immediate view yet to be uncovered by sophisticated analysis. It is possible,

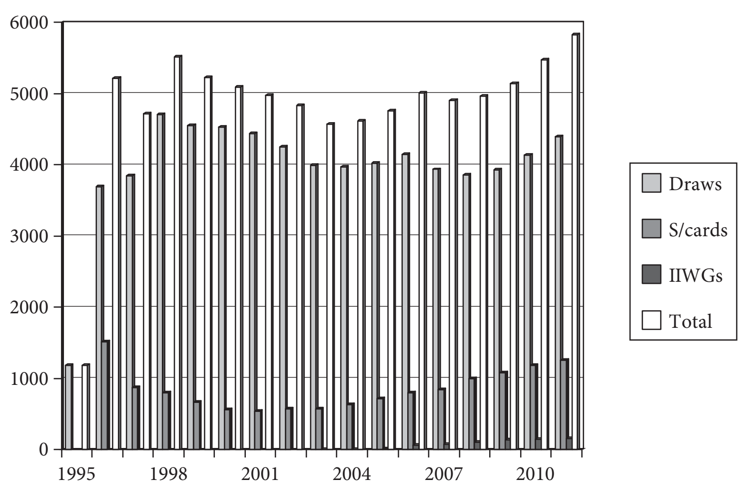


FIGURE 31.3 National Lottery sales, 1995–2011 (years ended March; £millions)

Source: Table 31.2.

therefore, that the Mikesellian relationship will emerge as the U.K. economic performance becomes less stable. Alternatively it may mean that robust causal explanations of National Lottery sales are more likely to be found among models of retail behavior than among those based on macroeconomic relationships.

Draw-Based Games

Theory of Draw Games

Philip Cook and Charles Clotfelter (1993) described the main draw game, lotto, as pari-mutuel gambling with long odds and large jackpots. They proposed that such games display a scale effect because bettors in lotteries serving large populations only take account of the size of the jackpot and do not consider that a larger population means a smaller chance of winning. Lotto has the feature that each bet increases the size of jackpot available to all players while simultaneously increasing the chance that the jackpot will have to be split with someone else. Normally the impact on the expected value of the jackpot is positive because the positive impact on jackpot size outweighs the negative impact of the increased chance of splitting.

In general, then, the expected value of a lottery bet is the probability of a chosen combination of winning the jackpot multiplied by the chance of the jackpot being won multiplied by the risk of having to split the jackpot. So the expression of expected value (EV) is

$$EV = PROB(WIN) \times PROB(JACKPOT) \times EXPECTEDSHARE.$$

The probability of winning the jackpot with a particular bet is the probability of any bet winning the jackpot (p) multiplied by the number of combinations (*boards*) purchased (W). Let the total number of combinations purchased by others be denoted as N. The size of the jackpot is the price of one combination (c) multiplied by the fraction of the board price devoted to the jackpot fund (k) multiplied by the total number of combinations purchased ($Q = W + N$) all added to the size of any rollover (R). The expected share of any jackpot won is more complex. If all combinations are chosen randomly the probability of a jackpot being won is very small, and a large number of independent trials indicate that the probability of a shared jackpot can be represented by a Poisson series. Hence there are Q combinations among which sharing might occur with probability p. The proportion of those not involving sharing is $\mathrm{e}^{-\mathrm{pQ}}$, so the expected proportion involving sharing is $(1 - e^{-pQ})$. The resulting expression for EV is

$$EV = pW \times (R + kc(Q)) \times (1 - e^{-pQ}).$$

If $W = 1$ and $R = 0$ EV is a monotonically increasing function of Q.

Cook and Clotfelter argued that this expression displays two features. First, if qQ is approximately 1 when there is no rollover, the EV of a single play ($W = 1$) is less than the jackpot payout rate (k). Why, then, should any rational individual buy a lottery ticket? Second, the probability distribution of the number of winners depends only on pQ, so larger games have the same distribution of winners as small ones but offer larger jackpots. To the bettor, larger games offer larger jackpots at reduced probability of winning with all other features unchanged. Because a bettor's evaluation of lotto is more sensitive to the size of the jackpot than the chance of winning, larger jackpots generate larger sales than smaller ones.

Lisa Farrell and Ian Walker (1999) and Farrell et al. (2000) questioned this analysis. They argued that the Cook and Clotfelter calculations did not apply to the situation when there was a rollover ($R > 0$). In the presence of rollovers, increased sales dilute the value of the jackpot rolled over from previous draws. Farrell and her colleagues postulated that the relationship between EV and sales for some finite Q and for sufficiently large R became monotonically decreasing as the dilution effect dominated the scale effect. For given sales the EV of rollover draws is always higher than for non-rollover draws, implying a random variation in EV due to rollovers. This provided an explanation for why a rational individual would participate in a draw when there are rollovers (see figure 31.4).

Nevertheless, for many draws EV is less than the price of the ticket or board, which is contrary to the tenets of demand theory. Farrell and Walker (1999) followed John Conlisk (1993), who suggested that that participation in the lottery yielded positive but unspecified non-pecuniary benefits in addition to the EV of the ticket. Farrell and Walker envisaged play as being determined by a reservation expected value (REV).

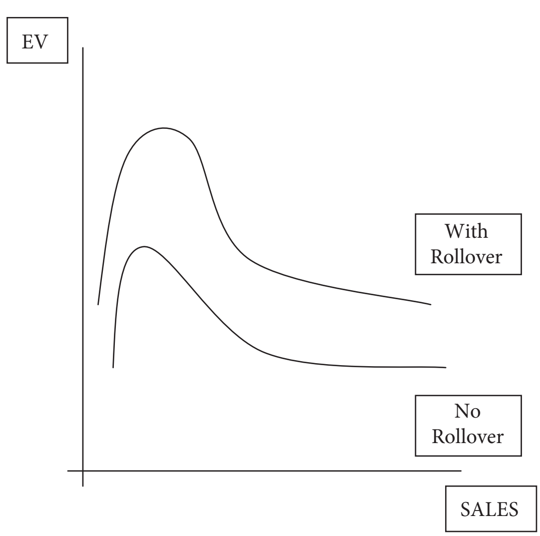


FIGURE 31.4 Expected value and sales

The *REV* of the ith individual depends on the number of combinations purchased (W) and a vector of characteristics of the ith individual and the W combinations. Non-pecuniary return is assumed to be a normal good so that the marginal benefit it yields falls as W rises. Hence the REV schedule shifts upward as income rises and if risk aversion declines. Farrell and Walker hypothesized that an individual will purchase lottery tickets to the point at which *EV* falls to the level of *REV*. However, this equality is randomly disturbed by increases to *EV* resulting from the occurrence of rollovers and by the increases in sales that result.

The variation in *EV* resulting from rollovers also allowed the elasticity of demand to be inferred. The variation in the difference between *EV* and the board price can be interpreted as the cost to the bettor of the lottery bet (P). When the price of a board is unity,

$$P = 1 - EV.$$

This relationship can be used to estimate price elasticity of demand when the board price does not change. Over the period November 1994 to February 1997 purchases when $EV = 0.45$ in non-rollover draws were compared with those when $EV = 0.63$ in rollover draws. Farrell and Walker calculated that the price elasticity varied between −1.785 and −1.456 and income elasticity between 0.449 and 0.132 depending on the estimation method. These estimates were then used to compute the consumer surplus of lotto (roughly £1 billion p.a.) and the annual deadweight loss (between £0.48 and £0.51 billion). By contrast, Farrell et al. (2000), using data from the period November 1994 to February 1996, found that the implicit price elasticity of lotto ranged between −0.80 and −1.06.

Farrell, Edgar Morgonroth, and Walker (1999) hypothesized that playing the lottery is addictive and adapted the Becker and Murphy (1988) model of rational addiction to the explanation of lotto sales. They reasoned that current consumption was determined, in part, by an individual's level of addiction generated by past consumption. Hence the regression models of the type reported above were augmented by a measure of the influence of past consumption on present purchases. This influence was assumed to vary according to size and presence of rollovers. Farrell, Morgonroth, and Walker (1999) compared the traditional specification of lottery demand with that including addiction using sales data for the period November 1994 to February 1997. They found that addiction was a significant explanatory variable in the lottery sales equation and that long-run price elasticity of lottery demand was −1.55.

This result was questioned by David Forrest, Robert Simmons, and Neil Chesters (2002), who argued that it should be treated with caution for three reasons. First, it was based on the first 116 weekly observations and so was heavily influenced by announcement effects. Second, the National Lottery was becoming embedded into the behavior of consumers with all the volatility and uncertainty that transition might imply. Third, the introduction was associated with large prizes associated with rollovers, which might shifted demand outward.

Forrest, O. David Gulley, and Simmons (2000) estimated the elasticity of demand for lottery tickets by computing the effective ticket price $P = 1 - EV$, where

$$EV = 1 - ((I/Q)(R + kcQ)(1 - e^{-pQ}) + EVs)$$

such that *EVs* is the expected value of the non-jackpot prizes. There is a general similarity with the formulation of Cook and Clotfelter. Forrest and his colleagues argued that only if the price elasticity of lotto demand was close to minus unity would the aim of sales maximization be achieved.

However, the effective price P is only calculable ex post and so cannot be used in the estimation of the lotto demand curve. Instead, a two-stage approach was adopted. First, the expected effective price was estimated from variables derived from information widely available to the public before the draws took place. This included the size of any rollover or superdraw and the average sizes of Wednesday and Saturday jackpots in non-rollover weeks. Forrest, Gulley, and Simmons found that estimated bettors' expectations were not significantly different from actual effective price calculated ex post. Second, actual sales were regressed against the effective price estimated in stage 1, together with other variables likely to affect lottery demand. It was found that the long-run price elasticity of demand was −1.03. This result was confirmed by Forrest, Gulley, and Simmons (2004) when demand functions for Wednesday lotto and Saturday lotto were separately estimated. The respective price elasticities were −0.709 and −1.074.

Nevertheless, this analysis does not consider the problem of why a rational bettor would willingly purchase a lottery ticket, the expected value of which is usually less than its money price. We have already reviewed Farrell and colleagues' proposed solution to this problem. Forrest, Simmons, and Chesters (2002) suggested, with Conlisk (1993), that the activity of engaging in the National Lottery itself yields positive consumption benefits (see also Hartley and Farrell 2002 and Hartley and Lanot 2003). According to Forrest and colleagues, these consumption benefits are derived from the prospect of buying the dream of winning the jackpot. This includes imagining how a jackpot might be spent or enjoying the prospect of quitting one's job. This view is supported by Don Slater and Eva Neitzert (2007), who reported that purchase of lottery tickets is associated with the formation of dreams by bettors. Moreover, they argued that the greater the jackpot the better the dream. (In fact, Emma Casey 2007 found that the relationship is probably not linear because most people appear to be afraid of winning too large a jackpot lest it lead to too great a disturbance in their relationships and their lives.) As a result, lottery sales are more closely related to jackpot size than effective price.

The hypothesis was tested by estimating two regression models. In the first, effective price appeared as an independent variable but not the size of jackpots and vice versa in the second. The results of Cox tests showed that each model was rejected against the other. Forrest, Simmons, and Chesters (2002) interpreted this result as suggesting that both effective price and jackpot size are important in explaining lottery sales. It is unlikely that any regulator or operator of lotteries would disagree with this conclusion.

Nevertheless, it has yet to be explored whether jackpot size is an accurate representation of the hypothesized consumer benefit of lottery participation or is a catchall for other factors yet to be individually identified. It is even perhaps conceivable that the demand for lottery tickets is consistently at least myopic, infested with money illusion and risk, and even possibly irrational.

Patrick Roger (2009) investigated the demand for EuroMillions lottery tickets across all European markets in which it was sold. He estimated a demand equation of the type proposed by Forrest, Simmons, and Chesters (2002). First, Roger estimated the anticipated value of a draw or effective ex ante price and then used the resulting price series as an independent variable in the estimated demand equation. He found considerable variation in the long-run price elasticity of demand between different jurisdictions. Most price elasticities ranged between −0.82 and −0.91. However, in Spain the long-run elasticity was −0.49, in Ireland it was −1.44, and in the United Kingdom it was −1.76. Roger suggested that the Ireland and U.K. elasticity is the result of competition with other forms of betting.

Lottery Design

The National Lottery has also been explored from the point of view of a rational designer seeking to maximize revenue. In 2001 Walker and Juliet Young argued that sales depend on the size of the set of number that bettors choose (n) and the total numbers available (NUM). For U.K. lotto $n = 6$ and $NUM = 49$, and it is termed 6/49 game. On the one hand, if the game is easy to win, rollovers are infrequent and there is a danger that the game will lose its attraction and sales will decline. On the other hand, if the game is so hard to win that long rollovers are very frequent, ultimately people will cease to buy tickets because they will believe they have little chance of winning. Game designers adjust the n/NUM ratio to match the conditions of the lottery market.

In addition, they must decide the proportion of the ticket price devoted to all prize pools, how skewed toward the jackpot prizes are, and how much weight should be given to middle-rank prizes compared to those on the extreme. Walker and Young called these, respectively, the mean, skewedness, and variance properties of the prize distribution. As rollovers occur all three properties increase in size for given sales. They found, using OLS regression, that sales are an increasing function of the mean and of skewedness of prize distribution. However, sales decrease with an increase in its variance.

Walker and Young attempted to model the effects of a change in the lotto format from 6/49 to 6/53. Unfortunately, the computational model used was very complex and did not always converge to a unique solution. Nevertheless, they felt able to tentatively conclude that a change to 6/53 would mean lower mean return and raised variance, both of which are likely to lower sales, and increased skewedness, which is expected to raise them. As a result, ceteris paribus sales would be lower with 6/53 than with 6/49.

An alternative approach to the problem was taken by Roger Hartley and Gauthier Lanot (2003), who, unlike Walker and Young, did not attempt to estimate bettor behavior. Instead, they built a theoretical dynamic model of consumption under uncertainty, which they then parametized using the facts of the National Lottery. Each individual is assumed to decide the optimal number of lottery tickets he or she will buy so as to maximize the expected value of lifetime utility within a binding lifetime budget constraint. Lifetime utility depends on the rate of discount, the time between draws, purchases of lottery tickets, and purchases of other goods. Following Forrest et al. (2002) and Hartley and Farrell (2003), the utility function has a "fun" term dependent on ticket purchases, external factors, and history. Rollovers are regarded as exogenous by consumers but are, in truth, the result of aggregate participation in the lottery. Hence the rollover process affects participation which, in turn, affects the rollover process. Expenditure decisions are assumed to be contingent on the number of weeks the lottery has rolled over, which is determined by a random number (RAND). Fun is assumed to be derived from winning a large jackpot and so is determined by RAND. If an individual wins a jackpot he or she is assumed to quit participation because all possible fun in contemplating the future has been extracted. This last aspect of lottery demand is contradicted by the existence of multiple jackpot winners, which suggests that winning a jackpot does not remove the fun of participating in the National Lottery.

Hartley and Lanot (2003) presented a series of simulations which show a variety of trade-offs between the probability of jackpot wins and tax receipts resulting from different model parameter settings. They found that increasing the probability of a jackpot win and decreasing the tax rate led to a rise in tax revenue. This occurs because, although the additional revenue associated with more rollovers falls, this effect is more than compensated by a rise in participation due to the greater probability of jackpot win. The greater the number of participants means more frequent sharing of jackpots. Reducing the tax rate also increased participation sufficient to offset the reduced tax per ticket sold. As a result, Hartley and Lanot suggested that whether the tax rate on the National Lottery is too high and the probability of winning the jackpot is too low are questions to be investigated further. So far as is known, neither the government nor the operator has seriously considered this suggestion.

David Paton, Donald Siegel, and Leighton Vaughan Williams (2003) found that price changes in the National Lottery had significant effects on the demand for other gambling, most notably betting. When the effective price of the National Lottery dropped the demand for gambling in general also dropped.

Lottery Addiction

We have already noted that Farrell, Morgonroth, and Walker (1999) found significant evidence of lottery "addiction." By contrast, Paton, Siegel, and Vaughan Williams (2004: 857) found no evidence that the price elasticity of lottery demand was higher in the long run than in the short run and so could be seen as addictive.[12] In reality, addiction, defined as the dependency of future sales on past consumption, is likely to reflect a range of explanatory factors and therefore be difficult to detect using econometric methods.

Supply of Lottery Products

The economics of the supply of National Lottery products has barely been subjected to economic analysis. Most economists take the nominal price (c) of a lottery board as given and usually assume that $c = 1$. Presumably, this assumption would be justified by assuming that the National Lottery acts as a pure monopolist that chooses to set the nominal board price at unity so that sales are determined solely by demand factors. While this may be valid for lotto in the United Kingdom since 1994, it is not true of all other draw games and has never been true of all types of scratchcards or IIWGs. The history of the National Lottery is replete with minor product changes involving different levels of c,[13] different k, and different EVs. On closer examination, therefore, a series of nuanced processes lead to the changing nature of the National Lottery product offer.

In this respect the National Lottery is no different from any other supplier of low value, mass-consumption products. However, the National Lottery is different from the likes of Coca-Cola in that these dynamic supply-side processes intimately involve the U.K. government in the form of the National Lottery Commission (NLC). The NLC is custodian of a set of policy outcomes which are couched in terms of its three statutory duties. Its mission is to appoint and supervise the lottery operator so that the outputs that are supplied lead to these outcomes. Thus specifying the National Lottery in terms of a sales-maximizing monopoly may not be an accurate model of reality. A simplified idea of the dynamic relationships that exist between the NLC and the lottery operator is given in figure 31.5.

Several features of this dynamic system should be remarked on. First, the powers of both parties are highly constrained. On the one hand, the NLC must ensure that the operator remains commercially viable in competition with other gambling companies. On the other hand, the operator cannot expect to pursue sales at the expense of its social responsibilities to prevent excessive gambling or to prevent sales to children. Second, the government has sought to partition the gambling market so that there is "clear, blue water" between the National Lottery and other gambling products.[14] Nevertheless, the reality of these dynamic relationships reflects, albeit imperfectly, the state of the U.K. gambling market. Third, the instruments each can call upon to achieve policy outcomes or business strategies are respectively limited by the damage that the unconstrained pursuit of one might do to the other. For example, the NLC does not instruct the operator to market certain products, though it does exercise its right to refuse to allow some products to be sold.[15] At the same time, the operator has not sought to actively undermine its special semimonopoly status in the gambling market. Fourth, the relative strengths of these various dynamic relationships vary over time. When sales are growing and the marketing strategies of the operator appear to be working well the government is likely to express less concern about net proceeds and the operator is likely to seek and obtain more commercial autonomy than if the opposite were the

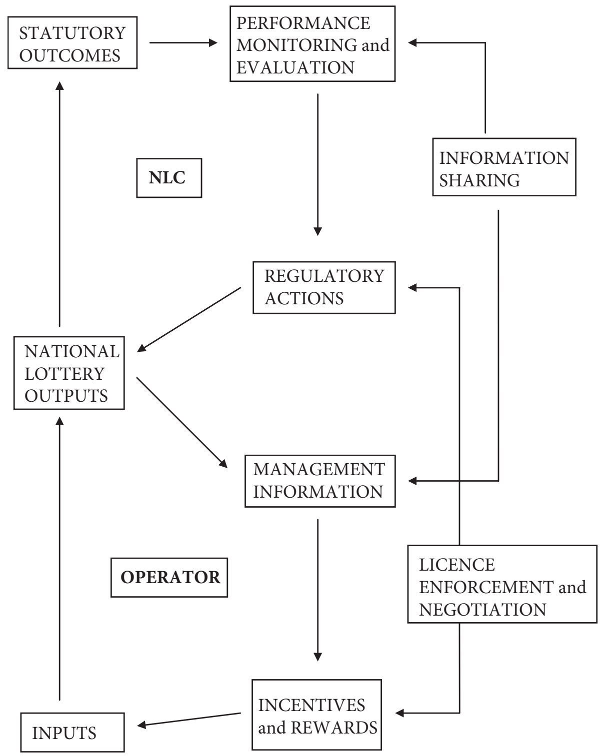


FIGURE 31.5 National Lottery supply

Adapted from Hancock (2011), figure 3.1, p. 30.

case. Nevertheless, if it were found that rapidly increasing sales were occurring at the expense of its primary policy outcomes it is likely that the NLC would seek to instill greater social responsibility in the operator.

The supply side of the National Lottery, therefore, takes on many of the features of a dynamic duopolistic system formed between the operator and the owner/regulator of the National Lottery. The regulator seeks to ensure that the outputs of the National Lottery are consistent with policy outcomes. However, in practice it cannot directly command the nature of those outputs, though it has refused to allow some to be supplied and has placed conditions on how others are to be sold. It neither has the ability to directly affect the mobilization and disposition of inputs commanded by the operator nor does it choose to dictate marketing strategies. Hence any consistency that exists between output and outcomes results largely from the way the incentives incorporated in the license to operate influence the day-to-day operations of the operator. These incentives are designed to ensure that, within the limits placed on its operations by the NLC, the operator seeks to sell as many National Lottery products as it can. In pursuing this aim, the operator is assured of, under the terms of the license, a guaranteed share of the after-tax, after-prize total revenue.

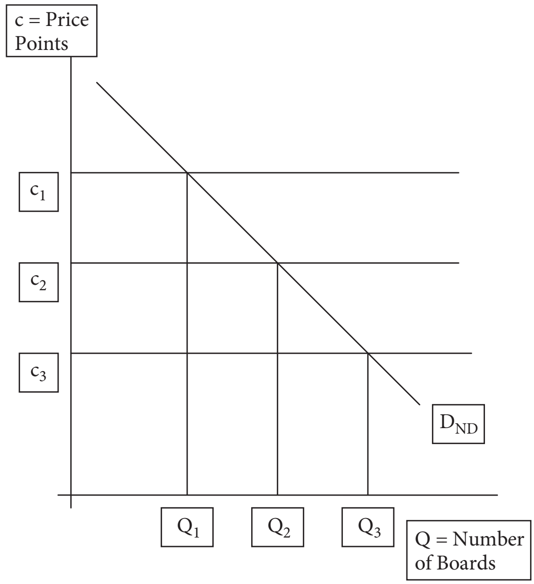


FIGURE 31.6 Demand and supply of scratchcards and instant win games

These revenue shares vary according to the type of product sold and the way in which they are distributed. This suggests that there are a number of supply curves for the various products sold by the National Lottery, each of which has different implications for the total revenue of the operator. In this respect the economics of scratchcards and IIWGs is no different from that of any other small outlay, frequently purchased object, such as a serving of Coca-Cola or a bar of chocolate. This is illustrated in figure 31.6.

If we consider a world in which scratchcards and IIWGs[16] are supplied at three separate price points, c_1, c_2, and c_3, then total sales of non-draw games, S_{ND}, are the multiple of c and Q. So

$$S_{ND} = c_1.Q_1 + c_2.Q_2 + c_3.Q_3.$$

However, with draw-based games there is a different relationship. If we limit consideration to two price points and one rollover at each we obtain the situation illustrated in figure 31.7.

There are four possible outcomes, A, B, C, and D. However, whether or not there is a rollover at a particular price point is literally a matter of chance. Let us assume that it has a probability P over a large number of trials. In that case, S bears a probabilistic relationship with c and Q.

$$S_D = c_4(P(Q_6) + (1 - P)(Q_4)) + c_5(P(Q_7) + (1 - P)(Q_5)).$$

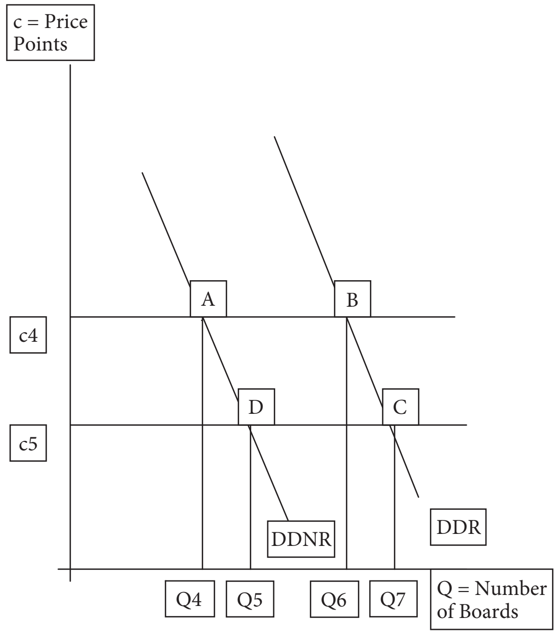


FIGURE 31.7 Demand and supply of lottery draws

In general terms, since

$$S = S_{ND} + S_D,$$

total sales at time t are calculated as

$$S_t = f(D_{ND};\ D_D;\ P)_t.$$

It follows that attempts to model National Lottery sales must be based on the normal demand function for draw games augmented by the following:

- the effect of the probability of the occurrence of a jackpot on demand for non-draw games,
- the effects of the demand for non-draw games on the demand for draw games and vice versa,
- the effects of any variation in the costs of supply, including changes to regulatory regime, and
- the effects of any variation in the price points at which supply takes place.

Summary and Conclusions

Since its modern inception in 1994, the National Lottery has become deeply embedded in the culture of the United Kingdom. It enjoys the largest turnover of any single brand of consumer goods and is, by far, the most available form of gambling. It is played at some time or another by nearly 70 percent of the adult population (Wardle et al. 2011) and, illegally, by perhaps 10 percent of British children (Ipsos MORI 2011).

