Testing Market Efficiency in Betting Markets: Does It Get Around the Joint Hypothesis Problem?

Shingo Goto*  Toru Yamada†

March 2018

Abstract

Testing market efficiency in betting markets does not necessarily get around the joint hypothesis problem because the result depends on the assumed role of bookmakers. We consider a simple model of fixed-odds betting markets in which bookmakers use their superior assessment of outcome probabilities to engage in monopolistic competition. Their rational pricing decisions can induce the Favorite-Longshot Bias even without bettors’ irrationality. Meanwhile, their pricing decisions accommodate bettors’ irrational biases to exploit betting demands. In European football betting markets, we find significant evidence for the Favorite-Longshot Bias and the Hot-Hand Bias, with the former being more significant than the latter. These biases persist over time despite the increased competition in online betting markets, which is difficult to explain only by bettors' irrationality but consistent with the predicted pricing behavior of bookmakers.

Testing market efficiency in betting markets has its own joint hypothesis problem, as both rational pricing effects and irrational biases can coexist in observed biases.

Keywords: Market Efficiency, Joint Hypothesis, Fixed-Odds Betting, Favorite-Longshot Bias, Hot-Hand Bias, Probability Weighting (Prospect Theory)

*Corresponding author. The University of Rhode Island; Email: shingo_goto@uri.edu. Ballentine Hall, 7 Lippitt Road, Kingston, RI 02881, USA. Phone: 401-874-4318.

†Nomura Asset Management Co., Ltd.; Email: t-yamada@nomura-am.co.jp.
Testing Market Efficiency in Betting Markets: Does It Get Around the Joint Hypothesis Problem?

Abstract

Testing market efficiency in betting markets may not necessarily get around the joint hypothesis problem because the result depends on the assumed role of bookmakers. We consider a simple model of fixed-odds betting markets in which bookmakers use their superior assessment of outcome probabilities to engage in monopolistic competition. Their rational pricing decisions can induce the Favorite-Longshot Bias even without bettors’ irrationality. Meanwhile, their pricing decisions accommodate bettors’ irrational biases to exploit betting demands. In European football betting markets, we find significant evidence for the Favorite-Longshot Bias and the Hot-Hand Bias, with the former being more significant than the latter. These biases persist over time despite the increased competition in online betting markets, which is difficult to explain only by bettors’ irrationality but consistent with the predicted pricing behavior of bookmakers. Testing market efficiency in betting markets has its own joint hypothesis problem, as both rational pricing effects and irrational biases can coexist in observed biases.

Keywords: Market Efficiency, Joint Hypothesis, Fixed-Odds Betting, Favorite-Longshot Bias, Hot-Hand Bias, Probability Weighting (Prospect Theory)

JEL Classification: G14, G19, G29, L83
1 Introduction

As the joint hypothesis problem (Fama, 1970) makes it difficult to address questions of market efficiency in financial markets, numerous studies have turned to sports betting markets as a ground for testing market efficiency.¹ Sports betting markets have an attractive feature of fixed goal-posts. That is, each bet has a well-defined termination point at which time its fundamental value is determined exogenously by the game outcomes and known with certainty, whereas the fundamental values are not directly observed in equity markets (e.g. Thaler and Ziemba, 1988). More importantly, since risks associated with game outcomes are mostly idiosyncratic and diversifiable, most predictable patterns in betting returns, if any, would be attributed to mispricing rather than to risk premiums. Consequently, systematic biases in betting prices (odds),² when observed, have been primarily attributed to the irrationality of the betting public (bettors in aggregate).

Testing market efficiency in sports betting markets, however, has its own challenges. In particular, the price formation process is often markedly different between sports betting markets and financial markets. Unlike market makers in financial markets, bookmakers in sports betting markets can incur substantial losses.³ According to Levitt (2004), bookmakers can offer bets only because they are more skilled in assessing the probabilities of game outcomes.


²A betting “price” is the reciprocal of the “odds” of the corresponding bet. We use the “price” and “odds” interchangeably when the contexts are clear.

³Leicester City’s winning of the English Premier League in 2015-16 was a big surprise to many. Major bookmakers like Ladbrokes Coral and William Hill offered fixed odds of 5,000-1 for the Leicester’s victory at the beginning of the season, “to rake in a little cash on something that would never happen.” This turned into one of the largest mistakes in sports betting history. UK bookmakers lost around $15 million for the bet on the 2015-16 Premier League Winner. Please see “Leicester’s Betting Line Was a 5,000-to-1 Blunder” by J. Robinson on Wall Street Journal, May 5, 2016.
outcomes than others. Through a detailed analysis of US NFL betting markets, Levitt (2004) shows that bookmakers do not necessarily “balance the books,” that is, they take some exposure in order to maximize their expected profits.\footnote{When bookmakers “balance the books,” that is, when they equalize the dollar amount of payouts across different outcomes, they receive the “vig” (commissions, margins) regardless of outcomes.} This evidence poses a significant challenge to the testing of market efficiency in sports betting markets, because questions of market efficiency can be addressed only when the books are balanced (Woodland and Woodland, 1994, footnote 7).

In light of Levitt’s (2004) insights, we take a few steps back and ask how bookmakers’ price-setting behavior can affect our evidence of market (in)efficiency in sports betting markets. Suppose we find significant evidence for systematic biases in sports betting markets. Without considering the role of bookmakers in a given structure of betting markets, can we use the evidence to conclude that the betting markets are irrational? Would the evidence help us gain insights into the sources of financial market anomalies, such as momentum and reversal (value) effects in asset returns?\footnote{Momentum and value effects are widely recognized as pervasive features of financial asset prices (e.g. Asness, Moskowitz, and Pedersen, 2013), though what drive these effects remains an enduring question. Moskowitz (2015) uses “value” effects and return reversal effects interchangeably.}

To address these questions, we focus on European football (soccer) betting markets that are primarily characterized by the Fixed-Odds Betting system. In this system, betting odds reflect the amount of dollars that bookmakers commit to pay out to bettors on successful bets per unit dollar. Thus, we can view Fixed-Odds Betting markets as state-contingent claims markets which determine the value of each game outcome. However, the values of game outcomes are not necessarily determined in a competitive equilibrium. In fact, bookmakers in the Fixed Odds Betting system (e.g. European football betting markets) are likely to choose “wrong” prices deliberately to maximize expected profits as their books are not necessarily balanced.\footnote{The concern of “non-market-clearing prices” is more pronounced in the Fixed Odds Betting system (e.g. European football betting markets) than in the Point-Spread Betting system (e.g. US NFL and NBA betting} As such, biased prices in Fixed-Odds Betting markets may
simply reflect the rational price-setting behavior of bookmakers rather than the irrational wagering behavior of the betting public.

This point, however, has been largely overlooked in the voluminous literature on the efficiency of European football betting markets.\(^7\) To highlight the role of bookmakers in Fixed-Odds Betting markets, we consider a simple model of bookmaker-determined prices. Using Shin’s (1991,1992,1993) model as a guidance, and appealing to the one-shot nature of a Fixed-Odds Betting, we focus on a minimal setup in which a risk-neutral bookmaker solves a single-period expected profit maximization problem. Motivated by Levitt’s (2004) insights, each bookmaker is assumed rational in a sense that she knows the objective probabilities of game outcomes. The bookmaker uses the knowledge to exert a market power in her specialties, but she also faces a severe competition from other bookmakers with different specialties. She cannot set the vig (commissions) freely, because bettors would simply walk away from bookmakers who charge higher vigs than others. She thus chooses her prices by maximizing her expected profits by taking the vig as given.\(^8\) Meantime, in solving her problem, the bookmaker takes into account that her pricing decisions affect bettors’ demands for her bets.\(^9\) The aggregate demand for her bets will increase in the ratio $\hat{p}/q$, where $\hat{p}$ is the subjective probability of the outcome perceived by the betting public, and $q$ is the corresponding probability implied by the betting price (odds-implied probability).

This simple model allows us to draw a few interesting implications about the behavior of

---


\(^8\)In Shin’s (1991,1992,1993) model, two bookmakers bid for the monopoly rights to the betting market in the first stage, and the one who submits the lower vig wins the bid and offer odds in the second stage.