Unlike most forms of gambling, women and men participate in the National Lottery with more or less equal enthusiasm. Like the BBC, It has earned high levels of public trust (Creigh-Tyte and Lepper 2004a, 2004b). Many who buy National Lottery tickets do not regard their actions as gambling, and some even believe it to be a means of investing for retirement. Most, whether they play or not, regard National Lottery proceeds with a proprietary interest and strongly disapprove of the government's use of them as an alternative source of revenue (Lepper and Hawkes 2007).

Nevertheless, despite their ubiquity, only two National Lottery products (Lotto and EuroMillions) have been subjected to rigorous economic analysis. In large part this is due to the lack of publicly available comprehensive time-series data which would make such an analysis possible (Creigh-Tyte and Farrell 2003; Forrest 2003). Despite the fact that the National Lottery yielded in excess of £2.2 billion to the Exchequer or to communal causes in 2010–2011, most of its activities are closed to public scrutiny. This arises because the National Lottery is run on behalf of the government by a privately owned company with its own commercial sensitivities unrelated to the government business it undertakes. No data are now published on the social or geographical location of sales, so it is not possible to determine the socioeconomic circumstances from whence they arose. No reliable data on participation, frequency of purchase, or purchase size are published, which means it is not possible to seriously analyze the various markets served by the National Lottery. No data on coverage are made available, thereby making it difficult to build models on the occurrence of jackpots. No data on boards sold in different games are published, which means that the quantity of National Lottery products is not precisely known and there is an imprecise reconciliation between sales data and the variables employed in economic theory.

Two general approaches could conceivably be adapted to the study of the economics of the National Lottery. Each complements the other, but neither can claim to present a full picture of all the economic influences that bear on it.

First, the National Lottery could be explored from the point of view of sources and uses of resources. We have attempted to outline a small part of the National Lottery flows of funds which could be expected to form a small element of such an analysis. However, much remains to be done. Currently there has been no systematic analysis of sources of National Lottery revenues or how they are distributed. There are few publicly available data on who buys lottery tickets and none on where those purchases take place. Hence the age-old question of whether or not the National Lottery is a tax on the poor cannot yet be systematically approached, let alone answered, with any accuracy. Moreover, economists have not attempted to assess whether or not the National Lottery is a relatively effective, efficacious, and efficient method of providing for society's merit goods. Finally, there have been few attempts to analyze the National Lottery in the context of the U.K. gambling industry (cf. Forrest et al. 2004, 2010). No doubt this is due to the general paucity of accurate, complete, consistent and timely publicly available time-series data on gambling in the United Kingdom.

Second, it is also possible to explore the demand for, and supply of, National Lottery products. We have seen that this is the approach taken by most economists when analyzing the National Lottery. It has led to deep and subtle analysis of the expected value, and robust estimates of short-term and long-term price elasticities, of draw tickets. This work has been employed to inform the design of draw-based games. However, lacunae remain. On the one hand, the demand for scratchcards and IIWGs has not been subjected to the same rigorous examination as Lotto and EuroMillions. As a consequence, the economics of the demand for National Lottery products remains substantially incomplete. Access to comprehensive information about the sales and PPPs of different lottery games and who buys lottery tickets and where would greatly aid investigation of the micro-economics of the National Lottery. On the other hand, there has been no attempt to explore the economics of supply of the National Lottery. No doubt this failure is the result of poor or nonexistent publicly available databases on cost conditions and the fact that the nominal ticket price of Lotto has remained at £1 since 1994, thus making simple the translation from sales to boards. Nevertheless, as new products are introduced with different operating costs, and at different price points, from Lotto, this lack of analysis of supply conditions represents a potential source of systematic error. This is particularly so as the importance of Lotto in total sales continues to weaken.

National Lottery proceeds are unusual among revenues of the general government sector because there is no reliable method of forecasting them. This is a direct result of the weaknesses we have noted in the economic analysis of the National Lottery. Robust dynamic models are not available, and it is not possible to rely on such shortcuts as correlations between National Lottery sales and on such macroeconomic variables as household consumption. No knowable degree of accuracy can be ascribed to the future course of National Lottery proceeds. As a consequence, the management by the general government sector of the uncertainty which variations in those proceeds engender is likely to be characterized by excessive risk adversity, imprecision, and high cost in terms of resources employed and revenues forgone.

Notes

The authors are grateful to Ben Haden for helpful and perceptive comments which have eliminated many errors from an earlier draft. The views expressed herein are those of the authors and are neither representative of the views of nor endorsed by the Department for Culture, Media and Sport or the U.K. Government and cannot be construed as if they are. Any errors of fact, logic, or judgment that remain are the sole responsibility of the authors.

1. All such conversions are based on average earnings.
2. By reducing the number and the value of the prizes the authorities imposed a double reduction in the expected value of the tickets in the Lotterie Generall.
3. This did not deter many private lotteries being launched over the subsequent two and half centuries (for a list see Ewen [1932] 1972). Notable examples included lotteries to pay for the settlement of Virginia in 1612, 1614, 1615, and 1618–1621, the London Aqueduct lotteries in 1635 and 1639, and the Royal Jewel Lottery 1685.
4. These boxes were six feet in diameter and wheel-shaped. They became know as wheels.
5. The Joint Stock Act permitting the issue of limited liability equity by joint stock companies was not passed until 1823. Before that companies were forced to operate under royal charter.
6. Roger Munting (1998, 629) reported that there were 200 insurance houses in London in the late eighteenth century which operated like betting shops. They were made illegal in 1802 but apparently continued until the State Lottery ended in 1826.
7. Echoes of the notoriety of the state lotteries of this period can be heard to this day. For example, the 1993 act setting up the modern National Lottery required that it be run and promoted with all due propriety.
8. Children are defined in the Gambling Act 2005 as those under 16 years of age.
9. Note that this does not prevent bookmakers from offering to lay bets on the outcomes of lotteries in other jurisdictions.
10. There are absolute limits to the number of successive rollovers that are allowed. With lotto it is four; in the case of EuroMillions, 12. If the jackpot is still not won once the rollover limit is reached, the jackpot pool that has accumulated rolls down and is shared among winners of the next highest prize tier.
11. Beginning on February 1, 2009, and for a further 12 quarters these transfers were at the rate of £73 million per quarter. In addition, on May 1 and August 1, 2012, two further transfers of £68 million were made. The London Olympics and Paralympics were held during August-September 2012.
12. One of the authors was once assured by a very senior psychiatric consultant that in more than 35 years she had yet to come across anyone who was addicted to lotteries. This view is confirmed by a succession of prevalence surveys which have shown that playing National Lottery games is associated with relatively low prevalence rates of problem gambling (Wardle et al 2011). Nevertheless, it should be remembered that the meaning of addiction in economic theory may bear only an indirect relationship with that employed in the study of problem gambling.
13. In the lottery trade a particular c is termed a price point. In the United States, large increases in sales of lotteries heavily dependent on scratchcards have been attained by inducing a rise through the price points over time. This has often been accompanied by a rise in k (or the prize percentage payout).
14. One example of this is the fact that the National Lottery operates largely outside the Gambling Act 2005 under which most of the rest of the British gambling industry is regulated.
15. In 2010, for example, the NLC refused Camelot Group permission to use National Lottery terminals to supply payment services.
16. In practice it is likely that the demand curve for scratchcards is different from that for IIWGs. Given the direct competition faced by National Lottery IIWGs from other gaming websites, the demand for IIWGs is likely to be more price-elastic than that for scratchcards.

For simplicity's sake, and because sales of IIWGs are small relative those of scratchcards, they are aggregated for the purpose of this argument.

References

Ashton, John. [1893] 2011. *A history of English lotteries: Now for the first time written.* Reprint, London: General Books.

——. [1898] 2011. *The history of gambling in England.* Reprint, Milton Keynes, U.K.: Lightning Source.

Becker, Gary S., and Kevin M. Murphy. 1988. A theory of rational addiction. *Journal of Political Economy* 96(4):675–700.

Casey, Emma. 2007. *Women and UK National Lottery play.* Birmingham, U.K.: National Lottery Commission.

Cohen, Jacob. 1953. The element of lottery in British government bonds, 1694–1919. *Economica* 20(79):237–246.

Conlisk, John. 1993. The utility of gambling. *Journal of Risk and Uncertainty* 6(3):255–275.

Cook, Philip J., and Charles T. Clotfelter. 1993. "The peculiar scale economies of lotto. *American economic Review* 83(3):634–643.

Creigh-Tyte, Stephen, and Lisa Farrell. 2003. Is the UK National Lottery experiencing lottery fatigue? In *The economics of gambling,* edited by Leighton Vaughan Williams. London: Routledge, 165–181.

Creigh-Tyte, Stephen, and John Lepper. 2004a. Survey of participation in, and attitudes towards, Gambling: Key findings from the 2004 NOP Survey. Technical Paper No. 4. London, Department for Culture, Media and Sport.

——. 2004b. Gender differences in participation in, and attitudes towards, gambling in the UK: Results from the 2004 NOP Survey. Technical Paper No.7. London: Department for Culture, Media and Sport.

Department for Culture, Media and Sport (DCMS). 2011. National Lottery distribution fund account 2010/11. London: Stationery Office.

Ewen, C. L'Estrange. [1932] 1972. *Lotteries and sweepstakes: An historical, legal, and ethical survey of their introduction, suppression and re-establishment.* Reprint, New York: Benjamin Blom.

Farrell, Lisa, and Ian Walker. 1999. The welfare effects of lotto: Evidence from the UK. *Journal of Public Economics* 72(1):99–120.

Farrell, Lisa, Edgar Morgonroth, and Ian Walker. 1999. A time series analysis of UK lottery sales: Long and short run price elasticities. *Oxford Bulletin of Economics & Statistics* 61(4):513–526.

Farrell, Lisa, Roger Hartley, Gauthier Lanot, and Ian Walker. 2000. "The Demand for Lotto: The Role of Conscious Selection", *Journal of Business & Economic Statistics* 18(2):228–241.

Forrest, David. 2003. Time-series modelling of lotto demand. In *The economics of gambling,* edited by Leighton Vaughan Williams. London: Routledge, 182–203.

Forrest, David, O. David Gulley, and Robert Simmons. 2000. *National Tax Journal* 53(4): 853–864.

——. 2004. Substitution between games in the UK lottery. *Applied Economics* 36(7):645–651.

——. 2010. The Relationship between betting and lottery play. *Economic Inquiry* 48(1):26–38.

Forrest, David, Robert Simmons, and Neil Chesters. 2002. Buying a dream: Alternative models of demand for lotto. *Economic Inquiry* 40(3):485–496.

Hancock, Linda. 2011. *Regulatory failure? The case of Crown Casino.* Melbourne: Australian Scholarly Publishing.

Hartley, Roger, and Lisa Farrell. 2002. Can expected utility theory explain gambling? *American Economic Review* 92(3):613–624.

Hartley, Roger, and Gauthier Lanot. 2003. On the design of lottery games. *Journal of the Operational Research Society* 54(1):89–100.

Ipsos MORI. 2011 Underage gambling in England and Wales: A research study among 11–16 year olds on behalf of the National Lottery Commission; www.natlotcomm.gov.uk/assets-uploaded/documents/ipsos-mori-report-on-behalf-of-the-commission_1316537803.pdf.

Lepper, John, and Deborah Hawkes. 2007. *Gender and the jackpot: Women, men and the National Lottery.* Birmingham, U.K.: National Lottery Commission.

Mikesell, John L. 1994. State lottery sales and economic activity. *National Tax Journal* 47(1):165–171.

Munting, Roger. 1998. The revival of lotteries in Britain: Some international comparisons of public policy. *History* 83(272):628–645.

National Lottery Commission. 2011. Guardian of a great national asset: National Lottery Commission annual report and accounts 2010/2011. Birmingham, U.K.: National Lottery Commission; www.natlotcomm.gov.uk/assets-uploaded/documents/annual-report-20102011_1311934839.pdf.

Officer, Lawrence H., and Samuel H. Williamson. n.d. Purchasing power of British pounds from 1264 to present. MeasuringWorth.com; www.measuringworth.com/ppoweruk (accessed Oct. 22, 2011).

Paton, David, Donald S. Siegel, and Leighton Vaughan Williams. 2003. The demand for gambling: A review. Chapter 18 in *The economics of gambling*, edited by Leighton Vaughan Williams. London: Routledge, 247–263.

———. 2004. Taxation and the demand for gambling: New evidence from the United Kingdom. *National Tax Journal* 57(4):847–861.

Raven, James. 1991. The abolition of the English state lotteries. *Historical Journal* 34(2):371–389.

Reith, Gerda. 1999. *The age of chance: Gambling in Western culture.* London: Routledge.

Roger, Patrick. 2009. *The demand for Euromillions lottery tickets: A European comparison.* Working Papers of LaRGE Research Center No. 2009-05. Laboratoire de Recherche en Gestion et Economie (LaRGE), Université de Strasbourg (France) ifs.u-strasbg.fr/large/publications/2009/2009-05.pdf.

Slater, Don, and Eva Neitzert. 2007. *What Does the Nation Dream About? Findings of the National Lottery Dreams Study.* Study funded by National Lottery/Camelot Group, London, Enterprise LSE.

Vaughan Williams, Leighton, ed. 2003. *The economics of gambling.* London: Routledge.

Walker, Ian, and Juliet Young. 2001. An economist's guide to lottery design. *Economic Journal* 111(475):F700–722.

Wardle, Heather, Alison Moody, Suzanne Spence, Jim Orford, Rachel Volberg, Dhriti Jotangia, Mark Griffiths, David Hussey, and Fiona Dobbie. 2011. *British gambling prevalence survey 2010.* London: National Centre for Social Research.

Williams, Penry. 1956. Lotteries and government finance in England. *History Today* 6(8):557–561.

Woodhall, Robert. 1964. The British state lotteries. *History Today* 14(7):497–504.

CHAPTER 32

THE BENEFITS AND COSTS OF SLOT MACHINE GAMBLING

SCOTT FARROW AND CHAVA CARTER

1 Introduction

Consider designing a meta-analysis to determine the net economic benefits and costs of slot machine gambling in a specific jurisdiction. The major steps would likely be (1) identifying studies that have internal and external validity, (2) collecting outcome data from these studies, (3) gathering the key characteristics of the studies, and (4) running a regression to identify the statistically significant determinants of the net benefits. Attempting such an analysis quickly founders on the paucity of studies that would pass step 1, in large part due to ongoing disputes among economists as well as other parties on the elements of an internally valid study.

Consequently, this chapter begins by introducing major issues and defining the slot machine segment of the gambling industry. Section 2 reviews basic economic welfare criteria as implemented via benefit-cost analysis, sections 3 through 9 summarize issues in the conceptualization of benefits and costs as well as framing issues, section 10 introduces a scorecard for key elements of a benefit-cost analysis, section 11 summarizes illustrative empirical studies, and section 12 concludes with areas for further research.

1.1 Who Defines Benefit-Cost Analysis?

Within economics and applied policy analysis there is a large body of literature on theoretical welfare economics and its implementation through benefit-cost analysis, including such texts and collections as Baumol and Wilson (2001), Boardman et al. (2011), Just, Hueth, and Schmitz (2004), Zerbe and Bellas (2006), Schmitz and Zerbe (2009), Brent (2006), and Jones (2005), among others. The core of the theory is static, partial equilibrium analysis with certainty, but the literature has expanded to include ever more complex analyses involving general equilibrium, dynamics, and uncertainty (e.g., Baumol and Wilson 2001; Arrow and Lind 1970; Goulder and Williams 2003; Graham 1981; Freeman 2003). Economists understand that the result of an applied welfare or benefit-cost analysis, while built from potentially positive or objective analyses of many components, ultimately contains a normative element when individual values are aggregated to form "bottom line" measures, such as the aggregated net present social value (e.g., Baumol and Wilson 2001; Foster and Sen 1997; Adler and Posner 2006; Lave 1996). Nonetheless, many economists and policy analysts, us included, advocate the use of benefit-cost analysis as providing an accounting structure for the integration of positive and negative impacts into monetary terms. Further, a demand exists for benefit-cost information as demonstrated when many governments and organizations require benefit-cost analysis as part of regulatory or investment processes (e.g., U.S. OMB 1992, 2003; EU–Regional Policy 2008; HM Treasury 2013; Treasury Board of Canada 2007).

Although general elements of theory exist, and even though textbooks outline many simplified applications, such as taxes, pollution regulations affecting health, changes in market structure, and so on, any particular application typically involves theoretical and empirical customizations to the problem context. For gambling, and slot machine gambling in particular, that customization has not yet solidified into a canonical form of analysis. Key elements without consensus remain the definition of price and its relation to consumer surplus measures, the integration of uncertainty, the classification of various impacts among transfers and externalities, and partial versus general equilibrium. This lack of consensus among economists, who often depend on multiple disciplines to identify and quantify the impacts of an action, has been further obscured by the interest and participation of noneconomists into the deceptively clear debate about benefits and costs. Alternative forms of integrative analysis, such as socioeconomic impact analysis or multi-attribute utility, have been used and suggested outside of the formal context of benefit-cost analysis but may at times use similar terminology. This chapter focuses on the welfare and benefit-cost issues defined by the economics literature in the hopes of improving clarity in that one domain without asserting that benefit-cost analysis is the only information that can or should be used in the analysis of decisions related to slots.

1.2 Slot Machines: Industry Definition and Relevance

The gambling industry around the world contains a huge variety of ways, legal and illegal, for people to place wagers on outcomes, including human sports, animal contests, card games, dice games, and other methods to generate random or nonrandom outcomes. Slot machines are one approach to generating a random outcome, if there is no tampering, which dates back over one hundred years (Fey 2006). The American Gaming Association (AGA 2010) defines a slot machine as "Any mechanical or electrical device in which outcomes are determined by a random number generator located inside the terminal" while defining the closely related video lottery terminal (VLT) as "an electronic game of chance played on a video terminal that is networked and can be monitored, controlled and audited by a central computer system. These games are authorized through the state lottery and considered by law to be lotteries, not commercial gaming." The designation electronic gaming device (EGD) covers both categories and represents "Any mechanical or electrical game of chance, including slot machines, video lottery terminals (VLTs), video bingo, video pull-tabs and video poker machines" (AGA 2010). For the purposes of this chapter, slots is the popular shorthand that will include both these definitions as well as the more generic EGD. Within the definition of slots there are many variations around the world which go by other names, including pokies, video poker games, fruits, and so on. Getting their name from the historical manner in which money was fed into a machine with mechanically spinning wheels (Fey 2006), modern slot machines typically accept wagers or bets as small as one cent ($.01) or as high as one hundred dollars. Current machines allow payoffs on one or more "lines" or combinations of symbols with a frequency of appearance controlled by a computerized random number generator (Turner and Horbay 2004; AGA 2010). Ultimately playing slots requires no special skills or prior knowledge of the game; neither does it require such of other players (Stewart 2010). The games have low stakes and a relatively high reward rate, and the machines themselves use sight and sound effects to encourage betting (Fisher and Griffiths 1995; Breen 2004).

While dependent on the legal regime of a given jurisdiction, slot machines can become dominant in a casino-like setting yet can also operate well in small, disbursed locations. Although this chapter focuses on physical slot machine gaming, the Internet provides new venues for slot machine gambling that have been little studied (National Gambling Impact Study Commission 1999; Australian Productivity Commission 2010; European Commission 2006). Evidence prepared for the American Gaming Association indicates that in a mature casino environment in the United States, such as Atlantic City, the share of revenue generated by slot machines has increased from 40 percent of revenue in 1978 to about 70 percent by 2010. Similarly, William Eadington (1999) reviewed data demonstrating that slots are the dominant revenue source in U.S. casinos while Richard Thalheimer and Mukhtar Ali (2003) reported that slots accounted for about 80 percent of riverboat revenues in 1998. Statistics Canada reported that about 40 percent of net revenue is earned by slot and VLT machines outside Canadian casinos and that slots dominate inside casinos, which, from slots and other games, generated 34 percent of net revenue in 2010 (Statistics Canada 2010). Thalheimer (2008) reported that after legalization of VLTs in the racetrack setting in West Virginia total wagering increased while pari-mutuel betting decreased. However, different casino and legal structures influence this balance, with historically lower levels of slots in the United Kingdom and in various locations within Europe (Eadington 2008). At the same time, a broader survey of gambling in the United States by John Welte et al. (2002) updating several earlier national surveys reported that 17 percent of respondents indicated that they had used gambling machines as compared to the two most popular forms of gambling, “Lottery,” with 66 percent, and “Office Pools, Raffles, Charity,” with 48 percent; 27 percent reported participating in casino gambling (generally including a large but undefined component for slots) as an independent category.

Although growth in a regulated market such as gambling need not follow the dictates of supply and demand, it is clear that slots are an important technology in the gambling industry. As such, it is likely to embody many of the issues that are barriers to a consensus in determining net benefits of gambling. For example, research is investigating hypotheses that slot machines may be disproportionately associated with problem gamblers as a result of the very rapid feedback from slot machines to the gambler and other elements of environmental control (Chóliz 2010; Harrington and Dixon, 2009; Williams and Wood 2004; Volberg 2001; Smith, Hodgins, and Williams 2007). Empirical results do reflect a high degree of correlation between slot machines and problem gambling. Katherine Marshall and Harold Wynne (2003) cited Canadian Community Health Survey data showing that approximately 25 percent of VLT players are at-risk or problem gamblers. Based on a prospective diary study, Robert Williams and Robert Wood (2004) attributed almost 60 percent of slots revenue in Ontario to moderate or severe problem gamblers. Rachel Volberg (2003) reported that in the United States problem gamblers are more likely to identify slot machines as their favorite type of gambling, whereas Elisardo Becoña et al. (1995) reported a disproportionately high incidence of pathological gamblers among slot machine players in Spain; notably, neither of these results appears inconsistent with the volume of gambling represented by slots. A recent review by Nicki Dowling, David Smith, and Trang Thomas (2005) concluded, however, that the existing evidence was insufficient to conclude that slots are relatively more addictive than other forms of gambling but that further research on addiction and on the characteristics of the machines and games themselves is warranted.

In a study of VLT players in Alberta, Garry Smith and Wynne (2004) found 61.1 percent to be either moderate-risk or problem gamblers (39.3% and 21.8%, respectively). About 79 percent of these problem gamblers reported spending over $300 per month (not including winnings) on gambling; in comparison, only 6.4 percent of the nonproblem gamblers spent over $100 per month on gambling, and none spent over $300. While the majority of all VLT players gave “winning” as their primary attraction to VLTs, 17.8 percent of the problem gamblers chose “excitement/thrill/rush” as compared to 5.1 percent of the nonproblem gamblers. Moderate-risk and problem gamblers most often played VLTs alone, while low-risk and nonproblem gamblers did not. Of problem gamblers 34 percent reported that VLT playing had created problems in their lives as compared to 2.4 percent of nonproblem gamblers and 0 percent of low-risk gamblers. About 58 percent of problem gamblers expressed a desire for VLTs to be removed from their communities.

Somewhat more broadly, research by Brad Humphreys et al. (2011) found substantial differences in the impacts of specific forms of gambling. The study differentiated between electronic gaming machines (EGMs) located at racetracks or casinos and those located in bars. The study, the results of which were generally consistent with those reported by Smith and Wynne (2004), also found that slot machine gambling is linked to reductions in self-reported stress levels but an increased probability of self-reported health problems. Another stream of economic research links "addiction" to a time dependence of purchases with increasing purchases over time (Becker and Murphy 1988; Guryan and Kearney 2010). With specific regard to slot machine gambling, newly detailed evidence from loyalty cards at a U.S. casino where 90 percent of the revenues accrued from slot machines indicates that about 8 percent of the gamblers could be classified as "economically addicted" (Narayanan and Manchanda 2011).

Consequently, slot machines appear to be a worthy subset of gambling for economic investigation.

2 Benefit-Cost Theory and Slots

A strength of benefit-cost analysis and its welfare foundation is that similar theoretical concepts are used across application areas. Textbooks in benefit-cost analysis typically place all benefit and cost impacts into the four categories of consumer surplus (CS), producer surplus (PS), government revenue (GR), and external effects (ES) (e.g., Boardman et al. 2011; Zerbe and Dively 1994; Bellinger 2007). Note is often made that the change in government revenue comes from the other categories, but interest in the distributional effect on government is sufficiently broad that government is standardly broken out in the above manner. Changes in these categories define the change in social welfare or social surplus (*SS*) from an action compared to a baseline, so that

$$\Delta SS = \Delta CS + \Delta PS + \Delta GR + \Delta ES = \text{Total Benefit} - \text{Total Cost}.$$

The same textbooks show that total benefit less total cost is equivalent to the surplus factors outlined above when benefits and costs are appropriately defined. Nonetheless, the broader use of benefits and costs is so intuitive that users may not realize the framework provided by the more jargon-based use of "surplus" measures from which the theory is developed. In general, "surplus" refers to the gain or loss to participants on the early or intra-marginal units obtained, even if they only break even on the final unit (Boardman et al, 2011; Australian Productivity Commission 1999; Walker 2007a; Grinols 2004). For producers, the surplus measure is akin to operating profit; for consumers, it is a monetary measure of getting a good deal as you would have been willing to pay more for it.