\(^9\)Betting demands of non-insiders in Shin’s (1991,1992,1993) model are exogenous and perfectly inelastic at the “balanced-book” level, as his model focuses on the adverse selection problem with insiders. We consider an elastic demand, as Suits (1979) finds that price elasticities of demand in horse-race betting markets are very high – considerably in excess of unity.
bookmaker-determined prices.\footnote{An interesting model extension would consider dynamic and strategic interactions among bookmakers in repeated games with asymmetric information. A price leadership may arise as an equilibrium (e.g. the implicit collusion model of Rotemberg and Saloner, 1986, 1990). The bookmaker in our simple one-short model resembles the price leader in a more sophisticated model with dynamic and strategic interactions. Please see Section 2.4.5 for a discussion.}

First, a bookmaker’s rational pricing induces the “Favorite-Longshot Bias.” That is, an odds-implied probability ($q$) tends to overstate the chance of winning by the Longshot (that has a lower probability to win) and understate the chance of winning by the Favorite (that has a higher probability to win). This implication corroborates Shin’s (1991, 1992, 1993) insight that bookmakers’ pricing decisions are at the origin of the Favorite-Longshot Bias. Consistent with the key role of bookmakers behind the Favorite-Longshot Bias, Bruce and Johnson (2000) show that the Favorite-Longshot Bias exists only in bookmaker-based betting markets but not in Pari-mutuel betting markets for UK horse-races for which both forms of betting are available.

Second, when bettors’ subjective probabilities are biased, bookmakers partially accommodate the biases in their pricing to exploit bettors’ demands. If a bookmaker were a pure monopolist who can set the vig freely, her pricing would disregard the bettors’ biases completely and rely only on her rational probabilistic assessments of the game outcomes. Faced with fierce competition, however, each bookmaker accommodates bettors’ biases, such as the Hot-Hand Bias or/and the Gambler’s Fallacy,\footnote{The “Hot-Hand Bias” refers to the belief in positive autocorrelation of a non-autocorrelated random sequence, and the “Gambler’s Fallacy” refers to the belief in negative autocorrelation of a non-autocorrelated random sequence.} partially in a manner that maximizes her expected profits under the vig constraint.\footnote{This result is broadly consistent with Pope and Peel’s (1989) and Levitt’s (2004) findings that bookmakers optimally exploit bettors’ biases and maximize their profits by setting prices between the efficient prices and those at which the book is balanced. Andrikogiannopoulou and Papakonstantinou (2017) note that the optimal price-setting behavior is in line with the actual practice of a bookmaker who provided their data.} Bettors’ misperceptions of probabilities, as suggested by the prospect theory (Tversky and Kahneman, 1992), also exacerbate the Favorite-Longshot Bias.
Third, increased competitions tend to sustain, rather than suppress, biases in bookmaker-determined prices.

Applying a log-linear approximation (Campbell and Shiller, 1988a, 1988b) to the first-order condition for the bookmaker’s optimal price-setting, we cast our model into a tractable multinomial logit model. We then implement the model in a sample of about 120,000 soccer matches on Football-Data.co.uk between 2000-01 and 2015-16. Evidence for the Favorite-Longshot Bias is very strong and prevalent across leagues and over time. We also find evidence for a significant Hot-Hand Bias, but not for a Gambler’s Fallacy. That is, odds-implied probabilities \(q\) tend to overstate the chance of winnings of teams with strong winning records, and understate the chance of winnings of teams with weak records. This evidence is broadly consistent with Andrikogiannopoulou and Papakonstantinou’s (2017) finding that nearly 80% of individual bettors systematically overweight (underweight) teams on long winning (losing) streaks in European football betting markets. The Hot-Hand Bias generates a reversal effect (e.g., DeBondt and Thaler, 1985; Rabin and Vayanos, 2010) in betting returns, similarly to Moskowitz’s (2015) finding of a value (reversal) effect in US sports betting markets, which he attributes to the bettors’ overreaction to information about game outcomes. Comparing the two biases, the Favorite-Longshot Bias has been larger and more consistent than the Hot-Hand Bias, though the Hot-Hand Bias has also been significant.

---

\(^{13}\)We also conduct our analysis separately using a sample of about 22,000 matches played in the Top Four European Football league divisions (England Premier League, German Bundesliga, Italian Serie A, Spanish La Liga). We use the Top Four Leagues to focus on the most active segments of the European football betting markets (similar to a “Large Cap” universe in stock markets).

\(^{14}\)Our empirical model can also capture the Home-Away Bias (over-betting on home-area teams) in the difference of two intercept terms. However, the intercept terms may also capture other unknown effects and we refrain from discussing the significance of the Home-Away Bias in this study.

\(^{15}\)Moskowitz (2015) examines how point spreads change between the “Open,” “Close,” and “End” (game outcome) of the betting and finds a momentum pattern between the Open and the Close, which is then reversed almost completely between the Close and the End. See also Gandar, Dare, Brown, and Zuber (1998) and Avery and Chevalier (1999) for earlier studies that examine the path of betting prices.
Our sample period (2000-2016) is marked by a significant growth and development of online betting markets. Sports betting markets have become increasingly competitive over time, as bettors’ search costs have diminished.\textsuperscript{16} To examine how systematic biases in sports betting markets have evolved during this period, we examine the performance of two hypothetical long-short portfolios.\textsuperscript{17} The first portfolio aims to exploit the Favorite-Longshot Bias by taking a long position in bets with high prices (low odds) and a short position in bets with low prices (high odds). The second portfolio aims to exploit the Hot-Hand Bias by taking a long position in bets on teams with weak betting returns and a short position in bets on teams with strong betting returns. Both hypothetical portfolio strategies would have produced positive long-short return spreads significantly and consistently between August 2000 and May 2016. The significant effects of the two biases have persisted throughout the sample period despite the rapid growth and developments in European football betting markets, especially on the internet. This finding is difficult to reconcile with models that rely solely on bettor irrationality by ignoring the role of bookmakers, but it is consistent with the predicted price-setting behavior of bookmakers.\textsuperscript{18}

If testing market efficiency in sports betting markets allowed us to get around the joint hypothesis problem, any predictable patterns in betting returns would be attributable to the irrationality of the betting prices. However, evidence of market irrationality in sports betting markets may need to be interpreted with a grain of salt, as sports betting markets and financial markets are organized very differently. This paper uses a simple model to show that studies of market efficiency in sports betting markets have their own joint hypothesis problems, as both rational pricing behavior of bookmakers and irrational biases of bettors

\textsuperscript{16}There are a number of websites aggregating odds across bookmakers in real time. One example is www.oddschecker.com.

\textsuperscript{17}This exercise is hypothetical because one cannot short bookmakers’ odds in practice.

\textsuperscript{18}Oikonomidis, Bruce, and Johnson (2015) document that a significant Favorite-Long Shot Bias prevails even in highly transparent internet-based European football betting markets. Puzzled by the finding, they conclude that “markets may invariably be susceptible to fundamental risk evaluation problems.” Our empirical results are consistent with theirs, but we attribute the findings to the price-setting behavior of bookmakers.
are likely to affect the observed biases in betting prices (odds).

The rest of the paper is organized as follows. Section 2 introduces our model of Fixed-Odds Betting markets and motivates our empirical hypotheses. Section 3 presents data and empirical results. Section 4 concludes.

2 A Model of Fixed-Odds Betting Markets

2.1 Nomenclature and Notation

2.1.1 Fixed-Odds Betting System vs. Point-Spread Betting System

The traditional form of the European football (soccer) betting market is characterized by the Fixed-Odds Betting system, whereas the traditional form of the US NFL and NBA betting markets is the Point-Spread Betting system.\footnote{Fixed-Odds Betting system and Point-Spread Betting system are very different from the Pari-mutuel Betting system in which odds are proportional to quantities of money wagered. In the Pari-mutuel Betting system, bookmakers can make money from the vigs regardless of the game outcomes.} In the Point-Spread Betting system, a bookmaker offers a point spread or a handicap.\footnote{MLB and NHL betting markets do not use the Point-Spread Betting system because scores in baseball and ice hockey matches tend to be much lower than those in American football or basketball games. MLB and NHL betting markets employ an odds or “money line” that resembles the Fixed Odds Betting system, though there are a few differences in rules and conventions. See Woodland and Woodland (1994) for a study of MLB betting markets.} For example, the bookmaker quotes “Home Team $-2.5$” and “Away team $+2.5$.” This means that a bet on the Home Team will win the bet if the Home Team defeats the Away Team by $3$ points or more. However, a bet on the Home Team will lose the bet even when the Home Team defeats the Away Team, if the point difference is $1$ or $2$ points.