This core theory however, ignores risk of varying outcomes, which depends on probability, clearly a central element of slot machine gambling as well as many other activities. However, it is the excitement and the bane of the current era that understanding and modeling decisions with risk (or uncertainty) remain unsettled. The theoretical framework is that of expected utility (Baumol and Wilson 2001; Friedman and Savage 1948; Markowitz 1952), which provides the well-known models of risk aversion, risk loving, and risk-neutral preferences, but sufficient observation of behavior counter to the theory has prompted proponents of the behavioral economics school to seek to develop alternative risk-based theories (Thaler 1992; Machina 1987; Camerer, Loewenstein, and Rabin 2004; Starmer 2000). As a consequence, measures that integrate risk have been proposed to replace some measures, especially consumer surplus, in equation 1. Each of the following sections takes one of the four components of equation 1 and identifies key theoretical issues in their application to slot machine gambling.

3 Consumer Surplus for Slots

This core theoretical element benefit-cost analysis has been much discussed in regard to gambling but is often omitted in practice (Australian Productivity Commission 1999; Walker 2007a; Grinols 2004; Crane 2008). To focus on consumer surplus the analyst assumes that gambling is like other products in which probability does not play an explicit role. In particular, the consumer is expected to gain satisfaction (utility) from the act of gambling regardless of winning or losing. This is gambling as entertainment. Further, the consumer must be responsive to the price of the activity as consumer surplus only exists with a downward sloping demand curve such that more is consumed if the price is decreased. The Australian Productivity Commission (APC) (1999, appendix C) used surplus as an element in its analysis of benefits while Yuliya Crane (2008) extended the consumer surplus analysis and applied it to the United Kingdom. It is important to note that these analyses were driven by information about the price elasticity of demand and the functional form of demand for gambling, about which relatively little is known. Yet the consumer surplus benefit with a linear demand curve can be estimated as gross gaming revenue divided by twice the absolute value of the elasticity (APC 1999, appendix D; Crane 2008, 161). Thus consideration of this factor can be a major component, in the neighborhood of the gross gaming revenue, if this framing of consumer benefits is utilized.

3.1 Non-normal Gambling and Consumer Surplus

Substantial interest exists across health care professions, policy analysts, economists, and the gambling industry regarding individuals who score high on diagnostic tests identifying behavior that defines a mental disorder focused on gambling or, in a possible revision of American psychiatric terminology, to addiction and other disorders (American Psychiatric Association 2010). A large literature exists on problem and pathological gamblers, those who have various personal and interpersonal difficulties associated with gambling and who are disproportionately the source of gambling expenditures and hence revenues (National Research Council 1999; Smith et al, Hodgins, and

Williams 2007). Economics deals somewhat differently with addiction, including its various forms, such as drug, alcohol, and tobacco use and, more recently, gambling. In the economic models of time-consistent and rational consumers there is the possibility of "rational addiction" where such behavior represents an individual's unique preferences (Becker and Murphy 1988; Gruber and Köszegi 2001). Related to the emerging behavioral economics literature are counterarguments as well as evidence that some "addicts" would prefer to exist in an unaltered state and are willing to pay to change their behavioral accordingly but are unable to do so (Vining and Weimer 2010). While the issue is still seeking resolution, an alternative consumer surplus model for some gamblers has been presented by the Australian Productivity Commission (1999), David Weimer, Aidan Vining, and Randall Thomas (2009), and Vining and Weimer (2010) in which an adjustment is suggested for those addicted whose demand does not reflect their "true" preferences. In such cases the addicted person, perhaps a gambler, is modeled as receiving the benefits that a "normal" gambler would receive should a normal gambler engage in a large quantity of gambling, but the addicted gambler is not accorded the full surplus that that gambler's "addicted" demand curve appears to imply. Consequently there is a downward adjustment in the surplus apparently accorded to addicted gamblers based on their observed behavior. This is illustrated in figure 32.1, where the blue area is the consumer surplus for a normal gambler, determined in the standard way as the area beneath the "normal" demand curve and above the price line. The "addicted" gambler has the larger demand curve, D_A, and consumes a larger quantity of gambling, Q_A, than does a normal gambler. The adjustment suggested, for which Weimer, Vining, and Thomas (2009) as well as the APC (1999) developed formulas for specific functional forms, is to use the normal demand as the reference point so that an addicted gambler still receives the blue-shaded consumer surplus accorded a normal gambler

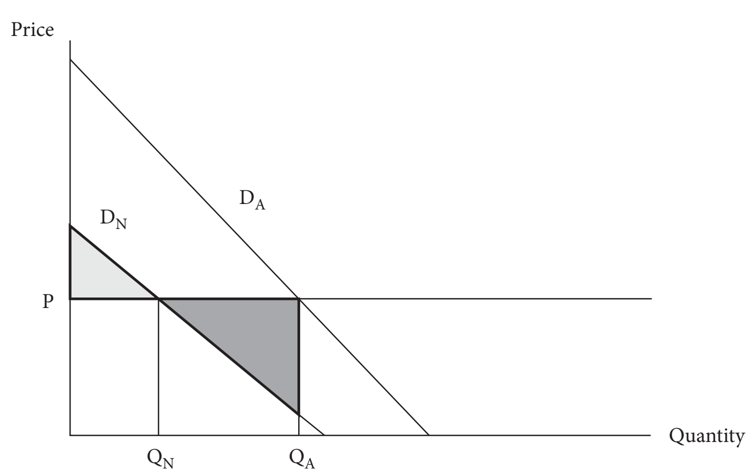


FIGURE 32.1 Surplus for a normal and addicted gambler

while the addicted gambler loses the value colored orange given the specific illustration in the figure.

This innovation in consumer surplus for non-normal gamblers usefully breaks up the monolithic composite consumer of gambling into two types while clearly paving the way for greater heterogeneity of gambling consumers.

The distinction between normal and addicted gamblers may be especially important for slot machines in ways that are not yet fully understood. First, if all types of gambling are equally "addictive," then the large proportion of slot machine revenues to the gambling industry would indicate an expected association with slot machines and the observed number of non-normal gamblers. However, work such as that by Mariano Chöliz (2010) and Harrington and Dixon (2009) suggest that slot machines may be more than averagely addictive. If so, any associated consumer surplus adjustments, and impacts discussed as externalities below, may be increased when analyzing slot machines if further research justifies such an empirical adjustment.

3.2 Location and Distance Consumer Surplus

Distance to a venue is an important determinant of consumer gambling behavior (e.g., Thompson, Gazel and Rickman 1995; Grinols 2004; Baker and Marshall 2005). An additional component of consumer surplus links consumer value to the distance to their gambling locations as introduced by Earl Grinols (1999, 2004) and used by PolicyAnalytics (2006) and Scott Farrow and Judith Shinogle (2010). Grinols developed a model where the usual consumer surplus based on the direct gambling price does not change, but instead policies that expand gambling locations appear to increase frequency and change the expenditures of existing gamblers and expand the set of those who gamble. Grinols modeled distance as entering directly into a utility function and used industry data to infer modest benefits for an average resident who may or may not gamble. He modeled a representative consumer as deriving utility from gambling, v, a composite of other goods,x, and enjoyment, $E(g,m)$, that is a function of the amount gambled per visit, g(also said to be price or cost), and distance, m, to define a utility function of the general form $U(x,v, E(g,m))$. Grinols (1999, 2004) continued by deriving an optimal expenditure function and specifying two constant elasticity forms for the utility function. The specific function was calibrated based on data such as income, the average amount gambled per visit, and the number of visits to derive monetary values for increased value, which are conveniently presented in tabular form. As this was derived for a representative consumer, it appears to apply to all citizens, though some number of people chose not to gamble.

While integrating site accessibility into value appears useful and insightful, Grinols's approach is unusual as utility is modeled as a function of enjoyment and prices are combined with quantities in the utility function. In contrast, there is an important history of using travel cost models to infer benefits for recreational consumers as surveyed in Freeman (2003). A direct travel cost approach, noted by Grinols, builds quality of a site into utility while random utility models build discrete choice for a consumer among qualitatively different sites. These models are not without problems, but Grinols's approach appears to be a different and relatively unexplored approach.

The role of distance and its link to consumer value appears to be a useful direction for further research designed to integrate the distance estimation into more widely accepted procedures. Internet slots or other online gambling also may open up new ways of assessing distance benefits, as there is effectively no distance between computer access and a gambling site, though the quality of site may differ from a traditional site. Speculatively, such an analysis will bring to the fore the previously ignored value of individual time as part of the price of gambling and highlight the question "What is the price people pay to gamble?" This question will be briefly investigated in section 4.

3.3 Risk-loving Preferences, Integrated Models, and Option Price

Modeling slot machine gambling as one of many sources of entertainment seems to capture some element of the activity, but would people gamble as much if there were zero probability of winning more than the amount wagered? Many survey respondents reply that they gamble in the hopes of winning, as discussed above. If so, the consumer surplus model of the gambler as only seeking entertainment omits an important component of gambling.

The willingness to give up money in order to accept a gamble, that is, risk-loving behavior, is a core part of basic expected utility theory (Eeckhoudt, Gollier, and Schlesinger, 2005). The amount a risk-loving person would give up in order to take an unfair bet, one whose statistical expected value is less than the wager, will vary by individual preferences and the probabilities and payoffs. A person with those preferences will rationally accept an unfair bet. Famous economists tried to resolve the paradox that many people both gamble and buy insurance (Friedman and Savage 1948; Markowitz 1952). That paradox has not been fully resolved, though there are many theoretical contenders (Starmer 2000).

Risk-loving preferences suggest an alternative measure to consumer surplus for the maximum willingness to pay in order to maintain the same level of utility. The ex-ante amount a person is willing to pay to accept a gamble has been called an option price. While usually framed as the amount a risk-averse person would give up to avoid variability in income, for risk-loving persons it is the amount they are willing to pay to have access to the gamble. This ex-ante option price has not to the author's knowledge been used to value the benefits of slots to consumers although it directly models the consumer's intent to gamble. Such an approach is implicit, however, in the market equilibrium of gamblers investigated as part of the longshot bias literature (Gandhi 2008; Humphreys and Weinbach 2010).

It is not surprising that models have been developed that allow gambling to generate both pleasure and variability in income (Conlisk 1993; Sauer, 1998). These models, however, have retained the risk-averse assumption for consumers such that gambling occurs because the pleasure of entertainment dominates the risk aversion. While this may be appropriate for some gamblers, it may also suggest that for some people, those who may be risk loving and gamble at least in part to improve their wealth, risk preferences and entertainment purposes mutually create demand for gambling instead of partially offsetting each other.

Further, behavioral economics identifies several ways that people have difficulty assessing probabilities in a purely rational way and may instead apply various heuristics to situations (Thaler 1992; Camerer, Loewenstein, and Rabin 2004; Starmer 2000). With respect to gambling, particular attention has been paid to implications of the behavioral "Law of Small Numbers" (Tversky and Kahneman 1971; Narayanan and Manchanda 2011). Generally, consumers may have difficulty with small samples and independent probabilities, mistakenly assuming either a negative correlation between outcomes (Gambler's Fallacy) or a positive correlation ("hot hand") in the context of what are actually random draws. A behavioral approach, and less than perfectly informed consumers, may well lead individuals to believe probabilistic properties of slot machines that are at odds with their actual design. For instance, following a mistaken "hot hand" belief, consumers may misinterpret such signals as "near misses" on a slot machine line as indicating that one is closer to winning on the next trial than is justified by a random draw.

Taken together, the theory of benefits to consumers indicates that consumers are heterogeneous, that multiple motivations may exist, and that consumer surplus or risk-based measures are the appropriate constructs to evolve into positive models of the gambling consumer, though no single widely accepted approach is agreed upon in the literature.

4 Framing Issues

The analysis of the consumer led directly into controversial issues in conceptual measurement and illustrates the usefulness of framing the background conditions for the analysis. Motivated by behavior of the consumer but also relevant to business (producer's surplus), government, and external factors, the issues of marginal effects, price, standing, partial and general equilibrium, and employment benefits are described here.

4.1 Marginal Effects

The incremental costs and benefits compared to a well-defined baseline are the desired measures for a benefit-cost analysis. The baseline of what would exist in the absence of a policy can be difficult to determine. The standard baseline is the status quo of what exists, for instance, what gambling may be permitted in and around a jurisdiction being analyzed. However, there may be some dynamic aspect to the baseline as well, for instance, if a neighboring jurisdiction might be expected to retaliate by altering their conditions for gambling if a new jurisdiction changed their position (Walker 2007a). Similarly, for an individual, the benefits and costs should be determined incrementally. The analyst may observe that a person becoming a problem gambler increases social costs by a particular magnitude, but what is the baseline? Problem gambling can often coexist with depression, drug and alcohol abuse, and higher levels of stress (Walker and Barnett 1999; Walker 2007b; Eadington 2003). Empirical difficulties arise in attempting to determine whether problem gambling acts as a catalyst for these conditions or vice versa, particularly if the direction of the effect varies by individual. Comorbidity likewise raises questions as to the degree to which problem gambling influences negative outcomes. The U.S. National Opinion Research Center (NORC) compares expected rates of negative outcomes for nonproblem gamblers with rates for problem gamblers, attributing the difference (after controlling for chance and for confounding factors) to gambling. The Australian Productivity Commission (1999, 7.11 and 9.9) employed a "causality adjustment" based on the premise that approximately 20 percent of problem gamblers would have experienced a given negative outcome even in the absence of their gambling problem.

4.2 Price

Price is fundamental to the analysis of both consumers and producers, but it is a slippery concept when one is buying a service that may have multiple outcomes, such as a payoff or no payoff. The *Oxford Dictionary of Economics* (2009) defines price as the "amount of money paid per unit for a good or service" while noting that defining price can be more complex for some goods. In the case of gambling, what appears to be the direct price for access to the gamble (the good or service) may in fact return money to the purchaser after the outcome is determined. The gross price initially bet is generally called the wager (American Gaming Association 2010), but most economic analysis (e.g., Eadington 1999; Paton, Siegel, and Vaughan Williams 2004; Thalheimer 2008) defines price as the long-run cost from repeated play expressed as a percentage of the amount wagered, though the price for any particular run of bets may vary substantially and, with some probability, be positive. Thus if a slot machine is designed to pay back on average 95 percent of the amount wagered, the price is said to be 5 percent of the amount wagered. The definition of price is important in the estimation of consumer surplus, which depends on consumer responsiveness to price; in the definition of effective tax rates; and in defining the seller's revenue and tax obligation (Clotfelter and Cook 1990; Paton, Siegel, and Williams 2004).

Travel costs add an additional complication to the price of gambling. The full cost of the gambling experience can involve some, perhaps large, travel costs though other co-activities also may also, such as dining, entertainment, and so on. The benefit-cost literature has made heavy use of the hedonic travel cost model whereby surplus is inferred from the costs incurred when people choose different destinations based on the characteristics of the destination (Freeman 2003). Such an approach seems not to have been used in the gambling literature, but Grinols's (1999, 2004) accounting for distance consumer surplus is a similar concept.

4.3 Standing

The issue of standing in benefit-cost analysis usually applies to a geographic area or to a characteristic of the consumer (Boardman et al. 2011; Zerbe and Bellas 2006). Defining whose benefits and costs are to be included through a determination of standing is particularly important to slot machine studies, which are often focused on a region or smaller political jurisdiction. In addition, the determination of standing can affect how transfers among parties are defined.

Regional analysis creates several problems in the definition of standing. If standing is defined as the citizens of a state, then only costs and benefits to those individuals count. This definition plays a central role in many gambling studies, as what is gained as a benefit by one jurisdiction, such as governmental revenues and producer surplus from locally owned businesses, may be lost as a cost to another jurisdiction. A regional analysis will only show the limited benefits or costs to its own citizens. The choice of standing may make it difficult to determine the "ownership" of some impacts. A problem gambler may be from another jurisdiction, or business owners may not in fact reside in the state. This issue can also be a policy motivation to legalize gambling when it appears that some benefits of gambling, such as government revenue, are "lost" to another jurisdiction, a framing that implies some limitation on standing.

One characteristic affecting the standing of a consumer that has generated debate is that of thievery. As some believe that gambling increases crime, the issue is whether the thief has standing in a benefit-cost analysis (Grinols 2007; Walker 2007b). If standing is granted, then crime is partially a transfer as the thief gets a benefit from what is stolen and the victim both loses what is stolen and may well suffer nonmonetary damages. The more common approach is to use the law as declaring a social value that declines standing for the thief and hence to not count the benefit, unless the law itself is being analyzed (Boardman et al. 2011, 39).

4.4 Partial and General Equilibrium, and Employment

The majority of benefit-cost analyses are partial equilibrium that includes impacts in one or a few markets, in contrast with general equilibrium, which seeks to take into account a larger number of interrelated markets. A partial equilibrium analysis is most easily justified when markets are believed to be reasonably competitive and without major distortions, including tax distortions (Goulder and Williams 2003; Boardman et al. 2011; Baumol and Wilson 2001; Chetty 2009). Time and resources may also limit analyses to the direct impacts modeled in a partial equilibrium analysis, and analytical methods in practice may include a behavioral response reflecting feedbacks from other markets into the main markets of concern (Boardman et al. 2011; Chetty 2009). However, the conceptual basis for general equilibrium analysis is well established, and in cases with large tax distortions, significant unemployment, or many and large substitutes or complements the general equilibrium analysis becomes more important (Hazilla and Kopp 1990; Goulder and Williams 2003). However, the information required to implement a general equilibrium analysis typically involves cross-market impacts which may not be available, such as cross-price elasticities. Econometric models are available for some markets but may lack the detail necessary for a relatively minor market such as gambling (e.g., Nevada Commission on Economic Development 1999; Treyz and Treyz 2002). Alternatively, analysts often resort to the assumption of fixed-proportion production relations as modeled in input-output analysis (e.g., Miller and Blair 2009; Regional Economic Applications Laboratory 2003). The work of Grinols (2004) builds on a macroeconomic trade approach that is more analogous to a general equilibrium approach, though the framework is relatively less common in benefit-cost analysis. Consequently, many applied general equilibrium analyses are substituting another set of assumptions for those used in partial equilibrium analysis. Ideally, the analyst may investigate the impact of alternative framing assumptions, but this is seldom done.

Whether or not employment benefits exist is a high-profile implication of the choice of partial or general equilibrium analysis and of the determination of standing. The costs of employment are almost universally included in costs of operation, part of the determination of producer surplus. However, advocates of gambling (as well as many other regional development projects) tend to list employment as an important benefit of the gambling industry. Standard guidance indicates that such benefits are to be included only in well-defined and limited circumstances. For instance, guidance from the U.S. government (U.S. OMB 1992, 6) states that generally "analyses should treat resources as if they were likely to be fully employed. Employment or output multipliers that purport to measure the secondary effects of government expenditures on employment and output should not be included in measured social benefits or costs." The default presumption in benefit-cost is that of well-functioning markets in which unemployment is at its natural rate due to turnover and transitional issues. Similar default guidance appears in textbooks, such as Boardman et al. (2011), with caveats as discussed below. In the natural rate of unemployment (full-employment) case, labor is paid its opportunity cost at the margin, the payment just compensates the employee for giving up his or her time, and no incremental employment benefits accrue in the labor market.

However, two important cases may justify an employment benefit that is a portion of labor expenditures. The core exception is a significantly higher rate of unemployment than the natural rate, with *significant* typically a matter of judgment. Such a higher rate of unemployment indicates disequilibrium in the labor market and the potential for labor expenditures to exceed the opportunity cost of labor, creating a partial, additional benefit to the worker in excess of the cost to induce the labor supply. Recent discussions of accounting for labor in times of high unemployment include Boardman et al. (2011) and Haveman and Farrow (2011), which contain guidance on assumptions for the amount of labor expenditures that might be considered as an additional benefit. The second case is a subset of the first when a regional analysis, as for a city or state, defines whose costs and benefits have standing in the analysis. While it would still be necessary for high unemployment to exist in the receiving jurisdiction to provide the potential for additional labor benefits, a regionally focused analysis may exclude lost jobs in other jurisdictions and thus overstate benefits from a broader perspective.

5 Producer Surplus

Conceptually similar to consumer surplus, producer surplus is typically measured as the excess of price over marginal cost for the appropriate market duration of short or long run. As the marginal cost is driven by technological and regulatory considerations in the industry, the result is case dependent. In a constant-cost, perfectly competitive industry, in the long run there would be zero producer surplus while in a market with regulatory protection against entry there may be significant producer surplus, though the potential exists that competition for economic rent may dissipate a potential surplus (Walker 2007a).

Regional gambling impact studies may include changes in local profits (Anielski and Braaten 2008), although some benefit-cost analysts assume with little justification that there is no producer surplus in the gambling industry while others apply an average rate of profit pending better information.

Two important modeling issues arise in estimating producer surplus beyond the issue of marginal cost. The first is whether to include taxes (gross surplus) or to exclude them (net surplus). As there is typically substantial interest in government revenues in the analysis of gambling, most producer surplus measures would be net of taxes, although at times it is useful to be clear that the tax revenues can also be accounted for as reductions in consumer and producer surplus (Boardman et al., 2011; Krutilla 2005). The second issue is the selection of partial or general equilibrium analysis and the role of substitution and complements to gambling. If a jurisdiction is considering gambling, other industries may be affected, such as substitute entertainment opportunities or other purchases (Grinols 2004, 2007; Walker, 2007a). If in moving toward a general equilibrium analysis one includes additional markets, then producer (and consumer) surplus may be lost or gained in other markets as well.

6 Net Government Revenue

The change in government revenues was identified as a core element of net benefits in equation 1. It is common in some analyses to read the statement that taxes are transfers from consumers and producers to the government and so cancel out as a negative item to one party and a positive item to another (U.S. OMB 1992, 2003; Boardman et al. 2011, 85–96). This "exact" netting out is only true if such transfers are costless to implement and have zero efficiency cost in contrast to an approach that explicitly recognizes such costs or the efficiency cost of taxation, termed the marginal excess burden (Boardman et al. 2011). For some applications, such as gambling, the impact on government revenue is a central issue of concern, and it is useful to reduce consumer or producer surplus and track the changes in government revenue. Also, in a regional analysis when those outside the jurisdiction are taxed, the "zeroing out" of transfers need not apply. Consequently, most gambling studies track the changes in governmental revenue.

Changes in net revenue may result both from added expenditures and added revenues. If gambling or its removal is being considered in a jurisdiction, then there are likely impacts on monitoring and enforcement costs incurred by the government. In some cases this may be transparent as new agencies are established by the government to oversee gambling activities; in other cases there may be a diversion of funds used for other purposes to cover new oversight. Increasing revenue is often an explicit purpose of the change. In such circumstances the tax or other named revenue from gambling is an important benefit so long as it is not double counted as part of consumer or producer surplus.

Several interpretative issues exist related to price, effects on competing goods and services, and the earmarking of funds. With regard to price, the effective tax rate is importantly determined by the specific regulation and the calculation of price. For instance, assume a tax regulation requires 25 percent of the casino slot machine "win" (wagers minus payout) to be paid to the government. Assume the win is 5 percent of the amount wagered; for a $1 bet the "tax" collection is 1.25 cents, a small percentage of the amount wagered (1.25%) but 25 percent of the price based on the payout rate. With regard to competing goods, the government may be competing with itself and observe a decline in some already legal types of gambling if new types of gambling are introduced (Walker 2007a). This is a specific illustration of general equilibrium concerns where the analyst may wish to include other significantly affected markets. Finally, changes to the legal status of gambling are sometimes justified by earmarking changes in government revenue to particular purposes, such as for the elderly or education. This political linking of source of funds and use of funds is counter to standard benefit-cost practice, which assesses the opportunity cost of funds based on a discount rate. Alternatively, the use to which the funds are to be put could be analyzed through a benefit-cost analysis by defining several different alternatives, such as, (1) raise funds via gambling (which can have a marginal burden of taxation), (2) raise funds directly via taxes, which is likely to include a marginal excess burden of taxation, and (3) combine a specific use of funds with both fund-raising methods and include doing nothing as an alternative. Such an analysis would at least indicate whether a potential efficiency improvement exists; it would not, however, indicate whether the largest potential improvement is being chosen, as the uses of the funds are restricted to a small set.