In a Fixed-Odds Betting for a soccer match, each bookmaker announces her odds a few days before the match is played. Bookmakers offer Fixed-Odds against each of three possible outcomes of the match, which we index by $j = 1, 0, -1$: the Home Team Win ($j = 1$), Draw
$j = 0$), or Away Team Win ($j = -1$).\textsuperscript{21} This system allows each bettor to make a detailed comparison between the probability of an outcome implied by the Fixed-Odds and his own assessment of the probability before making his wagering decisions.

### 2.1.2 Betting Odds

If an odd for the $j$-th outcome is $R_j$, a successful bet with a size of one yields a payoff of $R_j$ and a profit of $R_j - 1$ (when the $j$-th outcome obtains). $R_j$, $j = 1, 0, -1$, is the outcome-contingent “dividend” on a unit bet on the $j$-th outcome. We can also view $R_j$ as the gross return (one plus return) of a bet. For instance, betting on an outcome with the odds $R_j = 3.3$ will turn a £1 bet into £3.3,\textsuperscript{22} implying a return of 230% if the bet is successful (the $j$-th outcome obtains); and into nothing (£0) with a return of −100% if the bet is unsuccessful.

### 2.1.3 Betting Prices and the Vig

From an investor’s perspective, the market for bets in a football match corresponds to a market for contingent claims with three states of the world, $j = 1, 0, -1$. We can think of state-contingent claims (Arrow-Debreu securities) which pay £1 if the $j$-th state outcome obtains and nothing otherwise, and have their prices determined by $Q_i = 1/R_j$, $j = 1, 0, -1$.

The bookmaker offers odds on all three states, so our analysis rules out incomplete markets. $Q_i$ is the “state price” for the $j$-th outcome. We may simply call it the “betting price” of the $j$-th outcome. The sum of the state prices of all outcomes, $\sum_{j=-1}^{1} Q_j$, gives the price of a portfolio that pays £1 for sure at the end of the match (regardless of the outcome). This

\textsuperscript{21}Although more complicated bets can be placed on the score or on the half-time and full-time results, we will focus on the simplest formulation of the bet. There are papers that focus on econometric predictions of football scores (and hence game outcomes). Examples include Dixon and Coles (1997), Goddard and Asimakopoulos (2004) and Feng, Polson, and Xu (2016), though modeling football scores is outside the scope of this study.

\textsuperscript{22}This can be a Dollar, Euro, etc.
amount is larger than £1 because bookmakers charge “vigs” (profit margins) in offering the bets. Thus, we can write $\sum_{j=-1}^{1} Q_j = 1 + v$, where $v > 0$ is the vig.

### 2.1.4 Objective, Subjective, and Odds-Implied Probabilities

We can also define the “normalized” state prices of the three outcomes by $q_j = Q_j / (\sum_{k=-1}^{1} Q_k) = Q_j / (1 + v)$, so that $\sum_{j=-1}^{1} q_j = 1$, $q_j \in (0, 1)$ for $j = 1, 0, -1$. We can interpret $q_j$ as the “odds-implied probability” of the $j$-th outcome reflected in the bookmaker’s odd $(j = 1, 0, -1)$. Although several studies interpret the “odds-implied probability” as the market’s subjective probability for each outcome, this interpretation would be possible only when the books are balanced (Woodland and Woodland, 1994). Because bookmakers can influence the implied probability $q_j$, we distinguish it from the subjective probability of the betting public, $\hat{p}_j \in (0, 1)$, $j = 1, 0, -1$, $\sum_{j=-1}^{1} \hat{p}_j = 1$.

We use $p^o_j \in (0, 1)$ to denote the “objective (actual) probability” of each outcome, $j = 1, 0, -1$, $\sum_{j=-1}^{1} p^o_j = 1$. The ratio $\hat{p}_j / p^o_j$ captures the bias in the subjective probability $\hat{p}_j$. For example, $\hat{p}_j / p^o_j > 1$ implies that the subjective probability overstates the chance of the $j$-th outcome and $\hat{p}_j / p^o_j < 1$ implies that the subjective probability understates the chance of the $j$-th outcome.

### 2.2 Biases in Betting Markets and Their Possible Relations with Financial Markets

#### 2.2.1 The Hot-Hand Bias and the Gambler’s Fallacy

Several studies suggest that investors’ subjective probabilities ($\hat{p}_j$) deviate from the corresponding objective probabilities ($p^o_j$) in a systematic manner, and the biases are correlated with certain characteristics of the players/teams in the matches. According to an influential study by Gilovich, Vallone, and Tversky (1985), most individuals watching basketball believe in the “Hot-Hand”: they tend to believe that players who make a shot are more
likely to hit the next shot than players who miss a shot. Put differently, most individuals believe in a positive auto-correlation in basketball shots though outcomes of consecutive shots are close to independent, or slightly negatively auto-correlated (Gilovich, Vallone, and Tversky’s, 1985). Camerer (1989) studies US NBA (basketball) betting markets and finds that teams with winning streaks tend to do worse than expected (by point spreads), consistent with the Hot-Hand Bias.

While the “Hot-Hand” bias refers to the belief in positive serial-correlation of a non-serially correlated random sequence, the belief in negative serial-correlation of a non-serially correlated random sequence is called the “Gambler’s Fallacy” (e.g. Clotfelter and Cook, 1993; Terrell, 1994). “Gambler’s Fallacy” derives from a fallacious belief that a small sample should resemble closely the underlying population, which is often referred to as the “Law of Small Numbers” or “Local Representativeness” bias (e.g. Rabin, 2002).

If individuals are prone to the Hot-Hand Bias and/or the Gambler’s Fallacy, both betting returns and asset returns would exhibit return reversal effects (Jagadeesh and Titman, 1993) and/or return momentum effects (DeBondt and Thaler, 1985), respectively (Rabin and Vayanos, 2010). Since risk premiums play very little role in sports betting markets, return reversal and momentum effects in sports betting markets, if any, may reflect bettors’ biases in assessing probabilities. Our model explicitly allows for the possibility of these biases in bettors’ subjective probabilities.

As Camerer (1989), Rabin (2002), and Rabin and Vayanos (2010) note, the Hot-Hand Bias does not necessarily contradict with the Gambler’s Fallacy, as the Hot-Hand Bias may arise from the consequence of the Gambler’s Fallacy. Although individuals tend to exhibit the Gambler’s Fallacy after observing a short sequence of outcomes, they tend to believe in a Hot-Hand after observing long streaks (e.g. Edwards, 1961). Rabin and Vayanos (2010) develop a model in which agents exhibit a Gambler’s Fallacy in the short-run but a Hot-Hand Bias in the long-run, which they associate with return momentum effects in a relatively short horizon and return reversal effects in a long horizon.
2.2.2 The Favorite-Longshot Bias

The Favorite-Longshot Bias refers to an observation that odds-implied probabilities \( q \) overstate the chance of winning by the Longshot (that has a lower probability to win) and understate the winning by the Favorite (that has a higher probability to win). The following, due to Shin (1991,1992,1993), provides a precise definition of the Favorite-Longshot Bias:

**Definition 1** Odds-implied probabilities \( q \) exhibit a Favorite-Longshot Bias when \( q_j/q_i < p_j^o/p_i^o \) if and only if \( p_j^o > p_i^o \), for \( i,j = 1,0,-1 \), \( i \neq j \).

In stock markets, the Favorite-Longshot Bias resembles the phenomenon that stocks with greater skewness tend to have lower returns (“negative skewness premium”).\(^{23}\) Not surprisingly, both the “Favorite-Longshot Bias” and the “negative skewness premium” receive similar theoretical explanations. For example, Golec and Tamarkin (1995,1998) and Garrett and Sobel (1999) argue that the Favorite-Longshot Bias is consistent with some individuals’ skewness preferences. Mitton and Vorkink (2007) also emphasize the role of some investors’ skewness preferences in explaining the negative skewness premium in stock markets. They show that investors with greater demand for positive skewness will consciously choose to remain under-diversified because “diversification is a two-edged sword: it eliminates undesired variance in return distributions, but also eliminates desired skewness” (p.1256).

Julien and Selanie (2000) and Snowbert and Wolfers (2010) argue that the Favorite-Longshot Bias is consistent with bettors’ misperceptions of probabilities, as suggested by the prospect theory (Kahneman and Tversky, 1979; Tversky and Kahneman, 1992) and implemented with a probability weighting function. In explaining the negative skewness premium, Barberis and Huang (2008) develop a model with “cumulative prospecting” investors who are more unhappy for downside risk than happy about upside potential. They

---

\(^{23}\)See Mitton and Vorkink (2007), Brunnermeier, Gollier, and Park (2007), Barberis and Huang (2008), Boyer, Mitton, and Vorkink (2010), and Amaya, Christoffersen, Jacobs, and Vasquez (2015), among others.
show that, under the cumulative prospect theory preferences of Tversky and Kahneman (1992), investors transform objective probabilities using a weighting function that overweights the tails of the probability distribution, and this causes positively skewed securities to become over-priced and to earn negative average risk-adjusted returns.