7 External Effects and Shadow Prices

Market imperfections have a long history in economics, and various cases are well worked out demonstrating changes in economic efficiency and transfers from one party to another. Monopoly market power is one early example; the rise of environment and health issues brought externalities, public goods, and information asymmetries to the fore. A concern about gambling is that there may be effects, often detrimental, on other parties who are not part of the direct purchase of gambling services. Examples include concern about additional criminal activity, including white collar fraud or burglary; community well-being" or negative effects on household and friends often involving money or behavior linked to problem gamblers; and so on. Numerous papers, especially impact analyses identify various candidate effects (Anielski and Braaten 2008; Grinols 2004; Volberg 2003; Thompson, Gazel and Rickman 1997). In contrast to impact approaches, economic welfare approaches (Walker 2007b; Collins and Lapsley 2003; Eadington 2003) are likely to place impacts on those actually gambling primarily in the consumer or producer categories while putting into the external category those affected but who were not a part of the original gambling transaction between buyer and seller. The default economic position is that voluntary exchanges must benefit both parties in some way, not necessarily financial, while involuntary transactions may not be fully considered in prices and result in an externality. Douglas Walker and A. H. Barnett (1999), Walker (2007a), Grinols (2007), and Eadington (2003) worked to apply various definitions of social costs that are related to but somewhat distinct from those in broader use in benefit-cost analysis. For instance, Boardman et al. (2011, 91) used a typical definition of externalities from the environmental literature, stating that an externality is "an effect that production or consumption has on third parties-people not involved in the consumption or production of the good." John Roman and Graham Farrell (2002) linked this approach to the crime literature where businesses, government, or other actors may alter incentives for criminal activity and so externalize some of the actions of an industry. The potential for increased burglary or other criminal activity to support compulsive gambling is an example. Or consider fraudulent behavior to obtain money, illustrated in the nonfiction book and movie *Owning Mahowny*, which are about a compulsive gambler who creates fraudulent loans at his bank job in order to support his gambling (Ross 1987). Increased criminal activity, whether white collar or street level, would generally be considered to create involuntary harm to third parties and constitute an externality (whether a thief has standing was discussed in section 4.2).

Information asymmetries are also considered a market imperfection when there is unequal information on one side of the market, for instance on the probability of default. Consider, for instance, a gambler who borrows funds from a financial institution or informally from family or friends. The financial institution will gather what information it may and charge interest on the exchange that adjusts for risk. If there is asymmetric information, such as risky unobserved characteristics of the borrower, that market will work imperfectly in the sense that the price may be too low leading to inefficiency but also result in a transfer, if legal, to the buyer in this case or conversely if the seller has the imperfect information (Boardman et al. 2011, 89). Some of the gambling literature (APC 1999; PolicyAnalytics 2006; Crane 2008) has moved toward an empirical identification of the amount transferred between creditor and debtor in the case of default but not including that amount as a social cost, instead using the resource opportunity costs involved in a bankruptcy filing.

In the face of externalities and other market imperfections, individuals, or the government on their behalf, may take defensive action to reduce the frequency or severity of an impact. Individuals may invest in better locks or avoid some areas. The government may invest in social policies to reduce social costs or its own budgetary costs (which are separate objectives), such as providing therapy or increased policing. Such defensive expenditures are frequently considered a lower bound on the social cost of the causing activity. Particularly in the case of government one may question whether the defensive expenditures are economically optimal but do represent an existing expenditure of resources to reduce a problem (Walker 2007b). While such expenditures may be relatively easy to observe empirically, they are presumably reducing the costs relative to a no-policy equilibrium. Hence a government that spends nothing on treatment may have a different profile of social costs than one which does expend funds.

Finally, some authors include elements of personal or community morality as an external cost of legalized gambling. Two aspects of benefit-cost analysis appear relevant, though neither is common in the existing literature. The first is analysis related to the provision of public goods suggesting that different jurisdictions will specialize in certain characteristics and that people will sort themselves into those jurisdictions. At times the legalization of gambling is subjected to direct vote and so reflects a political-economic representation of community values. The second approach is the inclusion of "moral values" in benefit-cost analysis (Zerbe 2002; McConnell 1997; Flores 2002). Richard Zerbe has suggested that as long as people are willing to pay for certain states of the economy, say willing to pay based on their own preferences for others not to have gambling (a paternalistic cost), then those preferences should be represented in the analysis. The analysis becomes more complex when models of nonpaternalistic altruism are included but there are some conditions where nonpaternalistic preferences are appropriate to include (Flores 2002).

While conceptual issues can continue to be clarified, there remains the search for a consensus on shadow prices (Boardman et al. 2011, chap. 16) relevant to gambling where shadow prices are the marginal efficiency cost of an impact. This author's reading is that while there is some evolution of thought toward omitting the transfer component of some elements, more work remains to be done in the conceptualization and estimation of the shadow price of numerous impacts relevant to gambling.

8 Framing Issues between Impact Studies and Benefit-Cost Analysis

In contrast to the standard economic categories and questions, impact analyses appear to gather a number of potentially positive and negative impacts together with less conceptual structure to the organization, of which the Socio-Economic Impact Analysis of Gambling framework (SEIG) (Anielski and Bratten 2008) is a leading example. From an economist's perspective, impact studies often contain double counting of some impacts, as with gambling revenues and regional income; or include partial accounting, as with including gross labor expenditures as a benefit. For instance, Wynne and Howard Shaffer (2003) listed the following items as the most frequently cited impacts of gambling in the literature. To illustrate this tension, a welfare economic perspective is provided following each point on their list, and though stated below in an assertive way, many of the complexities of estimation justify further research.

Positive impacts (bullets identify impacts as presented in Wynne and Shaffer (2003):

- revenues for the public good, including health care, education, social services, and community infrastructure:
    - *Welfare approach*: change in government revenues (as appropriately netted from consumer and producer surplus), regardless of the use of the funds but net of governmental costs (Boardman et al. 2011).
- capital projects including parks, recreation facilities, museums, and cultural arts centers:
    - *Welfare approach*: Either omit the use of funds or define the alternatives of the project more carefully. The gambling policy presumably raises funds. A government decision to provide capital projects can be funded in several ways, including raising taxes. An alternative could be defined that allows gambling with no capital projects, that allows gambling with capital projects and that allows gambling, tax increases, and capital projects. In general, the use of funds is generally analyzed separately from the source of funds.
- job creation:
    - *Welfare approach*: The default in times of full employment is that there is zero benefit from employment. In times of high unemployment or when a regional analysis denies standing to those employed in another jurisdiction, there may be benefits from employment that should be estimated as net of the reservation price of labor (Boardman et al. 2011; Haveman and Farrow 2011).
- economic development:
    - *Welfare approach*: Changes in producer and consumer surplus appropriately discounted for impacts over time, though see "capital projects" or "employment" above.
- opportunities for indigenous peoples:
    - *Welfare approach*: Use distributional weighting on net benefits accruing to different groups (Boardman et al. 2011; Farrow 2011).

- the entertainment value that gambling affords to the many players:
  - *Welfare approach*: Analyze consumer surplus as a core concept but also likely including some element of risk preferences.
- "legal" gambling formats that keep "illegal" gambling in abeyance, thus reducing crime that can be associated with unsanctioned, illegal gambling alternatives:
  - *Welfare approach*: legalizing gambling in this context would be a defensive expenditure which should explicitly be compared to an alternative where gambling is illegal. As the alternatives in this case investigate the law itself, the alternatives would likely add insight if analyzed with standing both granted and not granted to the person doing illegal activity. When gambling is legal, there is a consumer surplus; when it is illegal the same surplus may be denied standing.

Negative Impacts:

- rise in the number of people with severe gambling problems:
  - *Welfare approach*: An increase in the number of people causing social costs increases the final outcome, social costs; the number of people is a cause and not an effect.
- the havoc that problem gamblers wreak on themselves, their families, and the community at large:
  - *Welfare approach*: The impact on gamblers themselves is either excluded in a standard analysis or compared to a "normal" gambler, as in the discussion of consumer surplus. Specificity is needed on "havoc" to identify the external impacts on others and their willingness to avoid the damage caused by the problem gamblers. Care should be taken with issues involving transfers, as there may be no net effect on the economy.
- lost productivity at work:
  - *Welfare approach*: The labor market is relatively highly developed and is expected to adjust to observed productivity. To the extent there is asymmetric information or illegality, analyses of asymmetric information or crime would apply or, possibly, any real additional costs, such as training or search costs, that are incrementally increased in timing.
- increased crime, notably fraud, theft, domestic violence, suicide, counterfeiting, and money laundering:
  - *Welfare approach*: Criminal activity is viewed as an externality for which there is a benefit-cost literature. Suicide has components of both externalities and costs internalized by the individual.
- the possible cannibalistic effects that large casinos, bingo halls, and electronic gambling in bars and lounges have on local small business revenues and employees:
  - *Welfare approach*: General equilibrium effects are well understood in concept but difficult to measure in practice. The term *cannibalistic* is generally not applied to business with positive and negative cross-price elasticities.

- increased need for health care, social service, policing, and other public service costs that governments must bear to deal with the negative fallout from legalized gambling:
  - *Welfare approach*: In general, changes in cost should be considered as part of the evaluation of net changes in government revenue. However, care should be taken in defining the baseline. A positive impact item listed above by Wynne and Shaffer includes reductions in illegal activity; the governmental costs should be assessed in comparison to the stated alternative.

An even broader set of impacts is contained in the SEIG framework (Anielski and Braaten 2008), which is based around six "impact themes." A total of 60 impacts are grouped into the themes of health and well-being, economic and financial, employment and education, recreation and tourism, and legal and justice.

To illustrate the contrast in framing, the major categories in the SEIG for a regional analysis are included in figure 32.2 while a possible structure for economic analysis that focuses on markets is included as figure 32.3.

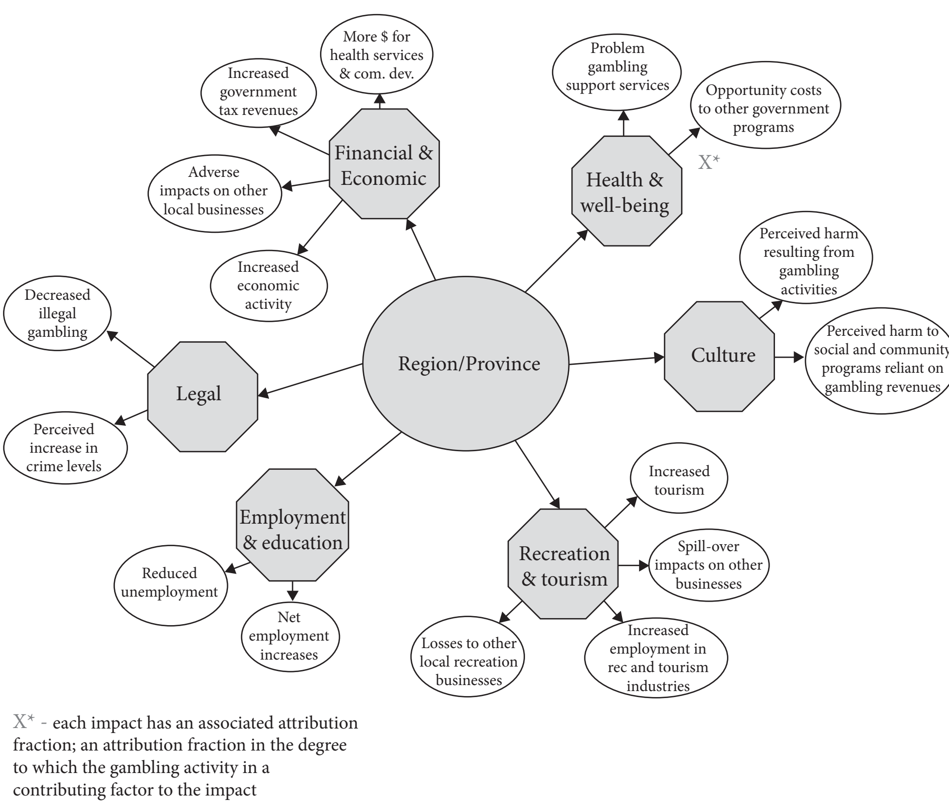


FIGURE 32.2 SEIG framing

Source: Anielski and Braaten (2008, 53)

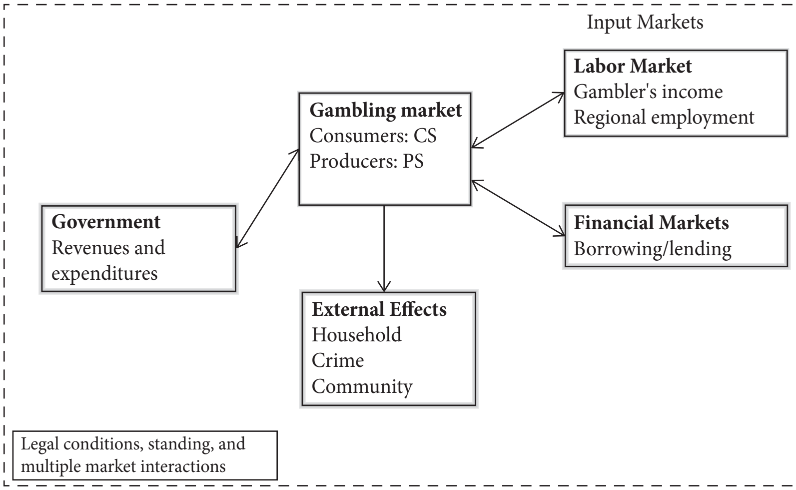


FIGURE 32.3 Economic framing

9 Distributional Effects

The default in benefit-cost analysis in the United States is that monetary valuations are added up equally across individuals with no adjustments for income status or other status aspects of individuals. Such an assumption implies an equal individual and social marginal utility of income, an assumption whose impact can be investigated through distributional analysis. In many international applications the default more frequently involves some kind of distributional weighting (Boardman et al. 2011; Farrow 2011; Brent 2006).

Governmental guidance varies significantly for the incorporation of distributional effects in a benefit-cost analysis. The U.K. government (HM Treasury 2013) specifies a particular function for weights on costs and benefits accruing to different income classes. The United States is more ambivalent, suggesting a supplemental analysis to the default of no distributional impact (U.S. OMB 2003). The people who play slot machines often span a wide range of the population, but there is evidence that gambling participation in the United States has trended such that those with lower socioeconomic status are participating more with higher financial involvement and are disproportionately affected (Welte et al. 2002). Consequently, there may be some usefulness in investigating distributional weighting of benefits and costs if the evidence suggests that benefits or costs accrue to others than the average of the population (Boardman et al 2011). As many gambling legalization issues have involved indigenous peoples, a separate weighting or net benefits analysis can be carried out for that population of concern. In an analysis of the benefits and costs of slot machine gambling in the U.S. state of Maryland, Farrow and Shinogle (2010) found that an initial determination of positive net benefits to Maryland could be reversed depending on the degree of distribution weighting, though there was significant uncertainty surrounding the estimate.

10 An Analytical Scorecard

Data collection instruments in the form of scorecards have been found useful to summarize the results of regulatory benefit-cost analyses (Hahn and Dudley 2007; Belzer 1999; U.S. OMB, 2010). Such an approach has not yet been applied to gambling studies, though there are bibliographic categorizations of gambling studies (Shaffer, Stanton, and Nelson 2006). Below is an adaptation of such scorecards to include issues relevant to gambling; in the interest of brevity, the line items are terse reminders of issues and are not repeated for what may be multiple benefit and cost categories. Such a scorecard may assist in distinguishing some impact approaches and in clarifying the nature of debates as for variations in surplus measures that have been considered (see table 32.1).

11 Illustrative Empirical Analyses of the Benefits and Costs of Slot Machine Gambling

Gambling has generated a large empirical literature on numerous subjects, including early work by Martin Weitzman (1965), which helped begin but not resolve empirical modeling debates that continue to this day. Adam Rose (1998), for a background paper for the National Gambling Impact Study Commission (1999), focused on regional impact studies as of that date and found 36 suitable for a meta-analysis, some of which incorporate slot machine use. Tom Coryn (2008) surveyed more recent analyses of various kinds but found few applications in Europe. There are relatively few complete benefit-cost analyses done specifically on slots. A computerized search for empirical studies was done using the databases Google Scholar and Econlit and search combinations including one term from each of the following two groups: Group I: economic analysis, economics, benefit cost, social costs, consumer surplus; and Group II: slot machine(s), fruit machine(s), VLTs, video lottery, EGMs, and gambling. The result yielded numerous studies but few specifically on slots. Many studies are done by consulting firms, advocates (for or against), and by state government and often reflect various problems associated with advocacy and time sensitivity. Consequently, only a few empirical cases are summarized below to illustrate the issues and results.

The illustrative benefit-cost analyses summarized in table 32.2 demonstrate the methodological variability that exists in the benefit-cost analysis of slots. The highest

Table 32.1 A Proposed Scorecard for Benefit-Cost Studies of Gambling

Item Number	Variables	Evaluation
1.	Is the economic issue identified, including whether there are market failures?	
2.	Is standing clear?... Whose benefits and costs count?	
3.	How is time incorporated? If a discount rate; source, nominal, real?	
4.	How are transfers to be accounted for, by including both sides, netted to zero or?	
5.	Is uncertainty in conditioning variables or parameters considered?	
6.	Does uncertainty alter the behavior of individuals and, if so, how is it included (e.g., risk loving)?	
7.	Is the choice of partial or general equilibrium explained?	
8.	Are distributional impacts considered?	
Estimation of Benefits		
9.	Is each benefit economically, conceptually justified?	
10.	Is each benefit quantified where possible using appropriate methods?	
11.	Is each benefit monetized where possible using appropriate methods?	
12.	If employment benefits are included, are they justified by the definition of standing or high levels of unemployment?	
Estimation of Costs		
13.	Is each cost economically, conceptually justified?	
14.	Is each cost quantified where possible using appropriate methods?	
15.	Is each cost monetized where possible using appropriate methods?	
Comparison of Costs and Benefits		
16.	Are net benefits calculated or is there a cost-effectiveness measure?	
Evaluation of Alternatives		
17.	Is at least one variation of the policy defined and analyzed?	
Clarity of Presentation		
18.	Contains executive summary or abstract	
19.	Reports impacts in natural units	
20.	Reports monetized impacts by category	
21.	Text contains summary net benefit table	
22.	Does the report appear credible and unbiased, including any appropriate consideration of source of funding?	
23.	Is the conclusion clear, as well as caveats?	
Other Comments		

Table 32.2 Illustrative Empirical Studies of Slot Machine Benefits and Costs

Application and Source	Standing	Benefit Factors	Key Cost Factors
Australian Gambling (APC 1999; Crane 2008)	Nationwide	Consumer surplus, government and community revenue	Bankruptcy, crime, emotional distress items, job-related costs, treatment costs
Indiana Riverboat Gambling (PolicyAnalytics 2006; Walker 2006)	State (subnational)	Distance CS, tax benefits, net change in profits, change in transactional constraints	Bankruptcy, crime, loss of productivity, health problems, divorce, regulatory costs net of transfers
Wisconsin Native American casinos (Thompson, Gazel, and Rickman 1995; NRC 1999)	State (subnational)	Casino spending in state and local economy	Forgone local business expenditures, social costs (crime and problem gambling)
Electronic gaming machines in Bendigo, Australia (Pinge 2008)	City	Expenditures in the local economy	Productivity loss, health costs, crime, gambling losses
Slots in Maryland (Farrow and Shinogle 2010)	State (subnational)	Consumer, producer surplus, government revenue, external costs	Numerous items associated with problem and pathological gamblers; secondary market impacts; uncertainty, distributional effects

geographic level of analysis observed is for a country, including Australia (APC 1999; 2010), the United Kingdom (Crane 2008), and the United States (Grinols 2004). The APC report is a touchstone report on many issues, and because it has an allocation of net benefits by mode of gambling, including slots, will be briefly summarized here with a more detailed critique embedded in Crane (2008). The APC report used a benefit-cost framework that included consumer surplus and change in government revenues (including community contributions that may be seen as coming from producer surplus) and considered but did not include general equilibrium effects. Producer surplus appeared to be assumed to be competed to zero, as only one mention, related to taxes, was found by an electronic search. The APC investigated social costs in great detail and formalized the adjustments for problem gamblers discussed above. Further, they considered the incremental effect of gambling and its frequent comorbidity with other socially costly outcomes and so applied a "causality adjustment" to many social costs on their reading that about 80 percent of the social costs of problem gamblers was due to gambling (APC 1999, 7.11; 9.9).

Numerous components were included in social costs driven by non-normal gambling while some debt transfers were excluded from social costs. Depending on one's framing regarding the rationality of problem gamblers, a portion of these costs may be excluded. Crane (2008) was able to analyze several different framings of these social costs. Ultimately, the APC concluded that its estimates of the net benefits of total gambling in Australia, nicely caveated, may be either negative or positive primarily depending on the elasticity of demand used in the calculation of consumer surplus. Benefits and costs were further allocated to gambling modes, with costs allocated on the basis of gambler's expenditures. Slots (gaming machines) had the largest potential to generate negative net benefits among the modes, though a range of positive net outcomes also was possible, including the second largest potential positive net benefit among the modes (APC 1999, 11.7)

The next two studies looked at the state-level impact of casino gambling while the latter two focused more specifically on slot machine gambling. The analysis of the state of Indiana's riverboat casinos, including but not limited to slots, developed a benefit-cost analysis in which only Indiana residents have standing (PolicyAnalytics 2006). Cost factors in this study included bankruptcy, crime, unemployment and loss of productivity, poor health and mental health problems, divorce, and regulatory costs with relatively careful attention paid to the role of transfers. In particular, estimates related to bad debts netted out in the estimates due to equivalent gains and losses from creditor to debtor, though resource costs related to bankruptcy were included. The report used a sensitivity analysis for social costs using two cost valuations, one based on the work of Grinols (2004) and the other based on research by the National Opinion Research Center. Benefit factors include distance consumer surplus, tax benefits (net change in state tax revenue), and the net increase in profits accruing to Indiana residents. In large part due to the importance of non-Indiana gamblers at the riverboat casinos, the net benefits to the state of Indiana are estimated at about $700 million (in 2005 dollars) with a benefit-cost ratio greater than 8. This report was reviewed by Walker (2006). While generally supportive of the analysis, including its analysis of transfers, Walker identified issues that he believed understated consumer surplus and overstated the social costs of crime due to gambling.

An earlier analysis of Native American casino gaming of all types, including slots, in Wisconsin is illustrative of hybrid impact studies (Thompson, Gazel, and Rickman 1995)and was also included in a review by the National Research Council (1999). This study is more consistent with an impact study framework than benefit-cost analysis in its definition of benefits; direct benefits were calculated as the summation of local casino expenditures, including wages, supplies, maintenance and new construction, and visitors' non-casino expenditures, such as lodging, dining, shopping, and transportation. Direct costs were calculated as the summation of forgone local business expenditures that result from residents' casino spending and expenditures by non-casino tourists who would have visited the area even in the absence of the casino and thus did not represent a new source of income. The report gives three levels of cost estimates (low, medium, and high) for comparative purposes. Indirect benefits and costs were calculated via the use of industry-based multipliers provided by the U.S. Department of Commerce's Bureau of Economic Analysis (BEA) and indicate a regional general equilibrium approach. The analysis is particularly notable for its survey of gamblers onsite to obtain information about travel, socioeconomic information, and gambling and non-gambling patterns. For the state of Wisconsin, the net effect could be either positive or negative, depending on uncertainty in the values of social cost that were used (estimates were also presented for more local economies). The National Research Council (1999) noted that this report was an advance on earlier efforts but still contained methodological weaknesses, such as in the definition of several major social cost items that likely overstated social costs.

The next two studies focused on slot machines and like the total gaming studies illustrated above, have promising aspects but also contain weakness. The benefit-cost analysis of EGM gambling in the city of Bendigo, Australia, employed an input-output methodology that hypothetically redistributed expenditures from the machine gaming industry to other industries. The goal of this impact study was to evaluate the effect of shutting down the gambling sector and redistributing expenditures to alternate sectors (retail trade and lodging and dining) within the city. While machine gambling did confer positive net benefits compared to a complete loss of gaming activity, the study showed that the gains were substantially lower than the net benefits that would be incurred by shutting down the EGMs and redistributing spending and savings to other sectors based on the amount of productivity and interregional leakage of the estimated gaming sector compared with other sectors. This is an example of consideration of an alternative which may or may not be feasible in the local context. While an advocacy document, the study illustrates the challenges in adapting available input-output models, leakages, and social costs to a smaller geographic area. In general, the benefits of those gambling were not considered and gambling losses by non-normal gamblers were counted as social costs in their entirety in an added section which, in terms of figure 32.1, would be equivalent to modeling social costs as the rectangle $P^* Q_A$ of the difference between the orange and blue areas in figure 32.1.

Research by Farrow and Shinogle (2010) continued to illustrate the problems in moving from an impact analysis to a benefit-cost analysis. Based on an impact analysis to inform a statewide vote on whether to allow slot machine gambling in the state (Shinogle et al. 2008), the benefit-cost structure was imposed while attempting to use estimates from the impact study. Heavily using results summarized in Grinols (2004) as adapted to Maryland, there was use of distance consumer surplus but not of standard surplus, of producer surplus, government net revenues, and external costs that included components of transfers in its measures of social costs. Extensions in the work included the use of Monte Carlo simulation to model uncertainty in many of the individual elements, of probabilistic benefits from employment in periods of high unemployment, and of distributional weighting based on lottery behavior in Maryland. While crudely indicating that a point estimate of the annual net benefit was likely positive, the simulation analysis indicated the wide range of possible outcomes while the distributional analysis indicated that the net benefits could become negative with distributional weighting.