From a similar but different angle, Brunnermeier, Gollier, and Park (2007) argue that investors tend to overstate the probabilities of large upside potential, because “a small optimistic bias in beliefs typically leads to first-order gains in anticipatory utility and only second-order costs in realized outcomes” (p.1092). In their model, investors optimize over beliefs of outcomes, as opposed to taking probabilities as given (Brunnermeier and Parker, 2005). This optimizing behavior of investors leads to, among other predictions, a strong preference for stocks with positively skewed distributions. This argument naturally extends to both the Favorite-Longshot Bias and the negative skewness premium.

In our model, these theoretical effects affect bookmaker-determined prices though biases in bettors’ subjective probabilities ($\hat{p}_j$) and betting demands both of which are described below. In light of recent evidence on bettors’ misperceptions of probabilities, our analysis will incorporate a probability weighting function that is suggested by Kahneman and Tversky (1992) and Camerer and Ho (1994).

2.3 Bookmakers’ Exposures to Betting Outcomes

Bookmakers can incur gains or losses when the books do not balance. In the Point-Spread Betting system, a bookmaker can seek a point spread to balance the book, i.e., to balance the dollar amount wagered on each outcome. If the books are balanced, bookmakers can make money regardless of the final game outcomes. We can thus view the “market clearing” spread as an expression of the dollar-weighted average opinion of the betting public (Camerer, 1989). In practice, bookmakers may choose “wrong” spreads to take speculative positions when they have superior information (Levitt, 2004).
Bookmakers in the Fixed-Odds Betting system are in different situations. They can experience large losses (or gains), because odds are set before the match. By way of example, let’s consider a match in which there are only two outcomes, $H$ (Home team win) and $A$ (Away team win). To simplify matters, we ignore the vig (i.e., $v = 0$) and consider actuarially fair gambles. Suppose that $2/3$ of the betting public (in terms of the quantity of money) bets on the outcome $H$ and the remaining $1/3$ bets on the outcome $A$. In this case, the actuarially fair odds are $1.5$ for $H$ and $3.0$ for $A$. ($H$ is the Favorite and $A$ is the Longshot in this example.)

Now suppose that the bookmaker fixes and announces the actuarially fair odds at $1.5$ for $H$ and $3.0$ for $A$. In order for the bookmaker’s book to balance, $2/3$ of the wagers should go to $H$ and $1/3$ of the wagers should go to $A$. In this case, $1.5 \times 2/3$ and $3 \times 1/3$ always equal one, thus the bookmaker’s profit is zero regardless of the game outcome. However, this balance hardly obtains because both bets have exactly the same expected return. Risk neutral bettors would be indifferent between the two bets, and hence it is more likely that similar amounts of wagers are placed on the two outcomes rather than they are allocated according to the respective probabilities ($2/3$ for $H$ and $1/3$ for $A$). When equal dollar amounts are wagered on both outcomes, bookmakers would lose money when the outcome $A$ obtains ($1 - 3 \times 1/2 = -0.50$), while they can earn profits when the outcome $H$ obtains ($1 - 1.5 \times 1/2 = +0.25$). As this simple example illustrates, bookmakers in Fixed-Odds Betting markets are exposed to game outcomes even when they offered actuarially fair gambles.

2.4 Model Setup

2.4.1 Bettors’ Subjective Probabilities

The majority of existing studies do not distinguish between the implied probability $q_j$ and the subjective probability $\hat{p}_j$. However, we explicitly consider possible biases in the subjective probability of the betting public ($\hat{p}_j$) and examine how the biases affect the
implied probability \( (q_j) \). We consider two sources that can distort the subjective probability \( (\hat{p}_j) \) from the objective one \( (p^o_j) \).

The first source is the Hot-Hand Bias or the Gambler’s Fallacy. We can incorporate the Hot-Hand Bias or the Gambler’s Fallacy as follows. Let \( X_j \) be the recent winning records of the Home team \( (j = 1) \) or the Away team \( (j = -1) \). We set \( X_0 = 0 \) for the “Draw” case. We then consider the following tractable specification of the subjective probability:

\[
\ln \frac{\hat{p}_j}{p_j} = X_j \delta + \text{const}_j, \quad \text{for } j = 1, 0, -1, \tag{1}
\]

where \( \hat{p}_j \) is a reference probability which may be subject to the probability weighting in the prospect theory, as we describe below. \( \text{const}_j \) is a constant term that plays no significant role in our analysis but ensures the subjective probabilities of the three outcomes to sum to one. \( X_1 \) (or \( X_{-1} \)) is higher when the Home (Away) team has a stronger winning record in recent matches. The Hot-Hand Bias implies \( \delta > 0 \) in (1), and the Gambler’s Fallacy implies \( \delta < 0 \), for \( j = 1 \) and \( j = -1 \). The sign of \( \delta \) may depend on the length of winning records in the measurement of \( X_j \), as the literature suggests that the Gambler’s Fallacy arises in the short run but the Hot-Hand Bias becomes dominant in the long-run (e.g. Rabin, 2002; Rabin and Vayanos, 2010). Although we focus on the Hot-Hand Bias and the Gambler’s Fallacy, due to their close association with the return reversal and momentum effects in financial markets, we can include a set of other team characteristics in \( X_j \) when we hypothesize them to affect the subjective probability of the betting public.

The second source of the bias in the subjective probability is the misperception of probabilities as summarized by the probability weighting function in the prospect theory of Tversky and Kahneman (1992). Recent studies demonstrate the relevance of the prospect theory, especially the significance of probability weighting, in characterizing the wagering behavior of bettors (e.g. Julien and Selanie, 2000; Snowbert and Wolfers, 2010; Andrikogiannopoulou and Papakonstantinou, 2016a, 2016b, 2017). To model the probability weighting, we assume that the reference probability \( \hat{p}_j \) takes the following form, which is suggested
by Tversky and Kahneman (1992) and Camerer and Ho (1994).\footnote{Barberis and Huang (2008) consider a two-parameter asymmetric probability weighting function in a model with cumulative prospect theory, but this study focuses on a single parameter case.}

\begin{equation}
\bar{p}_j = \frac{(p_j^o)^{1-\varepsilon}}{\sum_{i=-1}^{1} (p_i^o)^{1-\varepsilon}}, \quad j = -1, 0, 1. \tag{2}
\end{equation}

The prospect theory typically assumes $\varepsilon \in [0, 1]$ to promote the idea that bettors tend to overreact to small probability events, but underreact to large probabilities (i.e., the Favorite-Longshot Bias). When $\varepsilon = 0$, $\bar{p}_j$ is the same as the objective probability $p_j^o$. A larger deviation of $\varepsilon$ from zero toward one implies a larger distortion of $\bar{p}_j$ from $p_j^o$.

The two equations (1) and (2) imply the following relation between a log odds-ratio of the subjective probabilities of the betting public ($\hat{p}$) and that of the objective probabilities ($p_j$):

\begin{equation}
\ln \frac{\hat{p}_j}{\hat{p}_i} = (1 - \varepsilon) \ln \frac{p_j^o}{p_i^o} + (X_j - X_i) \delta + \text{const}, \quad \text{for } i, j = 1, 0, -1, i \neq j, \tag{3}
\end{equation}

where $\text{const}$ does not play an important role. In (3), $\delta > 0$ (or $\delta < 0$) captures the Hot-Hand Bias (or the Gambler’s Fallacy). $\varepsilon > 0$ is consistent with the Favorite-Longshot Bias, according to the definition in Section 2.2.2.

### 2.4.2 Betting Demands

Bettors subjective probabilities affect the aggregate demand. We consider the following specification of the aggregate betting demand for the $j$-th outcome ($j = 1, 0, -1$).

\begin{equation}
x_j (\hat{p}_j, q_j) = \left( \frac{\hat{p}_j}{q_j} \right)^\eta \quad \text{for } j = 1, 0, -1, \tag{4}
\end{equation}

where $\eta > 0$ is the elasticity of demand with respect to the “betting price per perceived chance,” $q_i/\hat{p}$. Suits (1979) reports that the elasticity $\eta$ is in excess of unity in horse-race betting markets. The downward-sloping iso-elastic demand function of the form (4) appears frequently in the macroeconomics literature with monopolistic competition (e.g., Dixit and Stiglitz, 1977) and in asset pricing models (e.g. Williams, 1993; Carlson, Fisher,
and Giammarino, 2004). It describes that bettors tend to wager more (less) when expected return of the bet, according to their subjective probabilities, \( \hat{p}_j R_j = \hat{p}_j/Q_j \), is higher. The benchmark case is when the implied probability and the subjective probability coincide, i.e., when \( q_j = \hat{p}_j \). In this case, we assume that the aggregate amount wagered on the bet is one.

The demand function (4) is a reduced-form specification of aggregate betting demands that is consistent with the demand of a “representative bettor” who maximizes \( \sum_{j=-1}^1 (\hat{p}_j x_j R_j - \frac{1}{\rho} x_j^\rho) \) ignoring the vig, where \( \rho = 1 + \frac{1}{\eta} > 1 \). This bettor maximizes his anticipated payoff (ignoring the vig), \( \sum_{j=-1}^1 \hat{p}_j x_j R_j \), after subtracting a convex betting cost, \( \sum_{j=-1}^1 \frac{1}{\rho} x_j^\rho, \rho > 1 \), that penalizes concentrated bets on particular game outcomes. We are not proposing this particular objective function as a model of bettors’ wagering decisions, but we can give a simple economic interpretation to the aggregate demand function (4). Being a reduced-form demand function, it is not meant to model interactions of bettors with different risk preferences, heterogeneous probabilistic beliefs, and varying budgets.

As Manski (2006) notes, however, aggregate demands should depend on the joint distribution of preferences, beliefs, and budgets across agents. For example, the literature has suggested locally convex utilities, skewness preferences, prospect theories with probability weighting functions, etc. to explain bettors’ wagering decisions, while Ali (1977) and Gandhi and Serrano-Padial (2015) stress the importance of the heterogeneity in bettors’ subjective probabilities. The list can go on, but the literature is yet to agree on a standard model that can be used to aggregate individual bettors’ wagering decisions to generate a tractable aggregate betting demand.\(^{25}\) Therefore, this study assumes a tractable single-parameter aggregate betting demand that allows us to focus on the price-setting behavior

\(^{25}\)More recently, Andrikogiannopoulou and Papakonstantinou (2016a, 2016b, 2017) analyze a unique dataset of wagering activities by 500 randomly selected individuals in European sports betting markets. Their studies point to the relevance of the prospect theory, especially the significance of probability weighting, in characterizing bettors’ risk preferences and betting demands. They abstract away from bookmakers’ price-setting behavior, however. Thus, their focus on individuals’ betting demands and our focus on bookmakers’ price-setting decisions are complementary.
of monopolistically competitive bookmakers.

Specifically, the single parameter $\eta$ captures the elasticity of the aggregate betting demand. The elasticity parameter $\eta$ increases when different bets (e.g. bets on Real Madrid, Bayern München, Chelsea, Juventus, etc.) become closer substitutes, i.e., when each bet becomes closer to financial securities. We expect $\eta$ to increase with the institutional and technological developments of the markets and with the number of professional bettors and information service providers. Thus, we take the liberty of interpreting $\eta$ as a measure of market sophistication in sports betting markets. In perfectly competitive securities markets, $\eta$ tends to infinity.

2.4.3 The Bookmaker’s Problem

Each bookmaker is more skilled in assessing the probability of each game outcome than others (Levitt, 2004). In our model, the bookmaker is rational in a sense that she employs the objective probability $p_{o}^{j}$ ($j = 1, 0, -1$) in setting her odds for each outcome. Since risks associated with betting outcomes are idiosyncratic and diversifiable across matches, we assume that each bookmaker is risk-neutral, similarly to Shin (1991,1992,1993).

The risk-neutral bookmaker’s problem is to maximize her expected profit by setting $Q_{j}$ for the three possible outcomes ($j = 1, 0, -1$):

$$\max_{\{Q_{j}\}_{j=1,0,-1}} E[\Pi] = \sum_{j=-1}^{1} x_{j}(\hat{p}_{j}, q_{j}) - \sum_{j=-1}^{1} p_{o}^{j} x_{j}(\hat{p}_{j}, q_{j}) \frac{1}{Q_{j}},$$

subject to the vig constraint, $\sum_{j=-1}^{1} Q_{j} = 1 + v$, where $\hat{p}_{j} \in (0,1)$, $p_{o}^{j} \in (0,1)$, and $Q_{j} \in (0,1+v)$. The demand function $x_{j}(\hat{p}_{j}, q_{j})$ is given in (4) with $q_{j} = Q_{j}/(1+v)$. We focus on the interior solution and preclude the unrealistic cases of $q_{j} = 0$ ($R_{j} \to \infty$) or $q_{j} = 1$ ($R_{j} = 1/(1+v) < 1$).
2.4.4 Monopolistic Competition among Bookmakers

If the bookmaker had no vig constraint, the unconstrained solution to her optimization problem would be $Q_j^* = (1 + \frac{1}{n})p_j^o$, $j = 1, 0, -1$, implying $\sum_{j=-1}^{1} Q_j^* = 1 + \frac{1}{n}$. Thus the optimal vig (markup) of the monopolist would be $v^* = \frac{1}{n}$. $Q_j^* (j = 1, 0, -1)$ is the optimal price if the bookmaker were a pure monopolist who could set the vig freely.

In practice, each bookmaker can exert her market power only within her specialized segment, and the vig is determined by competitive forces in the bookmaking market. Since bettors are free to walk away from high-vig betting opportunities for low-vig betting opportunities, each bookmaker has to take the level of vig ($v < \frac{1}{n}$) as given. We introduce $\theta = 1 - \eta v$ as a measure of the degree of monopolistic competition among bookmakers. Note that $v = (1 - \theta)/\eta = (1 - \theta) v^*$ where $v^*$ is the vig level chosen by a pure monopolist. Pure monopoly would imply $\theta = 0$, but the vig level will decrease with $\theta$. That is, a higher value of $\theta \in (0, 1)$ implies that more intense competition among bookmakers.

With the monopolistic competition, the bookmaker has to offer betting prices that are lower than the monopolistic prices, and hence $Q_j < Q_j^*$ or $q_j/p_j^o < \left(1 + \frac{1}{n}\right)/(1 + v)$ for $j = 1, 0, -1$.

2.4.5 A Possible Extension: Strategic Interactions among Bookmakers

The one-shot model of monopolistic competition among bookmakers, proposed above, is admittedly simple. The objective of the model is to highlight important differences in the price formation process between fixed-odds sports betting markets and financial markets. The model does not intend to capture the richness of the interactions that take place in actual betting markets.

Nevertheless, it is worthwhile discussing how strategic interactions among bookmakers and feedback effects of their odds might affect our model implications. A useful benchmark in this context would be a homogenous product Bertrand model, in which the vig is set at the
competitive level and odds-implied probabilities \( (q_j) \) coincide with objective probabilities \( (p_j^o) \) in equilibrium, as long as bookmakers’ assessment of those is correct. That is, there would be no systematic biases in betting prices under the Bertrand equilibrium. When all bookmakers charge a higher vig \( (v) \) and odds-implied probabilities that differ from the objective ones \( (q_j \neq p_j^o) \), any bookmakers would temporarily benefit from undercutting the vig and setting its \( q_j \)’s equal to \( p_j^o \)’s \( (j = -1, 0, 1) \). But this would lead to a “price war” which drives down future profits of the bookmakers. As expected future profits are high in rapidly growing betting markets with high information asymmetry, bookmakers interacting repeatedly and indefinitely would have little incentives to deviate as long as the vig is sufficiently low to deter new entrants. Following the argument of the “implicit collusion model” of Rotemberg and Saloner (1986), \( q_j \)’s that solve the profit maximization problem (5) may be sustained as a subgame perfect equilibrium, even when we account for dynamic and strategic interactions among bookmakers.

Moreover, in the presence of asymmetric information, a bookmaker with superior information in her specialized segment who solve (5) may emerge as a price leader. When a price leader emerges, other bookmakers have to trade off the temporary gain from unilaterally deviating against the costs of inflicting price wars that reduce future profits, but the calculation of the incentive is difficult when the price leader has superior information in her specialized segment (Rotemberg and Saloner, 1990). Thus implications of our simple model can carry through even in the presence of dynamic and strategic interactions among bookmakers. These implications can be cast into a few sensible empirical hypotheses which we can test with a tractable multinomial logit model.