12 Gaps and Directions for Research

This chapter articulated an economic approach to cost-benefit analysis of slot machines, though it should be clear that there is not an off-the-shelf template of conceptual and empirical guidance for the benefit-cost analysis of slots. Some views stated here may subsequently evolve with new research, and it is toward that end that several areas needing analytical improvement are briefly summarized as follows:

1. *Integrating conceptual consumer surplus models*: There are currently several surplus and risk-based consumer models, most obviously, "standard" surplus, addicted gambler's surplus, and distance surplus; and substitute ex-ante measures based on risk-loving behavior or a combination of risk aversion and entertainment. Each provides some insight, but none appears to fully capture the heterogeneity and motivation of observed gamblers or to clearly link the various observable concepts of price to the analysis.
2. *Empirical estimation of surplus concepts, including the role of distance and its implicit price as one element*: It may be possible in locations with state-run casinos and frequent player cards or with cooperating casino owners to develop travel cost models of surplus that are more common in other areas of recreational demand research.
3. *Are slot machines more or less addicting than other forms of gambling?* An important driver of external effects appears to be the number of "abnormal" gamblers. Determining whether and how much a differential impact exists by type of gambling is an important question. For instance, to obtain the partial derivative of the number of non-normal gamblers per change in amount wagered on different games could be informative or, more complexly, the total derivative of the number of non-normal gamblers in which game type is one factor. Newly available datasets, such as those for loyalty cards, may allow such analyses.
4. *Canonical accounting reporting formats*: Much discussion occurs in the literature about transfers as distinct from losses in economic efficiency. As noted in the discussion of market imperfections, some market failures have aspects of both. Developing the alternative accounting templates, one showing transfers and the other omitting them, could be a useful way of illustrating the importance of different impacts. In many cases the magnitude of transfers may exceed economic welfare losses and so has little effect on economic efficiency but possibly a large effect on the political economy of gambling debates.
5. *Shadow prices and transfers*. The empirical challenge in almost all application areas is to estimate credible shadow prices for the efficiency cost or benefit of what are typically nonmarket actions. Defining efficiency-based shadow prices generally implies a corollary definition of what transfers are. None of the impacts directly central to gambling are listed in a standard text as shadow prices, which tend to focus on injuries, recreation, and the environment. The gambling

literature appears to be reducing its variability of approach in some areas, such as bankruptcy, but disparate framings and estimates continue to exist on numerous topics. This likely indicates both a lack of consensus and perhaps coverage in the economics literature. Improvement is likely to depend on the usual slow accretion of science and review, but much appears to be at stake.

References

Adler, Matthew, and Eric Posner. 2006. *New foundations of cost-benefit analysis.* Cambridge, Mass.: Harvard University Press, Cambridge.

American Gaming Association (AGA). 2010. Facts at your fingertips: U.S. commercial casino industry. Washington, D.C.: AGA; www.americangaming.org/industry-resources/research/facts-your-fingertips.

American Psychiatric Association. 2010. *Diagnostic and Statistical Manual of Mental Disorders.* 5th ed. (DSM-5). Arlington, Va.: American Psychiatric Association; www.dsm5.org.

Anielski, Mark, and Aaron Braaten. 2008. *The socio-economic impact of gambling (SEIG) framework.* Report prepared by Anielski Management, Inc., for the Inter-Provincial Consortium for the Development of Methodology to Assess the Social and Economic Impact of Gambling; www.anielski.com/Documents/SEIG%20Framework.pdf.

Arrow, Kenneth J., and Robert C. Lind. 1970. Uncertainty and the evaluation of public investment decisions. *American Economic Review* 60(3):364–378.

Australian Productivity Commission (APC). 1999. *Australia's gambling industries: Inquiry report.* Commonwealth of Australia; www.pc.gov.au/projects/inquiry/gambling/docs/finalreport.

——. 2010. *Gambling: Inquiry Report.* Commonwealth of Australia; www.pc.gov.au/projects/inquiry/gambling-2009/report.

Baker, R.G.V. and D.C. Marshall. 2005. Modelling gambling time and economic assignments to weekly trip behavior to gambling venues. *Journal of Geographical Systems.* 7:381–402.

Baumol, William J., and Charles A. Wilson, eds. 2001. *Welfare Economics.* 3 vols. Cheltenham, U.K.: Edward Elgar.

Becker, Gary S., and Kevin M. Murphy. 1988. A theory of rational addiction. *Journal of Political Economy* 96(4):675–700.

Becoña, Elisardo, Francisco Labrador, Enrique Echeburúa, Elena Ochoa, and Miguel Angel Vallejo. 1995. Slot machine gambling in Spain: An important and new social problem. *Journal of Gambling Studies* 11(3):265–286.

Bellinger, William. 2007. *The economic analysis of public policy.* New York: Routledge.

Belzer, Richard. 1999. *CSAB Project on Regulatory Oversight: Study Protocol 1, Appendix C; Evaluative criteria for compliance [Executive Order 12866].* Center for the Study of American Business, Washington University in St. Louis; http://wc.wustl.edu/files/wc/imce/csab_project_on_regulatory_oversight.pdf.

Boardman, Anthony, David Greenberg, Aiden Vining, and David Weimer. 2011. *Cost-benefit analysis: Concepts and practice.* Upper Saddle River, N.J.: Pearson-Prentice Hall.

Breen, Robert B., 2004. Rapid onset of pathological gambling in machine gamblers: A replication, *International Journal of Mental Health and Addiction* 2(2):44–49.

Brent, Robert. 2006. *Applied cost-benefit analysis.* Cheltenham, U.K.: Edward Elgar.

Camerer, Colin, George Loewenstein, and Matthew Rabin. 2004. *Advances in behavioral economics*. Princeton, N.J.: Princeton University Press.

Chetty, Raj. 2009. Sufficient statistics for welfare analysis: A bridge between structural and reduced-form methods. *Annual Review of Economics* 1(1):451–488.

Chóliz, Mariano. 2010. Experimental analysis of the game in pathological gamblers: Effect of the immediacy of the reward in slot machines. *Journal of Gambling Studies* 26(2):249–256.

Clotfelter, Charles, and Philip Cook. 1990. On the economics of state lotteries. *Journal of Economic Perspectives* 4(4):105–119.

Collins, David, and Helen Lapsley. 2003. The social costs and benefits of gambling: An introduction to the economic issues. *Journal of Gambling Studies* 19(2):123–148.

Conlisk, John. 1993. The utility of gambling. *Journal of Risk and Uncertainty* 6(3):255–275.

Coryn, Tom. 2008. Casinos: Lessons learnt from cost-benefit analysis. In *Economic aspects of gambling regulation*, edited by Tom Coryn, Cryille Fijnaut, and Alan Littler. Leiden: Martinus Nijhoff, 31–55.

Crane, Yuliya. 2008. What are the costs and benefits of gambling in the United Kingdom? In *Economic aspects of gambling regulation*, edited by Tom Coryn, Cryille Fijnaut, and Alan Littler. Leiden: Martinus Nijhoff.

Dowling, Nicki, David Smith, and Trang Thomas. 2005. Electronic gambling machines: Are they the "crack cocaine" of gambling? *Addiction* 100(1):33–45.

Eadington, William R. 1999. The economics of casino gambling. *Journal of Economic Perspectives* 13(3):173–192.

———. 2003. Measuring costs from permitted gaming: Concepts and categories in evaluating gambling's consequences. *Journal of Gambling Studies* 19(2):185–213.

———. 2008. Gambling policy in the European Union: Monopolies, market Access, economic rents, and competitive pressures among gaming sectors in the member states. In *Economic aspects of gambling regulation*, edited by Tom Coryn, Cryille Fijnaut, and Alan Littler. Leiden: Martinus Nijhoff, 71–90.

Eeckhoudt, Louis, Christian Gollier, and Harris Schlesinger. 2005. *Economic and financial decisions under risk*. Princeton, N.J.: Princeton University Press.

European Commission. 2006. *Study of gambling services in the internal market of the European Union*. Report prepared by the Swiss Institute of Comparative Law, June 14; http://ec.europa.eu/internal_market/services/docs/gambling/study1_en.pdf.

European Union–Regional Policy. 2008. *Guide to cost-benefit analysis of investment projects: Structural funds, cohesion fund and instrument for pre-accession*, July; http://ec.europa.eu/regional_policy/sources/docgener/guides/cost/guide2008_en.pdf.

Farrow, Scott. 2011. Incorporating equity in regulatory and benefit-cost analysis using risk based preferences. *Risk Analysis* 31(6):902–907.

Farrow, Scott, and Judith Shinogle. 2010. Are there net benefits to slot machine gambling in Maryland? UMBC Working Paper 10–128; www.umbc.edu/economics/working_papers.html.

Fey, Marshall. 2006. *Slot machines: America's favorite gaming device*. Reno, Nev.: Liberty Belle Books.

Fisher, Sue, and Mark Griffiths. 1995. Current trends in slot machine gambling: Research and policy. *Journal of Gambling Studies* 11(3):239–247.

Flores, Nicholas. 2002. Non-paternalistic altruism and welfare economics. *Journal of Public Economics* 83(2):293–305.

Foster, James E., and Amartya Sen. 1997. *On Economic Inequality* after a quarter century. In Amartya Sen, *On economic inequality*, expanded ed. Oxford: Clarendon.

Freeman, Myrick, III. 2003. *The measurement of environmental and resource values*, 2nd ed. Washington, D.C.: Resources for the Future.

Friedman, Milton, and Leonard Savage. 1948. The utility analysis of choices involving risk. *Journal of Political Economy* 56(4):279–304.

Gandhi, Amit. 2008. *Rational expectations at the racetrack: Testing expected utility using prediction market prices.* University of Chicago working paper.

Goulder, Lawrence H., and Roberton C. Williams III. 2003. The substantial bias from ignoring general equilibrium effects in estimating excess burden, and a practical solution. *Journal of Political Economy* 111(4):898–927.

Grinols, Earl L. 1999. Distance effects in consumption: Measuring distance value with application to casino siting. *Review of Regional Studies* 29(1):63–76.

——. 2004. *Gambling in America: Costs and benefits.* Cambridge: Cambridge University Press.

——. 2007. Social and economic impacts of gambling. In *Research and measurement issues in gambling studies*, edited by Garry Smith, David Hodgins, and Robert Williams. Amsterdam: Elsevier; 515–541.

Gruber, Jonathan, and Botond Köszegi. 2001. Is addiction rational? Theory and evidence. *Quarterly Journal of Economics* 116(4):1261–1303.

Guryan, Jonathan, and Melissa S. Kearney. 2010. Is lottery gambling addictive? *American Economic Journal: Economic Policy* 2(3):90–110.

Hahn, Robert W., and Patrick Dudley. 2007. How well does the U.S. government do cost–benefit analysis? *Review of Environmental Economics and Policy* 1(2):192–221.

Harrigan, Kevin A. and Mike Dixon. 2009. PAR Sheets, probabilities, and slot machine play: Implications for problem and non-problem gambling. *Journal of Gambling Issues* 23:81–110.

Haveman, Robert and Scott Farrow. 2011. Labor expenditures and benefit-cost accounting in times of unemployment. *Journal of Benefit-Cost Analysis* 2:2.

Hazilla, Michael, and Ray Kopp 1990. Social cost of environmental quality regulations: A general equilibrium analysis. *Journal of Political Economy* 98(4):853–873.

HM Treasury. 2013. *The green book: Appraisal and evaluation in central government*; www.hm-treasury.gov.uk/data_greenbook_index.htm.

Humphreys, Brad, Rodney Paul, and Andrew Weinbach. 2010. *Consumption benefits and gambling: Evidence from the NCAA basketball betting market.* Working Paper 2010–7, University of Alberta, Department of Economics.

Humphreys, Brad R., Brian P. Soebbing, Harold Wynne, John Turvey, and Yang Lee. 2011. *Final report to the Alberta Gaming Research Institute on the socio-economic impact of gambling in Alberta*; dspace.ucalgary.ca/bitstream/1880/48545/1/Humphreys_et_al_final_May_25_2011.pdf.

Jones, Chris. 2005. *Applied welfare economics.* Oxford: Oxford University Press.

Just, Richard, Darrell Hueth, and Andrew Schmitz. 2004. *The welfare economics of public policy.* Cheltenham, U.K.: Edward Elgar.

Krutilla, Kerry. 2005. Using the Kaldor-Hicks tableau format for cost-benefit analysis and policy evaluation. *Journal of Policy Analysis and Management* 24(4):864–875.

Lave, Lester B. 1996. Benefit-cost analysis: Do the benefits exceed the costs? In *Risks, costs, and lives saved: Getting better results from regulation*, edited by Robert W. Hahn. Cambridge: Cambridge University Press, 104–134.

Machina, Mark. 1987. Choice under uncertainty: Problems solved and unsolved. *Journal of Economic Perspectives* 1(1):121–154.

Markowitz, Harry. 1952. The utility of wealth. *Journal of Political Economy* 60(2): 151–158.

Marshall, Katherine, and Harold Wynne. 2003. Fighting the odds. *Perspectives on Labor and Income* 4(12):5–13; www.statcan.gc.ca/pub/75-001-x/75-001-x2003112-eng.pdf.

McConnell, Kenneth. 1997. Does altruism undermine existence value? *Journal of Environmental Economics and Management* 32(1):22–37.

Miller, Ronald, and Peter Blair. 2009. *Input-output analysis: Foundations and extensions*. 2nd ed. New York: Cambridge University Press.

Narayanan, Sridhar, and Puneet Manchanda. 2011. An empirical analysis of individual level casino gambling behavior. Research paper series No. 2003 (R1), Graduate School of Business, Stanford University.

National Gambling Impact Study Commission. 1999. *National Gambling Impact Study Commission Final Report*; http://govinfo.library.unt.edu/ngisc/index.html (then select final report/research reports).

National Research Council (NRC). 1999. *Pathological Gambling: A Critical Review*. Washington, D.C.: National Academy Press.

Nevada Commission on Economic Development. 1999. *Does economic development pay for itself in Nevada?* Paper presented at the 1999 Annual Governors Conference on Economic Development, Las Vegas, Nevada, August 25.

Oxford Dictionary of Economics. 2009. Oxford University Press, Oxford. Available at http://www.oxfordreference.com/view/10.1093/acref/9780199237043.001.0001/acref-9780199237043-e-2415?rskey=nRfBPx&result=2280&q=)

Paton, David, Donald Siegel, and Leighton Vaughn Williams. 2004. Taxation and the demand for gambling: New evidence from the United Kingdom. *National Tax Journal* 57(4): 847–860.

Pinge, Ian, 2008. *Electronic Gaming Machines in Bendigo 2008: Assessing their economic impact.* Jimbrada Pty. Ltd. Available at http://www.pc.gov.au/__data/assets/pdf_file/0004/87277/sub099-attachment3.pdf.

PolicyAnalytics. 2006. *A benefit-cost analysis of Indiana's riverboat casinos for FY 2005: A report to the Indiana Legislative Council and the Indiana Gaming Commission*; www.in.gov/legislative/igareports/agency/reports/IGC02.pdf.

Regional Economics Applications Laboratory. 2003. *Economic impact of the casino gaming industry on the State of Illinois and local communities.* Urbana: University of Illinois; www.illinoiscasinogaming.org/press_files/clips/hewings_study.doc.

Roman, John, and Graham Farrell. 2002. Cost-benefit analysis for crime prevention: Opportunity costs, routine savings, and crime externalities. In *Evaluation for crime prevention.* Crime Prevention Studies, vol. 14, edited by Nick Tilley. New York: Criminal Justice, 53–92.

Rose, Adam, and Assocs. 1998. The regional impact of casino gambling: Assessment of the literature and establishment of a research agenda. Report prepared for the U.S. National Gambling Impact Study Commission; http://govinfo.library.unt.edu/ngisc/reports/ecoimprpt.pdf.

Ross, Gary. 1987. *Stung: The incredible obsession of Brian Molony*. Toronto: Stoddart Publishing Co.

Sauer, Raymond D. 1998. The economics of wagering markets. *Journal of Economic Literature* XXXVI (December):2021–2064.

Schmitz, Andrew, and Richard O. Zerbe, eds. 2009. *Applied benefit-cost analysis*. Cheltenham, U.K.: Edward Elgar.

Shaffer, Howard J., Michael V. Stanton, and Sarah E. Nelson. 2006. Trends in gambling publications. *Journal of Gambling Studies* 22(4):427–442.

Shinogle, Judith, Robert Carpenter, Scott Farrow, and Donald N. Norris. 2008. *An analysis of the impact of introducing video lottery terminals in Maryland*. Baltimore: Maryland Institute for Policy Analysis and Research; www.umbc.edu/mipar/research.php#research.

Smith, Garry, and Harold Wynne. 2004. *VLT gambling in Alberta: A preliminary analysis*. Edmonton: Alberta Gambling Research Institute, University of Alberta.

Smith, Garry, David Hodgins, and Robert Williams, eds., 2007. *Research and measurement issues in gambling studies*. Amsterdam: Elsevier.

Starmer, Chris. 2000. Developments in non-expected utility theory. *Journal of Economic Literature* 38:332–382.

Statistics Canada. 2010. Gambling, 2010. *Perspectives*. Stastistics Canada Catalogue no. 75–001–X

Stewart, David. 2010. *Demystifying slot machines and their impact in the United States*. American Gaming Association white paper. Washington, D.C.: AGA.

Thaler, Richard H. 1992. *The winner's curse: Paradoxes and anomalies of economic life*. Princeton, N.J.: Princeton University Press.

Thalheimer, Richard. 2008. Government Restrictions and the Demand for Casino and Pari-mutuel Wagering. *Applied Economics* 40(6):773–791.

Thalheimer, Richard and Mukhtar M. Ali. 2003. The Demand for Casino Gambling. *Applied Economics* 35:907–918.

Thompson, William, Ricardo Gazel, and Dan Rickman. 1995. The economic impact of Native American gaming in Wisconsin. Wisconsin Policy Institute Report 8/3; www.wpri.org/Reports/Volume8/Vol8no3.pdf.

——. 1997. The social and legal costs of compulsive gambling. *Gaming Law Review* 1(1): 81–89.

Treasury Board of Canada. 2007. *Canadian cost benefit analysis guide: Regulatory proposals*; www.tbs-sct.gc.ca/ri-qr/documents/gl-ld/analys/analys-eng.pdf.

Treyz, Frederick, and George Treyz. undated. *Evaluating the Regional Economic Effects of Structural Funds Programs Using the REMI Policy Insight Model*; ec.europa.eu/regional_policy/sources/docconf/budapeval/.../treyz.doc

Turner, Nigel, and Roger Horbay. 2004. How do slot machines and other electronic gambling machines actually work? *Journal of Gambling Issues* 11(July); jgi.camh.net/toc/jgi/11?seq=11.

Tversky, Amos, and Daniel Kahneman. 1971. Belief in the law of small numbers. *Psychological Bulletin* 76(2):105–110.

U.S. Office of Management and Budget (OMB). 1992. *Guidelines and discount rates for benefit-cost analysis of federal programs*. OMB Circular A-94. Washington, D.C.: OMB; www.whitehouse.gov/omb/circulars_a094.

——. 2003. *Regulatory analysis*. OMB Circular A-4. Washington, D.C.: OMB; www.whitehouse.gov/omb/circulars_a004_a-4.

——. 2010. Agency checklist: Regulatory impact analysis. Washington, D.C.: OBM; www.whitehouse.gov/sites/default/files/omb/inforeg/regpol/RIA_Checklist.pdf.

Vining, Aidan, and David L. Weimer. 2010. An assessment of important issues concerning the application of benefit-cost analysis to social policy. *Journal of Benefit-Cost Analysis*: 1(1):6.

Volberg, Rachel A. 2001. *When the chips are down: Problem gambling in America*. New York: Century Foundation Press.

——. 2003. *Gambling and problem gambling in Arizona: Report to the Arizona Lottery*. Northampton, Mass.: Gemini Research.

Walker, Douglas M. 2006. Review of PolicyAnalytics' "A benefit-cost analysis of Indiana's riverboat casinos for FY 2005." Prepared for the Casino Association of Indiana; walkerd.people.cofc.edu/pubs/IndianaReview(3–24F).pdf.

——. 2007a. *The economics of casino gambling*. Berlin: Springer.

——. 2007b. Problems in quantifying the social costs and benefits of gambling. *American Journal of Economics and Sociology* 66(3):609–645.

——. 2008. *Issues to consider in implementing the "Socio-economic Impact of Gambling (SEIG) Framework."* Prepared for the Canadian Gaming Association; www.walkerd.people.cofc.edu/pubs/CGA_SEIG_8-27.pdf.

Walker, Douglas M., and A. H. Barnett. 1999. The social cost of gambling: An economic perspective. *Journal of Gambling Studies* 15(3):181–212.

Weimer, David L., Aidan R. Vining, and Randall K. Thomas. 2009. Cost-benefit analysis involving addictive goods: Contingent valuation to estimate willingness-to-pay for smoking cessation. *Health Economics* 18(2):181–202.

Weitzman, Martin. 1965. Utility analysis and group behavior: An empirical study. *Journal of Political Economy* 73(1):18–26.

Welte, John W., Grace M. Barnes, William F. Wieczorek, Marie-Cecile Tidwell, and John Parker. 2002. Gambling participation in the U.S.—Results from a national survey. *Journal of Gambling Studies*, 18(4):313–337.

Williams, Robert, and Robert Wood. 2004. *The demographic sources of Ontario gaming revenue: Final report*. Prepared for the Ontario Problem Gambling Research Centre; www.uleth.ca/dspace/bitstream/handle/10133/380/?sequence=1.

Wynne, Harold J., and Howard J. Shaffer. 2003. The socioeconomic impact of gambling: The Whistler symposium. *Journal of Gambling Studies* 19(2):111–121.

Zerbe, Richard O. 2002. Can law and economics stand the purchase of moral satisfaction? *Research in Law and Economics* 20:135–172.

Zerbe, Richard, and Allen Bellas. 2006. *A primer for benefit-cost analysis*. Cheltenham, U.K.: Edward Elgar.

Zerbe, Richard, and Dwight Dively. 1994. *Benefit-cost analysis in theory and practice*. New York: Harper Collins.

CHAPTER 33

THE ECONOMICS OF LOTTERIES: A SURVEY OF THE LITERATURE

KENT GROTE AND VICTOR A. MATHESON

Introduction

Lotteries represent one of the oldest and most common forms of gambling around the world, with origins dating back at least to ancient Rome and possibly even earlier, to the Han Dynasty of China in the second century B.C. A lottery involves the sale by an organizing body, typically the government but also occasionally private businesses or charities, of a ticket giving the possessor a potential monetary reward. Lotteries differ from casinos in that lottery ticket sales generally do not take place at a location specifically set aside for gambling, and modern lotteries are usually operated by governments instead of by private firms.

Lotteries are of particular interest to scholars for a variety of reasons. First, they represent an important source of government revenues in many states and countries. Thus, they are of interest to public finance economists. Second, lotteries provide researchers interested in microeconomic theory and consumer behavior with a type of experimental lab that allows economists to explore these topics.

This chapter surveys the existing literature on lotteries organized around these two central themes. The first section examines the microeconomic aspects of lotteries, including consumer decision-making under uncertainty, price and income elasticities of demand for lottery tickets, cross-price elasticities of lottery tickets to each other and to other gambling products, consumer rationality and gambling, and the efficiency of lottery markets. The second section covers topics related to public finance and public choice, including the revenue potential of lotteries; the tax efficiency and deadweight loss of lottery games; the horizontal and vertical equity of lotteries, including potential externalities associated with gambling; earmarking and the fungibility of lottery revenues; and individual state decisions to participate in public lotteries.

The current literature on the economics of lotteries is so extensive that it is impractical to cover every paper on the topic. Thus, this chapter focuses on the most influential papers in the field. A more extensive bibliography of lottery related papers is available from the authors upon request or through Research Papers in Economics (RePEc at www.repec.org).

Microeconomics: Demand for Lottery

Much of the literature on lottery markets focuses on the demand for lottery products, be they in the form of estimating demand equations, determining the regressive nature of ticket purchases, or discussing the concepts of consumer rationality and market efficiency. Indeed, why people demand lottery tickets in the first place is a real question. Milton Friedman and Leonard Savage (1948) (and subsequently Harry Markowitz 1952) suggested that the curvature of individuals' utility functions changes as they get richer (or move away from their "normal" income), thereby providing a theory for why individuals exhibit risky behavior through their participation in lottery markets at the same time that they exhibit risk-averse behavior elsewhere. These theories provide motivation for the idea that lottery purchases can be considered rational behavior; if so, consumers of lottery products should have typical demand functions that include some familiar microeconomic variables, including price, income, consumer preferences, number of consumers, price of related products, and product characteristics.

Effective Price

The price of lottery tickets has received much attention in the literature, which may at first seem surprising since the actual price paid for tickets tends to remain constant unless the lottery authority decides to change it. The "effective price" of a lottery ticket, which considers the price as well as the return, however, may change over time and across lottery jurisdictions. Evidence on the effect of the effective price on ticket sales is mixed, with early studies (Vrooman 1976; Vasche 1985; Mikesell 1987) concluding that the effective price of tickets does not have a significant impact on the sales of tickets and later studies (DeBoer 1986; Clotfelter and Cook 1989; Miller and Morey 2003) finding a significant and negative relationship between the takeout rate and lottery sales.

While earlier studies used the takeout rate (or "vigorish") to calculate the effective price, subsequent studies have tended to use the difference between the nominal ticket price and the expected return as the measure of effective price. For many lottery games there is no difference between the expected return and the net difference between the ticket price and the takeout; however, lotto is a common lottery game that is distinguished by the characteristic that if there is no jackpot winner in a given drawing period, the prize pool rolls over into the next drawing, increasing the potential jackpot in the next period. Higher jackpots typically lead to higher expected values for ticket purchases, which in turn lead to lower effective prices even if the actual dollar price of a ticket remains constant. Because changes in game structures, ticket prices, and takeout rates are rare, many of the studies of lottery demand examine lotto games, taking advantage of the constant changes in effective price by including either jackpot or expected return as an explanatory variable.

Many studies of demand also estimate price elasticities in order to determine whether the existing lottery structures maximize the potential gaming revenues. Both Ian Walker (1998) and David Forrest, O. David Gulley, and Robert Simmons (2000a) concluded that the U.K. National Lottery has an optimal takeout rate of 50 percent based on an estimated price elasticity of demand that is close to −1. Many other empirical studies also estimated that price elasticities are approximately equal to −1; however, there are studies that found relatively more elastic demand (Farrel and Walker 1999 and Farrell, Morgenroth, and Walker 1999 for the U.K. lottery, particularly in the long run; Papachristou and Karamais 1998 for the Greek Lotto; and Gulley and Scott 1993 for the Mass Millions game), implying that a lower takeout rate would increase revenues. Other studies have suggested relatively less elastic demand, as low as −0.66 short-run price elasticity for the U.K. National Lottery as found by Forrest, Gulley, and Simmons (2000a), −0.382 in the Taiwan lotto game as measured by Chuan Lee, Chin-Tsai Lin, and Chien-Hua Lai (2010), and −0.19 for the Mass Millions game as measured by Gulley and Frank Scott (1993), implying higher takeout rates would increase government revenues.

Forrest, Simmons, and Neil Chesters (2002) argued that lottery demand depends more on jackpot size than expected value because players tend to participate in games with very low odds of winning in order to "dream big" about substantial winnings. Thus such studies as Larry DeBoer (1990) and Philip Cook and Charles Clotfelter (1993) included jackpot size and jackpot size squared to test for a nonlinear and positive relationship between jackpot size and ticket sales. Jackpot rollovers are such a distinctive part of the literature in terms of measuring the effect of "price changes" that Forrest, Simmons, and Chesters (2002), George Papachristou (2006), and George Geronikolaou and Papachristou (2007) calculated a "jackpot elasticity of sales (demand)" that has a similar interpretation as price elasticity of demand, except with a positive expected relationship with ticket sales.

Income Elasticity

Like the price of a product, income is another significant factor in the demand for any good and is particularly important for empirical studies on lotteries in order to determine whether the "lottery tax" in a particular jurisdiction is regressive. Studies typically use income level, per capita income, disposable income, or real income in order to estimate this effect, though some studies also use variables like the poverty rate in order to capture the regressive nature of lottery spending (Blalock, Just, and Simon, 2007). While the measurements of income elasticity vary from study to study, empirical research uniformly finds income elasticities less than one, indicating that a relatively greater percentage of income is spent on lottery products at lower income levels (Suits 1977; Clotfelter 1979; Clotfelter and Cook 1987, 1989). Instant games tend to have lower income elasticities than other games (Mikesell 1989; Jackson 1994; Garrett and Coughlin 2009) while lotto games with large jackpots tend to appeal to more affluent customers. Indeed, Emily Oster's (2004) study of Powerball sales in Connecticut predicted that at exceedingly high jackpot levels the Powerball game could actually become progressive, the only such finding in the literature.

Other indirect measures of income also tend to suggest that lotteries are a regressive form of taxation. Studies by John Laitner (1999), Allan Layton and Andrew Worthington (1999), and Cletus Coughlin and Thomas Garrett (2009) all found that individuals in government income assistance programs are more likely to participate in lottery markets. The observed effect of unemployment on ticket sales is mixed, with John Mikesell (1994) and Frank Scott and John Garen (1994) finding that unemployment rates tend to have a positive impact on lottery ticket sales, while Garrick Blalock, David Just, and Daniel Simon (2007) found a negative relationship and DeBoer (1990) found no correlation.

Demographics

Demographics also influence ticket sales, and empirical studies are in wide agreement as to their significant influences on ticket sales. The old adage that the lottery is a "tax on people who are bad at math" is borne out in the data. Level of education typically has a negative relationship with ticket sales (Clotfelter and Cook 1987, 1989; Kitchen and Powells 1991; Farrell and Walker 1999). With respect to race and gender, studies tend to find that black and Hispanic individuals are more likely than whites to buy lottery tickets (Jackson 1994; Scott and Garen 1994), and men are more likely to play than women are (Clotfelter and Cook 1987, 1990; Kitchen and Powells 1991; Farrell and Walker 1999), though the effect can vary by location, time period, and type of game. Studies also have found that people who live in urban areas and, therefore, are closer to more lottery vendors tend to buy more lottery tickets than do people in rural areas (Hersch and McDougall 1989; Clotfelter and Cook 1989, 1993; Kitchen and Powells 1991). Studies of other demographic variables, such as age and marital status, do not exhibit consistent effects on lottery ticket sales (Clotfelter and Cook 1989, 1990; Kitchen and Powells 1991; Jackson 1994; Farrell and Walker 1999).

Other Products: Substitutes and Complements

Lottery authorities typically offer multiple games, and lotteries may coexist with other types of gambling, so a final issue relating to the demand for lottery tickets is the extent to which other products are complements or substitutes for lotteries. The literature provides mixed empirical results on this issue. Cook and Clotfelter (1993), Gulley and Scott (1993), and Forrest, Gulley and, Simmons (2004) concluded that lotto rollovers do not impact sales of other lottery products in the lotteries they studied. Paul Mason, Jeffrey Steagall and Michael Fabritius (1997) found that the two Florida lotto games are substitutes for one another, while Forrest, Gulley, and

Simmons (2004) found some substitution effects between scratchcard purchases and the U.K. lotto. Conversely, in Ireland, Catriona Purfield and Patrick Waldron (1999) and, across the United States, Kent Grote and Victor Matheson (2006a) found that different lottery games serve as complements to one another. While it is more natural to suppose that lottery products are substitutes for one another, Grote and Matheson concluded that transactions costs and the ability to buy multiple types of game tickets at the same time are responsible for the complementarities exhibited by lottery ticket buyers.

A topic related to the concept of substitution is what happens to overall spending on lottery games in a lottery jurisdiction when new games are introduced. Presumably the purpose of introducing new lottery games should be to increase overall lottery spending, but if new lottery games merely attract ticket sales from already existing games, an effect often referred to as cannibalization, then the lottery authority has not benefited from introducing a new game to the lottery mix. Mikesell and C. Kurt Zorn (1987), Grote and Matheson (2006a), and Matheson and Grote (2007) all found that the introduction of new lotto games does have a negative impact on sales for existing lottery products, but the addition of new games increases overall lottery ticket sales. Matheson and Grote (2007) go on to note that the overall increase in ticket sales is larger if the new game is sufficiently different in odds or prize structure from the existing games. Finally, it is well documented that the introduction of lotteries in neighboring states serves to reduce lottery spending within a state, as people will cross state boundaries to buy lottery tickets (Suits 1979; Mikesell and Zorn 1987; Walker and Jackson 1999; Garrett and Marsh 2002).

Lottery ticket sales can also affect or be affected by the availability of other gambling activities in a jurisdiction. Some studies (Scott and Garen 1994; Calcagno, Walker, and Jackson 2010) have found that the presence of a lottery increases participation in other gambling activities, such as casino gaming and dog and horse racing, presumably reflecting a general attitude or preference toward gambling in a society. Most of the literature on gambling activities and their relationship to lotteries, however, find that either they are unrelated to each other or that they are substitutes for one another. Douglas Walker and John Jackson (2008) and Forrest, Gulley, and Simmons (2010) found that sales revenues in racing and lotteries are not strongly related. Donald Steinnes (1998) found that casino gambling does not have a significant impact on lottery sales, whereas Melissa Kearney (2005b) found that lottery spending does not significantly reduce spending on other forms of gambling. The remaining studies tended to find that lotteries and other forms of gambling are substitutes for one another (Gulley and Scott 1989; Siegel and Anders 2001; Elliott and Navin 2002).

Microeconomics: Lottery Structure and Demand

In addition to the other factors affecting demand for lottery tickets described in the previous section, other characteristics of lottery games, including the odds of winning, the prize structure, and the payout rate of the game, affect the demand for tickets as well. If consumers have preferences for certain lottery game characteristics, then it is logical to assume that states can and should structure their lottery games to attract the most consumers in order to maximize lottery revenues. This is often referred to in the literature as achieving an "optimal structure" for a lottery. A lottery association must determine a payout rate, the odds of winning, and the distribution of payouts among different size prizes for each game it offers.

While lotto games also offer smaller consolation prizes to ticket buyers who fail to win the largest prize, the jackpot prize is arguably the primary attraction of the game (Forrest, Simmons, and Chesters 2002), so it seems logical that states should structure the odds of winning the jackpot in order to attract the most consumers within a lottery district. Both DeBoer (1990) and Stuart Thiel (1991) concluded that the New York state lottery and Washington state lotteries, respectively, should provide worse odds of winning in order to attract more players to their lotto games. Longer odds would result in more rollovers leading to higher jackpots, and if consumers care more about the size of the lottery prize than about the odds of winning, such a strategy will result in higher sales.

The ability of lottery associations to generate additional demand by lengthening the odds of winning the jackpot is not unlimited, however, as eventually the odds of winning the grand prize become so low that the jackpot is won too rarely, causing players to lose interest (Forrest and Alagic 2007). Thus the optimal jackpot odds depend on how many potential consumers there are in the jurisdiction offering the tickets. Cook and Clotfelter (1993) referred to this as the "scale economies of lotto" and found that states often select their game formats so that the probability of winning the jackpot multiplied times the population within the state is approximately equal to one. Lottery associations have taken this finding to heart, and the past 20 years have witnessed a rise in multi-state or multi-country lotto games offering huge jackpots at increasingly long odds.

Cook and Clotfelter's findings are dependent on the particular risk preference of lottery consumers in the United States. The question of the optimal odds of winning has also been studied in the United Kingdom, Greece, and Spain. Walker and Judith Young (2001) used simulations of sales for the U.K. lotto to demonstrate that reducing the odds of winning may, in fact, reduce sales because reducing the odds of winning reduces the expected return on a lottery ticket while increasing the variance and skewness of expected return, which may not be favorable to the risk preferences of consumers in the United Kingdom. Papachristou (2009) demonstrated mathematically that scale economies likewise affect the mean, variance, and skewness of expected returns. Forrest, Levi Pérez, and Rose Baker (2010) found that sales for the Spanish National Lotto game increase when the odds of winning decline and additional lower tier prizes are added to the game structure.

Aside from the odds of winning the jackpot, a second important characteristic of lottery games is the prize structure offered. While the jackpot prize may be of primary interest to purchasers of lottery tickets, it is not the only characteristic, and the prizes offered as well as the percentage of sales used to fund the prizes offered (the payout rate) may have an impact on consumer preferences and consumer demand. John Quiggin (1991) has provided a mathematical model of lottery demand which shows that, though smaller prizes do not have much impact on the overall expected value of a ticket, they do reduce the expected losses so that consumers may prefer lottery games with multiple prizes and prize levels. The model also suggests that product differentiation of lottery tickets is particularly important when consumers of lottery products have very different risk preferences.

A number of empirical studies have been offered in the literature to test Quiggin's propositions. John Scoggins (1995) found that Florida lottery officials should increase the percentage of sales allocated to the jackpot prize from 25 to 30 percent in order to increase sales. Garrett and Russell Sobel (1999) found that lottery players in 216 U.S. games in 1995 appeared to be risk averse and to favor skewness of returns, recommending that lottery providers can achieve more skewness by offering smaller consolation prizes along with larger jackpots. Conversely, Walker and Young (2001) recommended that the U.K. lotto game reallocate the funding of prize money from the jackpot to smaller prizes in order to stimulate demand. Demand for tickets should increase due to higher overall expected return and lower variance, which should offset the reduced skewness of returns. Garrett and Sobel (2004) performed a statistical study on 135 U.S. lottery games and concluded that ticket sales for these games depend only on the size and odds of winning the jackpot prize, not the expected value of lower tier prizes.

A final characteristic that receives substantial attention in the lottery literature is the optimal takeout rate of lottery games. A higher takeout rate means a larger percentage of the ticket price is kept as revenue but also means lower ticket sales if consumers are responsive to effective price. A lottery association will maximize revenues when the effective price elasticity of demand nears a value of −1. Empirical tests of takeout rates have concluded that many lottery games approximate an optimal takeout rate, and researchers have been quick to recommend changes to takeout rates when effective price elasticities of demand deviate from the revenue-maximizing figure. (Refer to the previous section on *effective price* for more details.)

In one final contribution of note on this topic, Shu-Heng Chen and Bin-Tzong Chie (2008) demonstrated that there is an associated Laffer curve based on the takeout rate (or lottery tax rate) that is flat at the top, concluding that this provides a rationale for the varying takeout rates offered by different lottery games, since there is no single "optimal takeout rate" for all games.

Behavioral Economics: Rationality and Market Efficiency in Lottery Markets

While it is reasonable to question whether gambling by otherwise risk averse individuals can ever be considered rational, both Friedman and Savage (1948) and Markowitz (1952) offered theories about the shape of utility functions that establish a rational demand for lottery products across all income levels. Others have justified gambling as rational by assuming that gambling entails consumption benefits as well as expected winnings or losses (Conlisk 1993). From these bases rational participation in lottery markets can be tested empirically along with tests of efficiency in those markets. There are several different methods of testing for rationality in lottery markets that are related to the numbers that individuals choose to play, the consumer response to lottery rollovers and changes to the form of jackpot payouts, and where ticket buyers purchase their tickets, among others.

Related to the numbers selected, two particular types of irrational behavior are tested in lottery markets: the presence of the gambler's fallacy and the conscious selection of numbers. The gambler's fallacy occurs when players change their beliefs about the probability of a particular combination of numbers being drawn again after those numbers come up as the winning combination even though each drawing in the lottery is an independent random event (Vaughan Williams 2005b). Clotfelter and Cook (1993), Dek Terrell (1994), and Papachristou (2004) all found evidence of the gambler's fallacy in various games, that is, that players tend to not play numbers that have recently won. Jonathan Simon (1999), on the other hand, found evidence of an overselection of recent winning numbers in the U.K. lottery. On a similar note, Jonathan Guryan and Kearney (2005, 2008) found evidence of a "lucky store" effect in Texas. After a store sells a winning ticket, consumers increase their purchases of lottery tickets anywhere from 12 to 38 percent at that store relative to other stores in the community, an effect that cannot be explained with rational behavior.

The conscious selection of numbers in lottery games has received even more attention in the literature. If people have preferences for certain combinations of numbers based on such "lucky numbers" as birthdays, multiples of seven, or patterns on a play slip, then certain combinations of numbers will be selected more frequently while other combinations will be relatively ignored. If prizes are pari-mutuel in nature, however, "lucky numbers" will actually result in lower payouts, since consumers playing common numbers will have to share their winnings among more people, violating rationality.

Conscious number selection has been widely identified in lottery games. Clotfelter and Cook (1989) noted that lottery associations even encourage conscious selection in marketing related to their products perhaps to increase demand by players who wish to "control their destiny." Conscious selection also reduces the coverage of number combinations in any given draw, increasing the probability that a jackpot prize will roll over and potentially attract additional sales in future draws in response to the higher advertised jackpots. Walker (1998), Simon (1999), Farrell et al. (2000), and Ursula Hauser-Rethaller and Ulrich Konig (2002) all found that conscious selection results in more rollovers in lotto games but that the impact on ticket sales tends to be minor.

Other studies concentrate not on whether conscious selection occurs but whether its effect on expected returns is strong enough to allow individuals an opportunity to make money by betting on the unpopular numbers. While Papachristou and Dimitri Karamanis (1998) did not find that conscious number selection in the Greek lotto is large enough so that unpopular numbers ever become a fair bet (i.e., a bet with a positive expected return), other studies of the Canadian Lotto (Ziemba et al. 1986), the

Massachusetts numbers game (Chernoff 1981), 6/49 lotto games in the United States (Thaler and Ziemba 1988), and the U.K. lotto (Baker and McHale 2009) have all found potentially profitable bets among unpopular number combinations. All of the studies caution, however, that it is difficult to make profits over a reasonable timeline by playing these numbers because of the low odds of winning.

Other tests of consumer behavior relate to whether the bettors' responses to lottery rollovers are irrationally high (referred to as "lotto fever" or "lottomania") or irrationally low (referred to as "lotto apathy" or "jackpot fatigue") and how lottery players' behavior changes immediately subsequent to the jackpot prize being won, a phenomenon known as the "halo effect." Lotto fever occurs when a jackpot rollover attracts enough additional purchases of tickets to actually reduce the expected return on a ticket in spite of the higher jackpot being offered due to the increased probability of having to share a jackpot prize should it be won. Michael Beenstock and Yoel Haitovsky (2001) found statistical evidence of lotto fever consistently occurring after the third rollover in the Israel lotto game. Matheson and Grote (2004, 2005), in a cross-sectional study across U.S. lotto games, found that lotto fever is very rare, occurring in less than 0.1 percent of the drawings in their analyses of the phenomenon in more than 17,000 and 23,000 drawings, respectively.

Lotto apathy occurs when tickets sales do not increase despite an increase in jackpot and the expected return. Matheson and Grote (2005) found lotto apathy to be a much more common experience in U.S. lotto games, though it still occurs in less than 2 percent of the more than 10,000 drawings examined and is concentrated in states that simultaneously offer both a high-jackpot, multi-state lotto game and a smaller in-state game with a jackpot that can be easily overshadowed by the larger game. Similar to lotto apathy is the concept of jackpot fatigue, the concept that even with high jackpots lottery participants lose interest in the lottery after the lottery has been around awhile. Most models of lottery demand include a time trend as an independent variable in order to explain lottery ticket sales (Vasche 1985; Mikesell 1987; Mikesell and Zorn 1987; DeBoer 1990). Beenstock and Haitovsky (2001) commented particularly on the fact that the lotto in Israel does not have a positive time trend in spite of that country's growing economy. Stephen Creigh-Tyte and Farrell (2003) also proffered that it is the declining trend in sales for the U.K. lottery which has encouraged the Camelot Group to offer new lottery games and innovations to current games in order to keep the public interested in purchasing lottery tickets.

Several papers have examined the halo effect, the tendency for ticket purchases immediately following the award of a large jackpot to be unexpectedly high despite the jackpot resetting to a lower level. The finding of a halo effect demonstrates a degree of irrationality among lotto players and can be seen as a type of gambling addiction. Farrell, Edgar Morgenroth, and Walker (1999), Grote and Matheson (2007), and Guryan and Kearney (2010) all found degrees of addiction among various lotto games in the United States and the United Kingdom.

A final topic relating to bettor rationality is consumer understanding of lottery payouts. Allen Atkins and Edward Dyl (1995) demonstrated that individuals who win the lottery should choose the annuity, or the prize paid out over many years, rather than a lump sum paid out immediately because of the tax implications of the two payouts. However, they also concluded that most people will choose the lump sum (a conclusion that is validated by the empirical evidence), arguably because a lump sum can make them feel much wealthier, an argument consistent with Friedman and Savage's (1948) rationale for participating in lotteries in the first place. Irrespective of the wisdom of the lump sum versus the annuity, Matheson and Grote (2003) have shown that ticket buyers tend to be rational with respect to changes in the annuity lengths of jackpot prize payouts. They demonstrated that consumers are not fooled into buying more tickets when state lottery associations artificially increase advertised jackpots by increasing the annuity length of the prize payout.

Market Efficiency

If consumers, as a whole, display rationality in lottery markets, then lottery markets should also tend to be efficient. Leighton Vaughan Williams (2005a, 2005b) has discussed the concepts of weak form, semi-strong form, and strong form efficiency in gambling markets in general as well as the empirical literature related to efficiency. Weak form efficiency is stated to exist when there are no betting opportunities in lottery markets that offer positive expected returns (which means that, on net, there are no fair bets), and strong form efficiency exists when wagers have expected values of (1 minus the takeout rate) times the amount of the wager (Thaler and Ziemba 1988). If players are rational there should be few if any opportunities for lotteries to violate weak form efficiency, as the presence of a fair bet should attract more ticket purchases, reducing the expected value of a ticket back down to the ticket price (or lower). Aside from opportunities provided by purchasing rare number combinations discussed previously, studies covering a number of lotteries typically have found that opportunities for fair bets when purchasing a single randomly selected ticket are rare or nonexistent (Thaler and Ziemba 1988; Krautmann and Ciecka 1993; Papachristou and Karamanis 1998; Scott and Gulley 1995; Ciecka, Epstein, and Krautmann 1996; Grote and Matheson 2006b). In the most expansive study, Matheson and Grote (2005) found fair bets in roughly 1 percent of the more than 23,000 drawings in the U.S. state and multistate lotteries they examined, with positive expected values occurring most frequently in minor games in smaller states where relatively high jackpots attract little consumer attention.

Numerous authors also have examined the possibility of purchasing every number combination, a strategy dubbed the "trump ticket." In all cases these studies found that the purchase of the trump ticket is far more likely to provide a fair bet than is the purchase of a single ticket (Thaler and Ziemba 1988; Ciecka, Epstein, and Krautmann 1996; and Grote and Matheson 2006b), a result mathematically proven by Matheson (2001). Matheson and Grote (2005) found that 11 percent of all lottery drawings they examined would represent a fair bet with the purchase of the trump ticket; this is an astonishingly high number. They, like Thaler and Ziemba (1988), concluded, however, that this finding does not necessarily indicate a violation of weak form efficiency, as the transaction costs involved in purchasing all possible combinations are too high to make this strategy feasible, creating an effective barrier to purchasing the trump ticket. Anthony Krautmann and Ciecka (1993) noted that a consortium attempting to purchase every combination of a 1992 drawing in the Virginia lotto was unsuccessful in covering every combination despite enlisting help from numerous lottery retailers.

Tests for strong form efficiency in lottery markets are provided by Scott and Gulley (1995) and Forrest, Gulley, and Simmons (2000b). Both papers report evidence in favor of strong form efficiency and conclude that bettors are able to accurately forecast sales for a lotto drawing. In addition, the ability to forecast improves with the number of draws that have taken place, reaching a reasonable level of accuracy within the first 30 drawings of the game.

Public Finance: Revenues and Efficiency of Lotteries

As a significant contributor to government finances in the United States and the rest of the world, lotteries have been widely examined by public finance economists focusing primarily on their revenue potential and desirability as a method of taxation. Scholars generally acknowledge that even under the most optimistic assumptions, lotteries are unlikely to provide more than a few percentage points of the revenue needed for a modern state or national government (Humphreys and Matheson 2013; Mikesell and Zorn 1986, 1988) . However, it is important to note that they frequently approach or exceed in magnitude tax collections on goods such as alcohol or tobacco. Unique among tax collection agencies, revenue maximization is an explicitly stated goal of lottery organizations, and numerous papers have explored ways in which variations in product variety, lottery structure, and payout rates could be adjusted to increase revenues. All of these topics have been discussed previously in the sections on demand for lottery and lottery structure and will not be repeated here.

The efficiency with which lotteries generate revenues is a topic of some debate. As first noted by Roger Brinner and Clotfelter (1975), lotteries are an unusual form of taxation in that participation is voluntary and the government actually creates the consumption good that is then taxed. The creation of a new consumption good should raise welfare even if dead-weight loss is created when the good is taxed (Rodgers and Stuart 1995; Farrell and Walker 1997). John Livernois (1986) disputed the notion that lotteries should be considered voluntary, since spending on lotteries simply substitutes for spending elsewhere in the economy, and William Rodgers and Charles Stuart (1995) noted that while the creation of an untaxed lottery would raise welfare, the tax levels typically associated with state lotteries reduce welfare in comparison to other methods of taxation.

Aside from the high takeout rates of government lotteries, the large dead-weight loss of lottery taxation is also a result of high administrative costs, especially when payments to vendors (Mikesell and Zorn 1988) and advertising expenses (Heberling 2002) are considered. DeBoer (1985) and Stephen Caudill, Sandra Johnson, and Franklin Mixon

(1995), however, stressed that while administrative costs of lotteries are indeed relatively high, lottery associations generally experience economies of scale in lottery provision, and average administrative costs per ticket can be reduced by pooling resources with other agencies and by expanding sales.

Both Mikesell and Zorn (1988) and Szakmary and Szakmary (1995) have emphasized that an additional problem with lottery revenues is their volatility. Humphreys and Matheson (2013) countered that while lottery and gaming revenues may be subject to significant change from year to year, the variation in gaming revenues is negatively correlated with changes to other common revenue sources so that lottery revenues as part of the system of taxation serve to reduce the overall volatility of government revenues.

Public Finance: Incidence, Equity, and Externalities of Lotteries

One of the strongest criticisms of lotteries as a means of revenue collection is that they constitute a regressive tax. Indeed, on this point there is universal agreement among economists. Much of the literature on the correlation between income and lottery purchases has been reviewed in the previous section on income elasticity, and it is not necessary to revisit it here. Many lottery studies have focused specifically on the distributional incidence of lottery revenues instead of the general factors affecting lottery demand, including Wisman (2006), Kearney (2005a, 2005b), Campbell and Finney (2005), and Combs, Kim, and Spry (2008). Like those on general lottery demand, these studies have uniformly found that lotteries represent a highly regressive form of taxation, though individual products offered by lottery association may vary widely in their regressivity, with instant games generally faring the worst in terms of vertical equity. Elizabeth Freund and Irwin Morris (2005, 2006) found the presence of lotteries associated with higher levels of income inequality economywide.