2.5 Model Implications

The first-order condition for the optimality given the constraint \( \sum_{j=-1}^{1} Q_j = 1 + v \) (or \( \sum_{j=-1}^{1} q_j = 1 \)) and possible biases in bettors’ subjective probabilities (3) implies the fol-
lowing identity expressed in terms of “odds ratios”:26

\[ \eta \ln \frac{\hat{p}_j}{\hat{p}_i} - (1 + \eta) \ln \frac{q_j}{q_i} + \ln \left[ \frac{1 + \frac{1}{\eta} p_j^o}{1 + \frac{1}{\eta} p_i^o} \right] - 1 = 0, \text{ for } i, j = 1, 0, -1, i \neq j. \] (6)

The assumption that the bookmaker has to offer betting prices lower than the monopolistic prices (i.e., \( Q_j < Q_j^* \) and hence \( q_j/p_j^o < \left( 1 + \frac{1}{\eta} \right) / (1 + \gamma) \)) ensures that the term inside the square bracket in (6) is positive. This condition also satisfies the second-order condition, \( q_j/p_j^o < \left( 1 + \frac{2}{\eta} \right) / (1 + \gamma) \) for \( j = 1, 0, -1. \)

The first-order condition (6) is nonlinear and difficult to interpret. To obtain useful empirical predictions, we apply the log-linear approximation of Campbell and Shiller (1988a, 1988b) around \( p_j^o/q_j = 1 \) for \( j = 1, 0, -1. \)

**Proposition 1** Let \( p_j^o \) be the objective probability of the \( j \)-th game outcome, \( j = 1, 0, -1, \) which the bookmaker uses. Let \( \hat{p}_j \) denote the subjective probability of the betting public for the same outcome, where the log odds ratio \( \ln \frac{\hat{p}_j}{\hat{p}_i} \) is characterized by the relation (3). Then, solutions to the bookmaker’s constrained optimization problem (5) should satisfy the following approximate identity:

\[ \ln \frac{q_j}{q_i} \approx \frac{1}{1 + \gamma} \ln \frac{p_j^o}{p_i^o} + (X_j - X_i) \delta \phi + \text{const}, \text{ for } i, j = 1, 0, -1, i \neq j, \] (7)

where

\[ \gamma = \frac{\theta \left( 1 + \eta \varphi \right)}{1 + \eta + \eta \theta (1 - \varphi)}, \]

\[ \phi = \frac{\eta \theta}{(1 + \eta) (1 + \theta)} \]

with \( \theta = 1 - \eta \varphi \in (0, 1) \) capturing the degree of monopolistic competition. In (7), \( \text{const} \) is a constant term that plays no significant role in our analysis.

In the case of a pure monopoly (\( \theta = 0 \)), both \( \gamma \) and \( \phi \) become zero and the odds-implied probability \( q_j \) would reveal the bookmaker’s superior assessment of the objective probability

---

26“Odds ratio” here is a statistical/econometric terminology referring to the ratio of two probabilities. The meaning of “odds” here is different from betting odds.
$p_j^0$ fully for $j = 1, 0, -1$. That is, when the bookmaker could set the vig freely to maximize her expected profits, her pricing would reveal her superior probabilistic assessments of the game outcomes, regardless of the Hot-Hand Bias, the Gambler’s Fallacy, or the probability weighting. In another extreme case of perfectly inelastic betting demand (i.e., $\eta \to 0$), $\theta$ tends to one and hence we would have $\gamma = 1$ and $\phi = 0$. That is, a square root pricing rule that resembles Shin’s (1991) obtains. This would not be surprising because Shin (1991) assumes exogenous and perfectly inelastic betting demands. In practice, aggregate betting demands are elastic ($\eta > 0$), and bookmakers face competition from other bookmakers and potential entrants, $\theta \in (0, 1)$. In this setting, the log approximate identity (7) yields a few interesting implications:

First, $\gamma$ is positive, meaning that a Favorite-Longshot Bias obtains according to the definition in Section 2.2.2. The Favorite-Longshot Bias ($\gamma > 0$) arises even in the absence of bettor irrationality ($\delta = 0$ and $\varepsilon = 0$). Thus the bookmaker’s optimal price setting behavior under monopolistic competition gives rise to the Favorite-Longshot Bias. This implication is similar to Shin’s (1991,1992,1993) result, but our model does not assume the presence of insiders (bettors who have inside information about the game outcomes). This implication is also consistent with the empirical evidence of Bruce and Johnson (2000) who show that the Favorite-Longshot Bias exists only in bookmaker-based betting markets but not in Parimutuel betting markets for UK horse-races for which both forms of betting are available. We also note that $\gamma$ increases with $\varepsilon$, meaning that bettors’ misperceptions of probabilities, along the line of the prospect theory (Tversky and Kahneman, 1992), can exacerbate the Favorite-Longshot Bias.

Second, even when bettors’ subjective probabilities exhibit biases (e.g. the Hot-Hand Bias and/or the Gambler’s Fallacy), the bookmaker’s optimal price setting eliminates more than half of the biases, as $\phi \in (0, \frac{1}{2})$. However, the bookmaker will not correct the biases completely, because she can take advantage of bettors’ demands by accommodating their biases in order to maximize her expected profits. Thus, bookmaker-determined prices can
reflect biases in bettors’ beliefs. We cast these implications into empirical hypotheses below.

Third, for a given level of elasticity (or sophistication of the betting market) \( \eta \), the degree of the Hot-Hand Bias \( (\gamma) \) increases with the degree of monopolistic competition among bookmakers, \( \theta \). Also for a given level of elasticity \( \eta \) and bettors’ biases \( \delta \), the degree of the Hot-Hand Bias (or the Gambler’s Fallacy) \( (|\delta \phi|) \) increases with the degree of monopolistic competition, \( \theta \), as \( \phi \) is increasing in \( \theta \). That is, as the competition among bookmakers gets more intense, the degree of the Favorite-Longshot Bias and that of the Hot-Hand Bias (or the Gambler’s Fallacy) increase, rather than decrease.

Fourth, for a given level of \( \theta \), an increase in the elasticity of aggregate demands \( \eta \), reduces the vig and the degree of the Favorite-Longshot Bias \( \gamma \). This implication is consistent with the notion that each bookmaker’s market power in her specialty declines as \( \eta \) increases, i.e., as the betting market gets closer to competitive securities markets. However, the relation between the vig and the Favorite-Longshot Bias is ambiguous, as a decline in the vig can be associated with either an increase in \( \eta \) or an increase in \( \theta \). That is, the Favorite-Longshot Bias does not necessarily decrease with a decline in the vig. On the other hand, \( \phi \) increases with \( \eta \), but \( |\delta| \) is likely to decrease with \( \eta \). Thus, the relation between the betting market’s sophistication (\( \eta \)) and the degree of the Hot-Hand Bias (or the Gambler’s Fallacy) \( (|\delta \phi|) \) is ambiguous. Even when the betting market gets more sophisticated, the degree of the Hot-Hand Bias (or the Gambler’s Fallacy) does not necessarily decline, contrary to what we tend to anticipate.

Among these four implications, the first two give unambiguous empirical predictions, and we cast these into empirical hypotheses below. The remaining two implications, while surprising and intriguing, are difficult to test because we cannot observe the degree of competition among bookmakers (\( \theta \)), that of market sophistication (\( \eta \)), or that of bettors’ biases (\( \delta \)) directly. We only observe the vig, \( v = \frac{1-\theta}{\eta} \), which is decreasing in both \( \theta \) and \( \eta \). We thus ask how the recent growth in sports betting markets, as manifested by a steady decline in the vig, has been associated with the level of the Favorite-Longshot Bias and that
of the Hot-Hand Bias (or the Gambler’s Fallacy) in European football betting markets. We also examine if the degree of the Favorite-Longshot Bias and that of the Hot-Hand Bias (or the Gambler’s Fallacy) differ between games with high vigs and those with low vigs.