Others have noted that when lottery profits are earmarked, a proper accounting of where the spending goes is as important as who buys the tickets when assessing the income equity of the lottery system as a whole. Patrick Feehan and Forrest (2007) and Harriet Stranahan and Mary Borg (2004) found that wealthy individuals and regions tend to benefit disproportionally from money earmarked toward cultural programs and education, potentially exacerbating the regressivity of the revenue side of lotteries. Peter Gripaios, Paul Bishop, and Steven Brand (2010) and Noel Campbell and R. Zachary Finney (2005) noted that inequalities in the distribution of lottery proceeds are not limited to income levels but apply also to geography and race, though both studies suggest that inequalities in lotteries on the expenditure side are either nonexistent or at least less severe than inequalities on the revenue side.

The presence of lotteries may also affect other measures of well-being. On the downside, Kearney (2005b) found that the presence of a state lottery reduces expenditures on other consumption goods by up to 2.4 percent, echoing a general concern of Borg, Mason, and Stephen Shapiro (1991). As is often found in studies of casino gaming, Mikesell and Maureen Pirog-Good (1990) found a significant positive correlation between crime rates and the adoption of a lottery. Moreover, Robert Williams and Robert Wood (2007) noted that over one-third of gaming revenues in Ontario are generated by a small number of problem gamblers but that lottery sales are less prone to abuse than are casino gaming or horse racing. On the upside, Mark Skidmore and Mehmet Tosun (2008) found that the introduction of video lottery spurred general retail sales in West Virginia. Eric Lin and Shih-Ying Wu (2010) found that lottery sales are positively correlated with charitable giving, thus allaying fears that the establishment of a "good works" lottery would reduce other types of donations.

Public Finance: Earmarking and Fungibility

In order to encourage consumers to play (Landry and Price 2007) and to overcome opposition to state-sponsored gambling (Erekson et al. 1999; Pierce and Miller 1999; and Ghent and Grant 2007), governments frequently designate profits from lotteries toward specific agencies. Well over half of state lotteries in the United States, and some foreign lotteries, including the U.K. lottery, earmark all or part of the revenues generated for specific government programs, with education being the primary beneficiary (Matheson and Grote 2008).

An important empirical question is whether these earmarked funds actually enhance spending dollar-for-dollar for the designated programs or if governments simply substitute earmarked dollars for dollars that would have come from the state's general funds had earmarking not occurred. The extent to which different sources of state funds can substitute for one another is known as "fungibility." Studies of the fungibility of lotteries have focused on educational spending and nearly uniformly have found that the introduction of a state lottery increases total educational spending by less than the amount of the new earmarked lottery revenue, suggesting at least some degree of fungibility is present when funds are earmarked for specific state and local programs.

Mikesell and Zorn (1986), Borg and Mason (1988, 1990), and Garrett (2001a) all found that education spending in states that adopted earmarked lotteries for education failed to experience increases in educational spending despite the additional lottery funds. Borg, Mason, and Shapiro (1991) found that states with lottery funding earmarked to education have a statistically significantly lower level of spending per student. Taken as a whole, these works suggest that when lotteries provide a dedicated stream of revenue to education, lawmakers are able to divert general fund resources away from schools, leading to an indirect indication of fungibility.

Charles Spindler (1995), Steven Stark, R. Craig Wood, and David Honeyman (1993), Susan Summers et al. (1995), and Vance Land and Majeed Alsikafi (1999) all found that government appropriations to education in a number of different states tend to fall after the introduction of an earmarked lottery. O. Homer Ereksonet al. (2002) and Neva Novarro (2005) conducted cross-sectional, time-series analyses of all 50 states.

Erekson et al. (2002) found that the expenditure on education as a percentage of general revenues falls as lottery revenues per capita rise, indicating that fungibility is a real phenomenon. In addition, for every $1 per capita in lottery revenues generated as funding for a state, there is a loss of approximately 1 to 1.5 percent of education funding available. Novarro (2005) found 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 non-earmarked lottery profits tends to increase education spending by only 43 cents, on average, findings similar to those of William Evans and Ping Zhang (2007). Forrest and Simmons (2003) noted that earmarked funds for sport development in the United Kingdom raised total spending on sport while slightly reducing other local government spending on athletics but that the full degree of fungibility present in earmarked lotteries outside the United States is a largely open question.

Public Choice: State Adoption of Lotteries

Beginning with New Hampshire in 1964, lotteries have spread across the United States and Canada to the point where by 2011 governments offered lotteries in all Canadian provinces and 43 of U.S. states. Numerous studies have examined the factors influencing states to adopt lotteries, though the literature on what causes countries to adopt lotteries is lacking. Theoretical models of lottery adoption are grounded in the public choice models of regulation, first introduced by George Stigler (1971) and Sam Peltzman (1976), in which legislators seek to maximize political support through their legislative decisions. Several contributions provide differing rationales for the prediction that higher incomes should lead to lottery adoption. John Filer, Donald Moak and Barry Uze (1988) argued that concern for the regressivity of lotteries will limit their implementation in poor jurisdictions, whereas Robert Martin and Bruce Yandle (1990) suggested that the relative political power of the wealthy in rich areas will induce the adoption of lotteries as a means of transferring income from the poor to the wealthy. Erekson et al. (1999) argued that rich states will adopt lotteries simply due to the higher potential for lottery revenues.

Empirical studies typically have found strong connections between income, poverty levels, and income changes and the adoption of lotteries. Most papers have found a significant positive relationship between income and lottery adoption (Filer, Moak, and Uze 1988; Hersch and McDougall 1989; Martin and Yandle 1990; Erekson et al. 1999). James Alm, Michael McKee, and Mark Skidmore (1993) focused not on income levels but instead trends in income theorizing that a decline in income adds to the fiscal stress of a state, increasing the likelihood of a state to add a lottery.

Existing tax levels, debt, state government spending, and legal restrictions on the ability of a state to collect other forms of revenue all may affect the decision to offer a lottery, since fiscal stress can create a motivation for states to seek alternative forms of revenue. Numerous papers have proffered that if tax levels are high states are more likely to add a lottery as an additional or alternative form of funding (Filer, Moak, and Uze 1988; Alm, McKee, and Skidmore 1993; Jackson, Saurman, and Shughart 1994; Erekson et al. 1999). One significant deviation from the predicted relationship of variables measuring the fiscal stress in a state has been provided by Martin and Yandle (1990), who found a statistically significant negative relationship between per capita taxes paid and the decision to add a lottery, though they also found a positive relationship between the per capita debt of a state and the addition of a lottery.

"Tax exporting" is another significant factor in a state's decision to add a lottery. Tax exporting occurs when states earn tax revenues from constituents of other states or conversely lose revenues to other states. If a state is losing out on tax revenues because its residents are purchasing lottery tickets from other states that have already adopted lotteries, this may increase the likelihood of that state adding its own lottery. Numerous studies have considered the effect of the lottery status of neighboring states on lottery adoption and found statistically significant evidence of this effect (Davis, Filer, and Moak 1992; Alm, McKee, and Skidmore 1993; Jackson, Saurman, and Shughart 1994; Ghent and Grant 2007). Ronnie Davis, Filer, and Moak (1992) also measured tax exporting in a slightly different manner and found a direct relationship between the number of tourists in a state and a state's decision to add a lottery.

Since adding a lottery is a political decision by legislators and/or voters in a state, it is also necessary to consider organized opposition to lotteries. Many studies have found that the percentage of a state's population identifying themselves as a member of a conservative religious group (often Baptists) has a statistically significant negative effect on the adoption of lotteries (Hersch and McDougall 1989; Martin and Yandle 1990; Jackson, Saurman, and Shughart 1994; Erekson et al. 1999; Pierce and Miller 1999). In part to overcome some of the opposition to lotteries, some states have specifically earmarked lottery revenues to be used for specific (and relatively popular) state programs, often education. If earmarking can ease political opposition, states that earmark their funding for education should be more likely to adopt lotteries relative to states that do not intend to earmark lottery revenues, an effect uncovered by Erekson et al. (1999), Patrick Pierce and Donald Miller (1999) and Linda Ghent and Alan Grant (2007).

A final factor that is considered in many of the empirical models is the presence of other forms of betting in a state, though the hypothesized direction of the effect is not certain. On the one hand, if a state already allows for other forms of gambling, it is reasonable to assume that it may also be more willing to offer state lottery products to its constituents. On the other hand, other organizations offering gambling products represent an obvious interest group that would typically be opposed to the introduction of a competitor. As a case in point, Nevada, home to the largest casino industry in the United States, is also one of the few states that does not offer a state lottery. Empirically, Davis, Filer, and Moak (1992), Ernest Wohlenberg (1992), and Jackson, David Saurman and William Shughart (1994) all found that the presence of other forms of gambling increases the likelihood of a state adopting a lottery, suggesting that outside of Nevada, where casinos and other forms of gambling may be present on a smaller scale, there are less organized interests against competition and the presence of gambling indicates that there would be a demand for additional gambling products such as a state lottery.

Conclusions and Suggestions for Future Research

This chapter has explored the existing literature regarding the economics of lotteries. A substantial amount of research has focused on the demand for lottery products and the lottery's impact on public finances. More of this empirical evidence has been focused on lotteries in the United States, Canada, and the United Kingdom. There have been few studies of lotteries in other countries. Thus, analysis of international lotteries is a fruitful area for additional scholarly work. Even more pressing, while Garrett (2001b) and Matheson and Grote (2009) have provided cross-country studies of lotteries, very little work has been done comparing lottery demand, structures, and adoption among different countries. While the existing literature is extensive, many new frontiers yet exist.

References

Alm, James, Michael McKee, and Mark Skidmore. 1993. Fiscal pressure, tax competition, and the introduction of state lotteries. *National Tax Journal* 46(4):463–476.

Atkins, Allen B., and Edward A Dyl. 1995. The lotto jackpot: The lump sum versus the annuity. *Financial Practice and Education* 5(2):107–112.

Baker, R. D., and I. G. McHale. 2009. Modelling the probability distribution of prize winnings in the UK National Lottery: Consequences of conscious selection. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 172(4):813–834.

Beenstock, Michael, and Yoel Haitovsky. 2001. Lottomania and other anomalies in the market for lotto. *Journal of Economic Psychology* 22(6):721–744.

Blalock, Garrick, David R. Just, and Daniel H. Simon. 2007. Hitting the jackpot or hitting the skids: Entertainment, poverty, and the demand for state lotteries. *American Journal of Economics and Sociology* 66(3):545–570.

Borg, Mary O., and Paul M. Mason. 1988. The budgetary incidence of a lottery to support education. *National Tax Journal* 41(1):75–85.

——. 1990. Earmarked lottery revenues: Positive windfalls or concealed redistribution mechanisms? *Journal of Education Finance* 15(Winter):289–301.

Borg, Mary O., Paul M. Mason, and Stephen L. Shapiro. 1991. *The economic consequences of state lotteries.* New York: Praeger.

Brinner, Roger E., and Charles T. Clotfelter. 1975. An economic appraisal of state lotteries. *National Tax Journal 28*(4):395–404.

Calcagno, Peter T., Douglas M. Walker, and John D. Jackson. 2010. Determinants of the probability and timing of commercial casino legalization in the United States. *Public Choice* 142(1–2):69–90.

Campbell, Noel, and R. Zachary Finney. 2005. Mitigating the combined distributional consequences of the Georgia Lottery for Education and the HOPE Scholarship. *Social Science Quarterly* 86(3):746–758.

Caudill, Stephen B., Sandra K. Johnson, and Franklin G. Mixon Jr. 1995. Economies of scale in state lotteries: An update and statistical test. *Applied Economics Letters* 2(4):115–117.

Chen, Shu-Heng, and Bin-Tzong Chie. 2008. Lottery markets design, micro-structure, and macro-behavior: An ACE approach. *Journal of Economic Behavior and Organization* 76(2):463–480.

Chernoff, Herman. 1981. How to beat the Massachusetts numbers game. *Mathematical Intelligencer* 3(4):166–172.

Ciecka, James, Seth Epstein, and Anthony C. Krautmann. 1996. State lotteries and externalities to their participants. *Atlantic Economic Journal* 24(4):349–360.

Clotfelter, Charles T. 1979. On the regressivity of state-operated "numbers" games. *National Tax Journal* 32(4):543–548.

Clotfelter, Charles T., and Philip J. Cook. 1987. Implicit taxation in lottery finance. *National Tax Journal* 40(4):533–546.

——. 1989. *Selling hope*. Cambridge, Mass.: Harvard University Press.

——. 1990. On the economics of state lotteries. *Journal of Economic Perspectives* 4(4): 105–119.

——. 1993. The "Gambler's Fallacy" in Lottery Play. *Management Science* 39(12): 1521–1525.

Combs, Kathryn L., Jaebeom Kim, and John A. Spry. 2008. The relative regressivity of seven lottery games. *Applied Economics* 40(1–3):35–39.

Conlisk, John. 1993. The utility of gambling. *Journal of Risk and Uncertainty* 6(3):255–275.

Cook, Philip J., and Charles T. Clotfelter. 1993. The peculiar scale economies of lotto. *American Economic Review* 83(3):634–643.

Coughlin, Cletus C., and Thomas A. Garrett. 2009. Income and lottery sales: Transfers trump income from work and wealth. *Public Finance Review* 37(4):447–469.

Creigh-Tyte, Stephen, and Lisa Farrell. 2003. Is the UK National Lottery experiencing lottery fatigue? In *The Economics of Gambling*. New York, NY: Routledge, 165–181.

Davis, J. Ronnie, John E. Filer, and Donald L. Moak. 1992. The lottery as an alternative source of state revenue. *Atlantic Economic Journal* 20(2):1–10.

DeBoer, Larry. 1985. Administrative costs of state lotteries. *National Tax Journal* 38(4): 479–487.

——. 1986. Lottery taxes may be too high. *Journal of Policy Analysis and Management* 5(3):594–596.

——. 1990. Lotto sales stagnation: Product maturity or small jackpots? *Growth and Change* 21(1):73–77.

Elliott, Donald S., and John C. Navin. 2002. Has riverboat gambling reduced state lottery revenue? *Public Finance Review* 30(3):235–247.

Erekson, O. Homer, Glenn Platt, Christopher Whistler, and Andrea L. Ziegert. 1999. Factors influencing the adoption of state lotteries. *Applied Economics* 31(7):875–884.

Erekson, O. Homer, Kimberly M. DeShano, Glenn Platt, and Andrea L. Ziegert. 2002. Fungibility of lottery revenues and support of public education. *Journal of Education Finance* 28(2):301–311.

Evans, Willam N., and Ping Zhang. 2007. The impact of earmarked lottery revenue on K-12 educational expenditures. *Education Finance and Policy* 2(1):40–73.

Farrell, Lisa, and Ian Walker. 1997. *It could be you! But what's it worth? The welfare gain from Lotto*. Institute for Fiscal Studies Working Paper No. W97/04. London: Institute for Fiscal Studies.

——. 1999. The welfare effects of lotto: Evidence from the UK. *Journal of Public Economics* 72(1):99–120.

Farrell, Lisa, Edgar Morgenroth, and Ian Walker. 1999. A time series analysis of UK lottery sales: Long and short run price elasticities. *Oxford Bulletin of Economics and Statistics* 61(4):513–526.

Farrell, Lisa, Roger Hartley, Gauthier Lanot, and Ian Walker. 2000. The demand for lotto: The role of conscious selection. *Journal of Business and Economic Statistics* 18(2):228–241.

Feehan, Patrick, and David Forrest. 2007. Distribution of UK National Lottery grants across local authority areas. *Applied Economic Letters* 14(4–6):361–365.

Filer, John E., Donald L. Moak, and Barry Uze. 1988. Why some states adopt lotteries and others don't. *Public Finance Quarterly* 16(3):259–283.

Forrest, David, and Dika Alagic. 2007. The history of a lottery game that was seldom won. *Journal of Gambling Business and Economics* 1(1):57–68.

Forrest, David, and Robert Simmons. 2003. Sport and gambling. *Oxford Review of Economic Policy* 19(4):598–611.

Forrest, David, O. David Gulley, and Robert Simmons. 2000a. Elasticity of demand for UK National Lottery tickets. *National Tax Journal* 53(4):853–863.

——. 2000b. Testing for rational expectations in the UK National Lottery. *Applied Economics* 32(3):315–326.

——. 2004. Substitution between games in the UK National Lottery. *Applied Economics* 36(7):645–651.

——. 2010. The relationship between betting and lottery play. *Economic Inquiry* 48(1): 26–38.

Forrest, David, Levi Pérez, and Rose Baker. 2010. Evaluating the effects of game design on lotto sales: A case study from Spain. *Journal of Gambling Business and Economics* 4(2):1–19.

Forrest, David, Robert Simmons, and Neil Chesters. 2002. Buying a Dream: Alternative Models of Demand for Lotto. *Economic Inquiry* 40(3):485–496.

Freund, Elizabeth A., and Irwin L. Morris. 2005. The lottery and income inequality in the states. *Social Science Quarterly* 86(suppl.):996–1012.

______. 2006. Gambling and income inequality in the states. *Policy Studies Journal* 34(2): 265–276.

Friedman, Milton, and Leonard J. Savage. 1948. The utility analysis of choices involving risk. *Journal of Political Economy* 56(4):279–304.

Garrett, Thomas A. 2001a. An international comparison and analysis of lotteries and the distribution of lottery expenditures. *International Review of Applied Economics* 15(2): 213–227.

______. 2001b. Earmarked lottery revenues for education: A new test of fungibility. *Journal of Education Finance* 26(3):219–238.

Garrett, Thomas A., and Cletus C. Coughlin. 2009. Inter-temporal differences in the income elasticity of demand for lottery tickets. *National Tax Journal* 62(1):77–99.

Garrett, Thomas A., and Thomas 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, Thomas A., and Russell S. Sobel. 1999. Gamblers favor skewness, not risk: Further evidence from United States' lottery games. *Economics Letters* 63(1):85–90.

——. 2004. State lottery revenue: The importance of game characteristics. *Public Finance Review* 32(3):313–330.

Geronikolaou, George, and George Papachristou. 2007. On the demand for lotteries in Greece. *International Journal of Business and Economics* 6(3):255–259.

Ghent, Linda S., and Alan P. Grant. 2007. Are voting and buying behavior consistent? Evidence from the South Carolina Education Lottery. *Public Finance Review* 35(6):669–688.

Gripaios, Peter, Paul Bishop, and Steven Brand. 2010. A lottery within a lottery? An examination of the distribution of lottery funds in England. *Applied Economics* 42(1–3):63–71.

Grote, Kent R., and Victor A. Matheson. 2006a. Dueling jackpots: Are competing lotto games complements or substitutes? *Atlantic Economic Journal* 34(1):85–100.

——. 2006b. In search of a fair bet in the lottery. *Eastern Economic Journal* 32(4):673–684.

——. 2007. Examining the “halo effect” in lotto games. *Applied Economic Letters* 14(4–6): 307–310.

Gulley, O. David, and Frank A. Scott Jr. 1989. Lottery effects on pari-mutuel tax revenues. *National Tax Journal* 42(1):89–93.

——. 1993. The demand for wagering on state-operated lotto games. *National Tax Journal* 46(1):13–22.

Guryan, Jonathan, and Melissa S. Kearney. 2005. Lucky stores, gambling and addiction: Empirical evidence from state lottery sales. NBER Working Paper No. 11287. Cambridge, Mass.: National Bureau of Economic Research.

——. 2008. Gambling at lucky stores: Empirical evidence from state lottery sales. *American Economic Review* 98(1):458–473.

——. 2010. Is Lottery Gambling Addictive? *American Economic Journal: Economy Policy* 2(3):90–110.

Hauser-Rethaller, Ursula, and Ulrich Konig. 2002. Parimutuel lotteries: Gamblers’ behavior and the demand for tickets. *German Economic Review* 3(2):223–245.

Heberling, Michael. 2002. State lotteries: Advocating a social ill for the social good. *Independent Review* 6(4):597–606.

Hersch, Philip L., and Gerald S. McDougall. 1989. Do people put their money where their votes are? The case of lottery tickets. *Southern Economic Journal* 56(1):32–38.

Humphreys, Brad R., and Victor A. Matheson. 2013. Booms, busts, and gambling: Can gaming revenues reduce budget volatility? In *Boom and bust again: Policy challenges for a commodity-based economy*, edited by David L. Ryan. Alberta: University of Alberta Press.

Jackson, John D., David S. Saurman, and William F. Shughart II. 1994. Instant winners: Legal change in transition and the diffusion of state lotteries. *Public Choice* 80(3–4):245–263.

Jackson, Raymond. 1994. Demand for lottery products in Massachusetts. *Journal of Consumer Affairs* 28(2):313–325.

Kearney, Melissa S. 2005a. The economic winners and losers of legalized gambling. *National Tax Journal* 58(2):281–302.

——. 2005b. State lotteries and consumer behavior. *Journal of Public Economics* 89 (11–12):2269–2299.

Kitchen, Harry, and Scott Powells. 1991. Lottery expenditures in Canada: A regional analysis of determinants and incidence. *Applied Economics* 23(12):1845–1852.

Krautmann, Anthony C., and James E. Ciecka. 1993. When are state lotteries a good bet? *Eastern Economic Journal* 19(2):157–64.

Laitner, John. 1999. Means-tested public assistance and the demand for state lottery tickets. *Economic Dynamics* 2(1):273–290.

Land, Vance Q., and Majeed H. Alsikafi. 1999. A lottery’s impact on instructional and non-instructional expenditures and unrestricted revenues for education. *Journal of Education Finance* 25(2):149–174.

Landry, Craig E., and Michael K. Price. 2007. Earmarking lottery proceeds for public goods: Empirical evidence from U.S. lotto expenditures. *Economics Letters* 95(3):451–455.

Layton, Allan, and Andrew Worthington. 1999. The impact of socio-economic factors on gambling expenditure. *International Journal of Social Economics* 26(1/2/3):430–440.

Lee, Chuan, Chin-Tsai Lin, and Chien-Hua Lai. 2010. Jackpot promotion model for Taiwan lotto. *Applied Economics* 42(4–6):797–801.

Lin, Eric S., and Shih-Ying Wu. 2007. Lottery expenses and charitable contributions—Taiwan's experience. *Applied Economics* 39(16–18):2241–2251.

Livernois, John R. 1986. The taxing game of lotteries in Canada. *Canadian Public Policy* 12(4):622–627.

Markowitz, Harry M. 1952. The utility of wealth. *Journal of Political Economy* 60(2):151–158.

Martin, Robert, and Bruce Yandle. 1990. State lotteries as duopoly transfer mechanisms. *Public Choice* 64(3):253–264.

Mason, Paul M., Jeffrey W. Steagall, and Michael M. Fabritius. 1997. The elasticity of demand for lotto tickets and the corresponding welfare effects. *Public Finance Review* 25(5):474–490.

Matheson, Victor A. 2001. When are state lotteries a good bet? (Revisited). *Eastern Economic Journal* 27(1):55–70.

Matheson, Victor A., and Kent R. Grote. 2003. Jacking up the jackpot: Are lotto consumers fooled by annuity payments? *Public Finance Review* 31(5):550–567.

———. 2004. Lotto fever: Do lottery players act rationally around large jackpots? *Economics Letters* 83(2):233–237.

———. 2005. Rationality and efficiency in lotto games. In *Information efficiency in financial and betting markets*, edited by Leighton Vaughan Williams. New York: Cambridge University Press.

———. 2007. Gamblers' love for variety and substitution among lotto games. *Journal of Gambling Business and Economics* 1(2):85–99.

———. 2008. *U.S. lotto markets.* In *Handbook of sports and lottery markets*, edited by Donald B. Hausch and William T. Ziemba. Amsterdam: Elsevier, 503–524.

_______. 2009. *Spreading the fortune: The distribution of lottery prizes across countries.* Working Paper No.904. Department of Economics, College of Holy Cross.

Mikesell, John L. 1987. The effect of maturity and competition on state lottery markets. *Journal of Policy Analysis and Management* 6(2):251–253.

———. 1989. A note on the changing incidence of state lottery finance. *Social Science Quarterly* 70(2):513–521.

———. 1994. State lottery sales and economic activity. *National Tax Journal* 47(1):165–171.

Mikesell, John L., and Maureen A. Pirog-Good. 1990. State lotteries and crime: The regressive revenue producer is linked with a crime rate higher by 3%. *American Journal of Economics and Sociology* 49(1):7–19.

Mikesell, John L., and C. Kurt Zorn. 1986. State lotteries as fiscal savior or fiscal fraud: A look at the evidence. *Public Administration Review* 46(4):311–320.

———. 1987. State lottery sales: Separating the influence of markets and game structure. *Growth and Change* 18(4):10–19.

———. 1988. State lotteries for public revenue. *Public Budgeting and Finance* 8(1):38–47.

Miller, James D., and Matthew R. Morey. 2003. Power markets: Transferring systematic risk to lottery players. *Public Budgeting and Finance* 23(2):118–133.

Novarro, Neva K. 2005. Earmarked lottery profits: A good bet for education finance? *Journal of Education Finance* 31(1):23–44.