2.6 Empirical Hypotheses

2.6.1 Main Hypotheses

We can cast (7) naturally into a multinomial logit regression model,\(^{27}\) which can be implemented in data without any ad hoc modifications. Specifically, using the “Draw” outcome \((j = 0)\) as the “pivot” category and setting \(X_0 = 0\), we can express the log odds ratios of the “Home team win” \((j = 1)\) and the “Away team win” \((j = -1)\) as:

\[
\ln \frac{p_1}{p_0} = (1 + \gamma) \ln \frac{q_1}{q_0} - X_1 \beta + a_1, \\
\ln \frac{p_{-1}}{p_0} = (1 + \gamma) \ln \frac{q_{-1}}{q_0} - X_{-1} \beta + a_{-1},
\]

(10)

where \(\gamma\) is given by equation (8) and

\[
\beta = \frac{\delta \eta \theta}{1 + \eta + \eta \theta (1 - \varepsilon)}.
\]

\(a_1\) and \(a_{-1}\) are constants whose difference may capture the “Home-Away Bias.” But we refrain from pushing this interpretation as these intercept terms may capture other effects.

We implement the empirical model (10) with maximum likelihood to test the following hypotheses:

**Hypothesis 1 (The Favorite-Longshot Bias)** Bookmakers’ price-setting behavior in monopolistic competition induces a Favorite-Longshot Bias, \(\gamma > 0\). The null hypothesis is \(\gamma = 0\).

\(^{27}\) Former studies that use multinomial logit models in betting markets include Figlewski (1979), Vlastakis, Dotsis, Markellos (2009), and Nyberg (2014), among others.
Hypothesis 2 (The Hot-Hand Bias or/and The Gambler’s Fallacy)  The team’s past performance can bias the subjective probabilities of the betting public and affects the odds-implied probabilities, after controlling for the Favorite-Longshot Bias, and hence $\beta \neq 0$. $\beta > 0$ implies a Hot-Hand Bias whereas $\beta < 0$ implies a Gambler’s Fallacy. The null hypothesis is $\beta = 0$.

The Favorite-Longshot Bias ($\gamma > 0$) can arise from the bookmakers’ price setting behavior, even in the absence of bettors’ biases. Thus the bias may not be transferable to financial markets. The probability weighting, as suggested by the prospect theory (Tversky and Kahneman, 1992), can exacerbate the Favorite-Longshot Bias as $\gamma$ increases with $\varepsilon$. Monopolistic competition among bookmakers $\theta > 0$ plays a key role in the generation of the bias.

On the other hand, a nonzero value of $\beta$ is originated from biases in bettors’ subjective probabilities. For example, a return reversal (value) effects associated with the Hot-Hand Bias, if any, may help us understand similar effects in financial markets (Rabin and Vayanos, 2010). The probability weighting $\varepsilon > 0$ can again intensify the observed degree of the Hot-Hand Bias or/and the Gambler’s Fallacy, as the absolute value of $\beta$ increases with $\varepsilon$.

2.6.2 Auxiliary Questions – Effects of Increased Competition and Market Sophistication, Calendar Seasonality

The rapid growth of remote betting markets on the internet has undoubtedly increased the competition among bookmakers. Moreover, an increasing number of professional bettors seek to predict game outcomes more accurately others. There are also an increasing number of information services that help the bettors develop their hunches and predictions and assist them with actual betting strategies. These should contribute to the sophistication of sports betting markets. Although we cannot observe the degree of monopolistic competition among bookmakers ($\theta$) or that of market sophistication ($\eta$) separately, we can observe the
vig \( (v = \frac{1-\theta}{\eta}) \) that is negatively associated with both \( \theta \) and \( \eta \).

The relation between the vig \( (v) \) and the degree of the Favorite-Longshot Bias, \( \gamma \), is ambiguous, because \( \gamma \) is increasing in \( \theta \) but decreasing in \( \eta \). Similarly, the relation between the vig \( (v) \) and the degree of biases in bettors’ beliefs, \( |\beta| \), is also ambiguous, because observed vigs \( (v) \) and unobserved biases \( (\delta) \) are likely to be correlated with each other. These ambiguous relations are interesting because we tend to think that a decline in the vig \( (v) \), due to increased competition and/or improved market sophistication, should be associated with a reduced level of biases in betting prices. It is thus interesting to investigate how \( \beta \) and \( \gamma \) vary over time, and how they differ between matches with high vigs and those with low vigs in the cross-section.

Another interesting empirical question is if bettors’ biases exhibit a calendar seasonality. Doran, Jiang, and Peterson (2012) show that gambling preferences of investors increase at the start of a New Year. Although it is unclear how the increased gambling preference can affect bettors’ biases, it would be interesting to examine if the observed biases, if any, exhibit some calendar seasonality. We investigate this possibility by conducting the analysis by calendar quarters.

2.6.3 Notes on the Effects of Betting Exchanges

This study builds on Levitt’s (2004) insights that bookmakers possess superior skills in assessing the probabilities of betting outcomes. A recent study by Smith, Paton, and Vaughan-Williams (2009) challenges this view by showing that prices in person-to-person betting exchanges are more informative about game outcomes than bookmaker-determined prices. Franck, Verbeek, and Nüesch’s (2010) finding in European football betting markets also supports this view. These results, however, do not contradict Levitt’s (2004) or our arguments, because rational bookmakers offer “inefficient” prices deliberately to begin with.

Furthermore, the person-to-person betting exchanges have not grown as fast as we had
anticipated. According to the UK Gambling Commission, the bookmaker-based betting markets and betting exchanges had Gross Gambling Yields (GGYs) of £0.64 billion and £0.14 billion, respectively, in 4/2008-3/2009 in the fast growing remote (online) gambling industry. In 4/2015-3/2016, the bookmaker-based betting markets and betting exchanges had GGYs of £1.61 billion and £0.17 billion, respectively. Betting exchanges remain small relative to the size of bookmaker-based betting markets.\footnote{Interestingly, Betfair, the world’s largest internet betting exchange operator that offers person-to-person gambling, decided to launch Fixed-Odds Betting services to compete with traditional bookmakers in 2012. By becoming a bookmaker in Fixed-Odds Betting markets, the exchange can offer much more flexible betting platforms and entice larger betting demands than the person-to-person betting exchange, but it is exposed to the risk of unbalanced books. The company had built a trading team, headed by a former executive at Ladbrokes (one of the largest bookmakers), to prepare for the move.} While it would be definitely interesting to study the competition and interaction (or learning) between bookmakers and betting exchanges, betting exchanges do not pose a major threat to bookmaker-based sports betting markets at this stage.

3 Empirical Evidence

3.1 Data and Descriptives

Our dataset contains about 120,000 soccer matches for which betting odds are reported on Football-Data.co.uk.\footnote{Studies analyzing the same dataset include Franck, Verbeek, and Nüesch (2010), Koning (2012), Dierer (2013), and Nyberg (2014), among others.} The dataset covers matches played in 22 football league divisions in the following countries: Belgium, England (5 leagues), France (2 leagues), Germany (2 leagues), Greece, Italy (2 leagues), Netherlands, Portugal, Scotland (4 leagues), Spain (2 leagues), and Turkey. The dataset does not contain data for Scandinavian countries, Switzerland, or East European countries and Russia. Our sample includes only domestic league matches. It does not include domestic cup matches, international matches, friendly matches, or national team matches.
We conduct most of our empirical analysis using the full sample (matches played in all 22 Leagues) and using the sample of more than 22,000 matches played in the Top Four League divisions: England (Premier League), Germany (Bundesliga), Spain (La Liga), and Italy (Serie A). We analyze the Top Four League separately because they tend to have larger, more active, and more sophisticated betting markets.

Each league (e.g., in the top division) is typically comprised of 18 – 20 teams and organized on a double-robin basis, in which every team plays all others in its league once at home and once away.

The dataset contains the odds offered by 13 different bookmakers, though not all of them post odds in a given period. The dataset collects odds on Fridays in the afternoon for weekend matches, and odds on Tuesdays in the afternoon for midweek matches. We calculate the average odds of different bookmakers for each match in our analysis.

The sample period runs from the 2000-2001 season to the 2015-2016 season, with a total of 16 seasons. Each season runs from summer to next spring.

Panel A of Table 1 tabulates the number of matches in our sample in the sample of all 22 Leagues and that of the Top Four Leagues between 2000-01 and 2015-16. In the whole sample, Home teams won 45.3% of the matches, 26.9% of the matches were draws (tied), and Away teams won 27.8% of the matches.