Oster, Emily. 2004. Are all lotteries regressive? Evidence from the Powerball. *National Tax Journal* 57(2):179–187.

Papachristou, George. 2004. The British gambler's fallacy. *Applied Economics* 36(18): 2073–2077.

——. 2006. Is lottery demand elasticity a reliable marketing tool? Evidence from a game innovation in Greece. *International Review of Economics and Business* 53(4):627–640.

——. 2009. Scale economies of lotto once more. *Applied Economic Letters* 16(1–3):319–323.

Papachristou, George, and Dimitri Karamanis. 1998. Investigating efficiency in getting markets: Evidence from the Greek 6/49 Lotto. *Journal of Banking and Finance* 22(6):1597–1615.

Peltzman, Sam. 1976. Towards a more general theory of regulation? *Journal of Law and Economics* 19(2):211–240.

Pierce, Patrick A., and Donald E. Miller. 1999. Variations in the diffusion of state lottery adoptions: How revenue dedication changes morality politics. *Policy Studies Journal* 27(4):696–706.

Purfield, Catriona, and Patrick Waldron. 1999. Gambling on lotto numbers: Testing for substitutability or complementarity using semi-weekly turnover data. *Oxford Bulletin of Economics and Statistics* 61(4):527–544.

Quiggin, John. 1991. On the optimal design of lotteries. *Economica* 59(229):1–16.

Rodgers, William M., and Charles Stuart. 1995. The efficiency of a lottery as a source of public revenue. *Public Finance Quarterly* 23(2):242–254.

Scoggins, John F. 1995. The lotto and expected net revenue. *National Tax Journal* 48(1): 61–70.

Scott, Frank, and John Garen. 1994. Probability of purchase, amount of purchase, and the demographic incidence of the lottery tax. *Journal of Public Economics* 54 (1):121–143.

Scott, Frank A., Jr., and O. David Gulley. 1995. Testing for efficiency in lotto markets. *Economic Inquiry* 33(2):175–188.

Siegel, Donald, and Gary Anders. 2001. The impact of Indian casinos on state lotteries: A case study of Arizona. *Public Finance Review* 29(2):139–147.

Simon, Jonathan. 1999. An analysis of the distribution of combinations chosen by UK National Lottery players. *Journal of Risk and Uncertainty* 17(3):243–276.

Skidmore, Mark, and Mehmet Serkan Tosun. 2008. Do new lottery games stimulate retail activity? Evidence from West Virginia counties. *Journal of Regional Analysis and Policy* 38(1):45–55.

Spindler, Charles J. 1995. The lottery and education: Robbing Peter to pay Paul? *Public Budgeting and Finance* 15(3):54–62.

Stark, Steven R., R. Craig Wood, and David S. Honeyman. 1993. The Florida Education Lottery: It use as a substitute for existing funds and its effects on the equity of school funding. *Journal of Education Finance* 18(Winter):231–242.

Steinnes, Donald N. 1998. Have Native American casinos diminished other gambling in Minnesota? An economic answer based on accessibility. *Regional Analysis and Policy* 28(1):18–32.

Stigler, George J. 1971. The theory of economic regulation. *Bell Journal of Economics and Management Science* 2(1):3–21.

Stranahan, Harriet A., and Mary O. Borg. 2004. Some futures are brighter than others: The net benefits received by Florida Bright Futures scholarship recipients. *Public Finance Review* 32(1):105–126.

Suits, Daniel B. 1977. Gambling taxes: Regressivity and revenue potential. *National Tax Journal* 30(1):19–35.

——. 1979. Economic background for gambling policy. *Journal of Social Issues* 35(3):43–61.

Summers, Susan R., David S. Honeyman, James L. Wattenbarger, and M. David Miller. 1995. An examination of supplantation and redistribution effects of lottery allocations to a community college system. *Journal of Education Finance* 21(Fall):236–253.

Szakmary, Andrew C., and Carol Matheny Szakmary. 1995. State lotteries as a source of revenue: A re-examination. *Southern Economic Journal* 61(4):1167–1181.

Terrell, Dek. 1994. A test of the gambler's fallacy: Evidence from pari-mutuel games. *Journal of Risk and Uncertainty* 8(3):309–317.

Thaler, Richard H., and William T. Ziemba. 1988. Parimutuel betting markets: Racetracks and lotteries. *Journal of Economic Perspectives* 2(2):161–174.

Thiel, Stuart E. 1991. Policy, participation and revenue in Washington state lotto. *National Tax Journal* 44(2):225–235.

Vasche, Jon David. 1985. Are taxes on lotteries too high? *Journal of Policy Analysis and Management* 4(2):269–271.

Vaughan Williams, Leighton. 2005a. Semi-strong and strong form information efficiency in betting markets. In *Information efficiency in financial and betting markets*, edited by Leighton Vaughan Williams. Cambridge: Cambridge University Press.

——. 2005b. Weak form information efficiency in betting markets. In *Information efficiency in financial and betting markets*, edited by Leighton Vaughan Williams. Cambridge: Cambridge University Press.

Vrooman, David H. 1976. An economic analysis of the New York State Lottery. *National Tax Journal* 29(4):482–489.

Walker, Douglas M., and John D. Jackson. 1999. State lotteries, isolation and economic growth in the U.S. *Review of Urban and Regional Development Studies* 11(3):187–192.

——. 2008. Do U.S. gambling industries cannibalize each other? *Public Finance Review* 36(3):308–333.

Walker, Ian. 1998. Lotteries: Determinants of ticket sales and the optimal payout rate. *Economic Policy: A European Forum* 27:357–392.

Walker, Ian, and Juliet Young. 2001. An economist's guide to lottery design. *Economic Journal* 111(475):700–722.

Williams, Robert J., and Robert T. Wood. 2007. The proportion of Ontario gambling revenue derived from problem gamblers. *Canadian Public Policy* 33(3):367–387.

Wisman, Jon D. 2006. State lotteries: Using state power to fleece the poor. *Journal of Economic Issues* 40(4):955–966.

Wohlenberg, Ernest H. 1992. Recent U.S. gambling legalization: A case study of lotteries. *Social Science Journal* 29(2):167–184.

Ziemba, William T., Shelby L. Brumelle, Antoine Gautier, and Sandra L. Schwartz. 1986. *Dr. Z's 6/49 lotto guidebook*. Vancouver: Dr. Z's Investment, Inc.

CHAPTER 34

THE TAXATION OF GAMBLING MACHINES: A THEORETICAL PERSPECTIVE

LEIGHTON VAUGHAN WILLIAMS AND DAVID PATON

1 Introduction

The U.K. gambling sector has experienced a number of regulatory shocks over the past decade, which have led to considerable debate and controversy within the industry and policy-making communities. Although there is a well-established literature on the economic impact of the growth of gambling facilities on local and regional economies in the United States (e.g., Walker and Jackson 2011; Kearney 2005; Siegel and Anders 2001; d'Hauteserre 1998) and the United Kingdom (e.g., Paton, Siegel, and Vaughan Williams 2002, 2004; Paton and Vaughan Williams 2013; Forrest, Gulley, and Simmons 2010), there has been relatively little research on optimal taxation of gambling machines within these facilities.

In this chapter we address this gap by examining the theoretical arguments for taxing gambling machines by means of a levy on machine takings, rather than by means of a license fee levied per machine. Recent tax debates in the United Kingdom provide an ideal context for such a discussion.

A particular feature of the tax debate in the United Kingdom since 2000 has been a stated desire by the government to use economic theory and evidence as a basis for policy changes. As a result, several gambling sectors (in particular betting and bingo) have moved to a system of taxation based on gross profits (gross win)—see Paton, Siegel, and Vaughan Williams (2002) for an introduction to this context. The gambling machine sector proved more resilient, however, to taxation reform for reasons that include successful challenges to tax policy in the courts.[1]

Prior to the implementation of changes announced in the 2012 budget statement in the House of Commons, U.K. gambling machines were taxed in two ways. A value added tax (VAT) was levied at the standard rate on machine takings (cash-in-box) and could be partially offset by recovery of VAT paid on inputs. The Amusement Machine Licence Duty (AMLD) also was also levied as a fixed amount paid for each machine up front on an annual basis. The level of AMLD was linked to the type of machine (higher for higher stakes and prizes machines). In the 2012 budget, however, the Chancellor of the Exchequer announced the introduction of the Machine Games Duty (MGD), a new gross profits tax for machines, to replace AMLD and VAT on machines. The date announced for the introduction of the new duty was February, 1, 2013.

In the following section we present the theoretical arguments underlying the case for a switch from a license fee system of taxation (as with AMLD) to one based on machine takings, that is, a gross profits tax (GPT). In section 3 we consider a few other relevant issues, including equity and externalities. The new duty was introduced on February 1, 2013.

2 Economic Basis for GPT on Gambling Machines

2.1 Background

Since 2001 the U.K. government has moved toward the use of a GPT for gambling. Partly this was in response to arguments (see Paton, Siegel, and Vaughan Williams 2000) that a GPT would lead to a more allocatively efficient outcome than would a tax on stakes. A GPT represents a tax on margins and is likely to encourage firms to shift from a high-price/low-quantity strategy to a low-price/high-quantity strategy. Under quite general assumptions, David Paton, Donald Siegel, and Leighton Vaughan Williams (2000, 2001, 2002) have shown that a tax revenue-neutral shift from a stakes tax to a GPT leads to lower prices in equilibrium and a reduction in the dead-weight loss.

There is a range of published evidence suggesting that the benefit from the introduction of GPT on general betting has indeed been realized (e.g., Paton, Siegel, and Vaughan Williams 2003; National Audit Office 2005).

The question we seek to answer here is whether there is anything in the particular institutional framework surrounding gambling machines which means that the theoretical arguments in favor of a shift to GPT do not apply or that they apply to a lesser extent.

2.2 Economic Efficiency

When economists refer to allocative efficiency, they are referring to the allocation of scarce resources to their best available uses. A basic premise (under certain assumptions) is that, unimpeded, the economy produces optimal outcomes as the forces of demand and supply interact so that resources can be allocated to where they are most desired and, therefore, yield the highest benefit. Because taxes will distort this result, it is generally optimal to keep this distortion to a minimum.

It is useful to model the decisions taken by firms in the gambling machine market in two stages. In the first stage firms decide how many machines to operate and where to locate them. In the second stage firms decide on the optimal pricing strategy for each machine, *given the number of machines chosen in stage* 1. We analyze the impact of a switch from AMLD to a form of GPT on each stage separately. Note that in the analysis below we define price in the standard way in the gambling literature as 1 minus the expected value of a £1 bet. For example, if a machine pays out, on average, 70 pence for each £1 wagered, then the price of a £1 bet on that machine in our model would be 30 pence.

2.2.1 *Stage 1: Decision on Number of Machines*

We consider the case of a switch from a system in which all tax is levied by means of AMLD to a system in which all tax is levied by a GPT. In either case it is optimal for firms to decide to install machines until the marginal revenue, which declines with each additional unit, is equal to the marginal cost. For simplicity we consider the case of a linear marginal revenue curve and constant marginal cost of installation, c.

Consider first the case of AMLD. Denoting marginal revenue as MR, marginal cost of production as MC, the cost of the license as L, and the number of machines as N, we have

$$MR = a - bN$$
$$MC = c + L.$$

The equilibrium number of machines under AMLD (N^{AMLD}) is then found, where $MR = MC$.

$$a - bN^{AMLD} = c + L$$
$$N^{AMLD} = \frac{a - c - L}{b} \tag{34.1}$$

In the case of a GPT, levied at a rate t, we have

$$MR = (a - bN)(1 - t), \text{ where } t = \text{the rate of GPT.}$$
$$MC = c.$$

The equilibrium number of machines, N^{GPT}, is again where $MR = MC$.

$$N^{GPT} = \frac{a - c - at}{b(1 - t)} \tag{34.2}$$

In the simplest case where c equals zero, it is easy to show that any value of t less than 1 will lead to a greater number of machines under GPT than under AMLD. Put another way, if t is set so that $N^{AMLD} = N^{GPT}$, then tax revenue under a GPT will always be greater than under AMLD. A consequence of this is that total surplus (tax + producer surplus) will be greater under GPT, and a GPT is allocatively more efficient than AMLD.[2]

If $c > 0$, it will not always be the case that a shift to GPT will improve allocative efficiency. To see this consider the case in which GPT is set at a rate that maintains the same number of machines as under AMLD, that is, $N^{AMLD} = N^{GPT}$, so that

$$a - c - at = (1 - t)(a - c - L)$$

$$t = \frac{L}{c + L} \tag{34.3a}$$

or

$$L = \frac{ct}{1 - t} \tag{34.3b}$$

If t is levied at a lower rate than $\frac{L}{c+L}$, then the quantity of machines will increase under a GPT. A corollary of this is that if the revenue-neutral rate of GPT (t^*) is such that $t^* < \frac{L}{c+L}$, then the shift to GPT will lead to an increase in the total number of machines.

The example in figure 34.1 helps to illustrate the intuition behind this result. The result holds true also for the case of variable costs, that is, that an increase in the number of machines is more likely the larger the license fee is relative to the marginal cost. However, in this case, the condition under which a shift to GPT will increase the number of machines will be less easy to satisfy.

The example shows that under a GPT the firm is relatively better off (post-tax revenue of £464 [= 1,200–736] instead of £305 [= 1,000–695], the government benefits from higher revenue (£736 under GPT compared to £695 under AMLD), and consumers benefit by being provided with a preferred activity (inferred from the fact that consumers are willing to pay more to participate in it).[3] In sum, AMLD distorts the economy from the optimal outcome.

It can also be argued that AMLD constitutes an entry barrier. In general entry barriers impede the competitive process and innovation. The point is that there is a dynamic

- A firm is choosing between installing machines in its urban pubs or its rural pubs; it has £2000 capital available.
- In rural pubs, each machine costs £1000 and yields gross profit of £1600.
- In urban pubs, each machine costs £2000 and yields gross profit of £3000.
- Under AMLD with a licence cost per machine of £695: the firm sets up a single machine in an urban pub where the added value to the economy is £1000. No machines are set up in the rural pubs. The total tax raised is £695.
- Under GPT with a rate of 23%: the firm sets up two machines in the rural pubs and none in the urban pubs. The added value to the economy is 2 × £600 = £1200. Total tax collected is £736.

FIGURE 34.1 The AMLD versus GPT comparison loss to efficiency, in addition to the static loss considered here. For this reason AMLD acts in favor of incumbents—a classic case, indeed, of regulatory capture.

In summary, a key question is whether a switch from AMLD to GPT is likely to lead to an increase in the number of machines. If we assume that the unit of analysis is an individual firm, the equilibrium number will increase subject to our condition, but this may not translate to any additional machines, especially if N is small. The effect of the switch would be that some firms will increase the number of their machines, some will not, while some new operators will set up. Aggregating up to industry level, the net effect is that the total number of machines increases. The second question is whether we would expect a price reduction in those venues where the number of machines has not increased. Unless all venues are treated as monopolies, we would expect prices overall to decrease following the industry-wide increase in numbers. However, the size of any price decrease will vary between venues and will be related to the nature and extent of competition.

2.2.2 *Stage 2: Pricing Decision*

Once a machine is installed, variable costs associated with extra stakes are close to zero, and hence the profit-maximization problem is equivalent to revenue maximization, noting that gambling revenue is by convention measured as "cash-in-box," that is, total stakes net of any payouts to punters. It is easy to show that the revenue-maximizing price (the margin) given the number of machines is not related to the choice of AMLD or GPT. In contrast, for a quantity-setting firm with some monopoly power, the equilibrium price is affected by a stakes tax.

Where a firm has price p and sales q (where for gambling p refers to the margin and q refers to the stakes), revenue can be given by $R = q \cdot p(q)$.

In the case of no taxation, a firm will maximize revenue by identifying the quantity level, p^* and associated price (p^*), that maximizes R. Under a GPT tax is levied directly on revenue at a rate, t. The revenue function now becomes $R = (1 - t)q \cdot p(q)$. As $(1 - t)$ is a constant, the revenue-maximizing quantity and price remain unchanged at (p^*, q^*). Under AMLD levied at a fixed value of L per machine, the revenue function becomes $R = q \cdot p(q) - L$ and again the equilibrium is (p^*, q^*).

The case of a stakes (or quantity) tax, like the old general betting duty, is somewhat different. A tax levied at a rate of g per unit leads to a revenue function of $R = (1 - g) \cdot q \cdot p[(1 - g) \cdot q]$. As long as the demand curve facing the firm is downward sloping, the revenue-maximizing equilibrium must result in a price, $p > p^*$. The intuition behind this result is that, as the tax is essentially levied on quantities, there is an incentive for firms to prefer a high-price, low-quantity strategy.

In sum, a GPT and an AMLD are less distorting, in terms of price and quantity decisions, than a stakes tax, which leads to a relatively higher price than either a GPT or an AMLD. This analysis is relevant, however, only for the pricing decision *given a fixed number of machines.* Inasmuch as abolishing AMLD leads to an increase in the number of machines at stage 1, it will likely lead to a reduction in price. Thus, in the long run, an AMLD works very much like a stakes tax—it is (indirectly at least) a tax on quantity. In particular, if the number of machines increases, we would expect a lower equilibrium price. If the total number of machines in a region were to increase, then the profit maximizing price will alter. This point is clearest in the context of a single venue.[4] For example, if the number of machines in a casino increases from 1 to 20, then it is clear that the profit-maximizing price for each machine will decrease.

Given this, the theoretical conclusion is similar to that used in the context of LBO's (licensed betting offices); if there is monopoly power, AMLD is allocatively inefficient compared to GPT.

3 Other Issues

3.1 Tax Pass-on

We would expect that if one firm sets the "monopoly" profit-maximizing price that its neighbor firm would run its machine on a lower margin and capture most of the market (assuming there is no product differentiation), thereby forcing the first firm to also lower its margin. This would continue until any further lowering of the margin makes the machine unviable (this story fits with industry suggestions that players are sensitive to odds). In this case a tax increase would mean that the margin would have to be increased by all to retain viability, and a tax cut would cause margins to fall. It would also mean that machines in places with high costs, like town centers, would justify their position with higher margins. To suggest that prices would not be altered is to assume that players are not sensitive to odds in other venues and behave irrationally.

3.2 Equity

In terms of fairness, tax levels under a GPT would be linked to the revenue-generating potential of machines and decisions would not be distorted as much as under the license regime. For example, although the demand for a machine may be low, the machine may still be viable insofar as its costs are covered, notably rent and tax. The key point, therefore, is that while under the current system the machine may not cover the fixed cost of the license, its viability may well be maintained under an equivalent GPT.

3.3 Externalities in Gambling

High taxation on machines (relative to other gambling products) may be justified on the basis that they pose particular risks to players, in term of the potential for problem gambling (part of the "social cost" of gambling). This is part of the wider literature on problem and compulsive gambling (e.g., Ladouceur, Lachance, and Fournier 2009). The problem may perhaps be linked also to the number of machines available, insofar as this is linked to access. If so, this would seem to be a matter for regulatory rather than tax policy.

3.4 VAT

Given a wish to tax gambling machines as a proportion of machine takings, a further issue is whether this should be done on the basis of VAT, by GPT (as with other gambling sectors), or by a mixture of both.

A consequence of removing VAT would mean that many firms that currently operate machines would be brought into partial exemption for the first time, as machines form only a part of their business.

Allowing firms to recover the tax they have paid on inputs will lead to less distortion within an industry than will a straight output tax with no recovery—a situation where a firm has positive added value but where tax solely on output would be greater than that added value. A lack of provision for input tax recovery will also create distortions that encourage vertical integration—a firm that produces and operates machines will not have to pay a VAT on machine rent while a firm that only operates machines will. VAT recovery also provides an incentive to invest, through the capital goods scheme. Removing this provision is likely to lead to less investment, a point that could, of course, be made for all betting sectors.

VAT is already levied on an efficient basis and makes up a significant proportion (over half) of tax on machines. This means that the arguments put forward for a GPT as linking to externalities and removing distortions to decisions are slightly tempered—the fact that VAT is already levied means that the advantages are less than for a total move from AMLD to GPT.

The VAT treatment of gambling machines has, however, been subject to legal challenges, a factor highlighted by the Chancellor of the Exchequer in his 2012 budget statement (HM Treasury 2012).

> One area where I am today making substantial changes is gambling duties. The VAT treatment of gambling machines is being repeatedly challenged by operators in the courts. So I will introduce a new Machine Games Duty—with a standard rate of 20% and a lower rate for low stakes and prize machines of 5% of net takings.

4 Conclusions

Our theoretical analysis leads us to conclude that there are strong arguments from economic theory for the replacement of the AMLD with a system in which all taxation is levied as a proportion of machine takings. In particular, it is likely that, with respect to some firms and some machines, the AMLD constitutes an entry barrier. On its abolition, the most plausible scenario is that the number of machines in operation would increase. This, in itself, provides an economic argument in favor of the abolition of the AMLD. We also argue that the change in the equilibrium number of machines is likely to lead to a reduction in the average price (or an increase in the average payout) and that this will lead to a net welfare gain.

A possible implication of the above analysis is that an increase in the number of machines will be associated with a rise in negative externalities associated with machine gambling. Although this outcome is a possibility, licensing individual machines is unlikely to tackle the problem in an efficient way. It is reasonable to suppose that this issue may be better dealt with by "social regulation" of the location of these machines.

In terms of equity, a move toward taxation of a proportion of machine takings would also be likely to bring this sector more in line with other gambling sectors.

The issue of a VAT is less clear-cut. There are theoretical arguments as to why a VAT may be more efficient than a GPT. However, it might be argued that all gambling sectors should be treated on the same basis, in which case this is an argument for shifting from a GPT to a VAT across the gambling sector.

Another option is to retain a VAT on machine takings and to replace the AMLD with a GPT. Having two tax systems may seem overly complex and lead to dead-weight loss in itself, but this issue is outside the immediate scope of this analysis. Given that a VAT is already levied on cash in-box under the current system and that many operators will be liable for a VAT anyway in other areas, it may be that the complexity of operating two systems will be less than the practical problems of partial exemption. There is also the legal context regarding the VAT treatment of gambling machines, highlighted by repeated challenges by operators in the courts.

In conclusion, there is a clear and strong case for gambling machines to be taxed in proportion to machine takings rather than with a fixed tax per machine. The appropriate balance between the use of a GPT and a VAT as the means of taxing machines is less clear-cut, however, and should take account of legal as well as practical considerations, most notably the ease, cost, and convenience of administration and collection.

Notes

1. For an example of this see Paton and Vaughan Williams (forthcoming).
2. If an increase in the number of machines leads to an increase in negative externalities from gambling, then we cannot be sure that this increase in allocative efficiency will be welfare improving. The issue of problem gambling is briefly considered in section 2.4.
3. An alternative setup is to remove the capital restriction and let the £1,000 and £2,000 figures be annual running costs—in this case only the urban machine is installed under AMLD while under GPT the rural machines are also installed, and there is benefit for consumers, producers, and government from their installation.

4. Note, though, that we are not assuming that every venue will increase the number of machines, only that the total number increases. In practice we might expect some venues to increase the number, others not to change, and others still to enter the market.

References

d'Hauteserre, Anne-Marie. 1998. Foxwoods Casino Resort: An unusual experiment in economic development. *Economic Geography* 74:112–121.

Forrest, David, O. David Gulley, and Robert Simmons. 2010. The relationship between betting and lottery play. *Economic Inquiry* 48(1):26–38.

HM Treasury. 2012. Budget 2012 statement by the chancellor of the Exchequer, the Rt. Hon. George Osborne MP, March 21; www.hm-treasury.gov.uk/budget2012_statement.htm.

Kearney, Melissa S. 2005. The economic winners and losers of legalized gambling. *National Tax Journal* 58(2): 281–302.

Ladouceur, Robert, Stella Lachance, and Patricia-Maude Fournier. 2009. Is control a viable goal in the treatment of pathological gambling? *Behaviour Research and Therapy* 47(3): 189–197.

National Audit Office. 2005. *HM Customs and Excise: Gambling duties*, HC 188 session 2004–2005, January. London: National Audit Office.

Paton, David, and Leighton Vaughan Williams. Forthcoming. Do new gambling products displace old? Evidence from a postcode analysis. *Regional Studies*, 47(6): 963–973.

Paton, David, Donald S. Siegel, and Leighton Vaughan Williams. 2000. *An economic analysis of the options for taxing betting: A report for HM Customs and Excise*, September. London: HM Customs and Excise.

——. 2001. Gambling taxation: A comment. *Australian Economic Review* 34(4):437–440.

——. 2002. A policy response to the e-commerce revolution: The case of betting taxation in the UK. *Economic Journal* 112 (480): F296–F314.

——2003. Evaluation of the gross profits tax on betting: A report for HM Customs and Excise, January. London: HM Customs and Excise.

——. 2004. Taxation and the demand for gambling: New evidence from the United Kingdom. *National Tax Journal* 57(Dec): 847–861.

Siegel, Donald S., and Gary Anders. 2001. The impact of Indian casinos on state lotteries: A case study of Arizona. *Public Finance Review* 29(2):139–147.

Walker, Douglas M., and John J. Jackson. 2011. The effect of legalized gambling on state government revenue. *Contemporary Economic Policy* 29(1):101–114.

Name Index

General Index