Panel B of Table 1 provides summary statistics for the betting prices \( Q_j = 1/R_j \) for \( j = 1, 0, -1 \) in the full (pooled) sample of matches played in the Top Four Leagues between 2000-01 and 2015-16. At the average, \( \sum_{j=-1}^{1} Q_j = 110.2\% \) in the whole sample (all 22 leagues) meaning that the average vig in the entire sample is 10.2%. When we limit to the sample of the Top Four Leagues, the average vig is a little lower at 9.0%. After adjusting for the vig, the average odds-implied probabilities for the Home Team Win (\( j = 1 \)), Draw (\( j = 0 \)), and Away Team Win (\( j = -1 \)) are 44.4%, 27.2%, and 28.4%, respectively, in the whole sample. Compared to actual frequencies of the three outcomes, the odds-implied
probabilities \((q_1, q_0, q_{-1})\) tend to understate the probability of a Home Team Win (44.4\% vs. 45.3\%), and overstate the probabilities of a Draw (27.2\% vs. 26.9\%) and an Away Team Win (28.4\% vs. 27.8\%) in the whole sample. These descriptive statistics suggest that, on average, \(q_1\) tends to understate the probability of a Home Team Win while \(q_0\) and \(q_{-1}\) tend to overstate the probabilities of Draws and Away Team Win, respectively. Since \(q_1\) tends to be larger than \(q_0\) and \(q_{-1}\), these descriptive statistics are broadly consistent with the Favorite-Longshot Bias.

In our multinomial logit model (10), we measure each team’s recent game performance \(X_1\) and \(X_{-1}\) by the trailing average betting returns on the team. Let \(j = 1\) and \(j = -1\) denote the Home team and Away team, respectively, facing \(t\)-th game. We calculate the team’s recent game performance by its trailing average of betting returns in the current season:

\[
X_{j,t} = \frac{1}{T} \sum_{l=1}^{\min(T,t)} r_{j,t-l} \text{ for } j = 1, -1
\]

where \(T\) is the number of games and \(r_{j,t-l}\) is the return on team \(j\) in \((t-l)\)-th game in the current season, \(l = 1, 2, ..., \min (T, t)\),

\[
r_{j,t-l} = \begin{cases} 
R_{j,t-l} - 1 & \text{if current team } j \text{ wins} \\
(j = 1 \text{ [-1] for current home [away] team.}) \\
-1 & \text{otherwise.}
\end{cases}
\]

Suppose we consider a match between Barcelona and Real Madrid in Spanish La Liga. When Barcelona is the Home team and Real Madrid is the Away team, both playing \(t\)-th match in the season, we calculate \(X_{1,t}\) as the average betting return on Barcelona in matches Barcelona played in the season (either as the Home team or as the Away team) and \(X_{-1,t}\) as the average betting return on Real Madrid in matches Real Madrid played in the season.

Teams tend to win more often in Home games than in Away games (Table 1, Panel A). Thus, using the number of wins (and losses) may give a misleading picture of the team’s recent performance. We thus use average betting returns to measure a team’s recent performance. We consider returns in the current season only, because the team management and composition often change dramatically during off seasons.
(either as the Home team or as the Away team). The calculation of average returns uses up to and including $T$ recent games in the current season.

We focus on the case of $T = 20$, meaning that the maximum number of matches used in the calculation of average return for each team is 20, but we also investigate the robustness of our results to different choices of $T$. This exercise is interesting because the literature suggests that the Gambler’s Fallacy arises after observing short sequences of outcomes (the Law of Small Number) but the Hot-Hand Bias tends to prevail after observing long sequences. (e.g., Edwards, 1961; Camerer, 1989; Rabin, 2002; Rabin and Vayanos, 2010).

To suppress the effects of outliers, we winsorize $X_{j,t}$ in the top 1% of the sample in each multinomial logit regression.

Panel C of Table 1 reports descriptive statistics of betting returns.

Table 1 about here.

Table 2 reports average vigs over time and across the Top Four Leagues under consideration. Average vigs in different leagues are similar, but they have decreased steadily over time from close to 13% in the 2000-01 season to around 5% in the 2015-16 season. Apparently European football betting markets have become more competitive and more sophisticated in recent years.

Table 2 about here.

3.2 Evidence

3.2.1 Hypothesis 1 and Hypothesis 2

Table 3 reports the results of the multinomial logit model (10) implemented in a sample of 117,467 matches in the whole sample and in a sample of 22,208 matches in the Top Four League. The estimated $\gamma$ coefficients are significantly positive at 0.19 ($t = 16.6$) in the whole sample and 0.15 ($t = 6.86$) in the Top Four League, providing a strong evidence for the Favorite-Longshot Bias (Hypothesis 1).
The estimated $\beta$ coefficients are also significant at 0.09 ($t = 8.35$) in the whole sample and at 0.09 ($t = 4.13$) in the Top Four Leagues. The significantly positive $\beta$ suggests the presence of the Hot-Hand Bias, but not the Gambler’s Fallacy in European Football markets (Hypothesis 2).

In Table 4, we examine if the estimated $\beta$ and $\gamma$ coefficients vary with the length of the measurement interval for each team’s trailing performance, i.e., $T$ in equation (11). Both coefficients are consistently positive and significant. This exercise indicates that the Hot-Hand Bias ($\beta$) is quite robust to the choice of $T$, suggesting a consistent reversal pattern in betting returns. Although Rabin and Vayanos (2010) suggest that the Gambler’s Fallacy tends to arise in the short-run before the Hot-Hand Bias prevails, we do not find evidence for the Gambler’s Fallacy even when we use a short measurement interval like $T = 1$ or $T = 5$. But the evidence for the Hot-Hand Bias is very strong, after controlling for the effects of the Favorite-Longshot Bias.

We also conduct our tests in each of the 22 leagues in our sample. Table 5 tabulate our results. As we can see, estimated $\gamma$ coefficients are all positive and significant (at the 5% level) in 16 leagues, suggesting that the Favorite-Longshot Bias is a pervasive feature in European football betting markets. Estimated $\beta$ coefficients are also positive with only one exception (Belgium), and 10 out of 22 coefficients are significant at the 5% level. The Favorite-Longshot Bias is more consistent across leagues than the Hot-Hand Bias. There are a few variations in the significance of the biases across leagues (countries), especially in the significance of the Hot-Hand Bias. While it would be interesting to ask why the Hot-Hand Bias is particularly strong in some leagues/countries (e.g. Top 3 leagues in England, Greece) but very weak in others (e.g. leagues in Germany, France, Netherlands, Belgium),
we refrain from speculating on the causes in this paper.

Table 5 about here

3.2.2 Calendar Seasonality

To see if there are any calendar seasonality effects (e.g. Doran, Jiang, and Peterson, 2012) in our results, we split each season into four periods by calendar months: July-September, October-December, January-March, and April-June, and then estimate the multinomial logit model for each sub-sample.\textsuperscript{31} We tabulate our estimation results in Table 6. In the whole sample, the Favorite-Longshot Bias is prevalent and significant in each period, though the bias is somewhat weaker at the start of the season (July-September) when we restrict our analysis to the Top Four League. In the Top Four League, the bias gets stronger towards the end of the season. For example, the $\gamma$ coefficient increases from 0.10 ($t = 1.8$) in July-September to 0.22 ($t = 4.9$) in April-June.

We observe similar patterns for the Hot-Hand Bias. In the whole sample, the bias is somewhat weaker at the start of the season. Interestingly, the Hot-Hand Bias gets stronger after the turn of the calendar year. In the Top Four Leagues, the $\beta$ coefficient is not significantly different from zero in the July-September and October-December sub-samples, but gets very significant in the January-March ($\hat{\beta} = 0.23$, $t = 3.8$) and April-June ($\hat{\beta} = 0.19$, $t = 2.5$) sub-samples. This may reflect the fact that the Hot-Hand Bias gets stronger when bettors observe more game outcomes in each season.

Table 6 about here

3.2.3 Effects of Increased Competition

As indicated by the steady decline in average vigs in Table 2, the European football betting markets have become increasingly competitive and more sophisticated over time. To

\textsuperscript{31}Although our dataset includes matches in July, they are not friendly games. We consider competitive matches in domestic leagues only.
examine time-series variations in the level of the Favorite-Longshot Bias and that of the Hot-Hand Bias, we split our sample into four 4-year periods and implement the multinomial logit model in each of the 4-year subsample periods (Table 7).

While we observe some variations in the $\gamma$ and $\beta$ coefficients across different subsample periods, their estimates are significant in all of the four 4-year subsample periods in the whole sample. Both Favorite-Longshot Bias and Hot-Hand Bias have been significant and consistent during the period of rapid developments in European football betting markets. In the Top Four League sample, the Favorite-Longshot Bias has been significantly positive in all sub-sample periods, while, the Hot-Hand Bias has been consistently positive but less significant between 2004-05 and 2011-12.

Table 7 about here

We then introduce a cross-sectional view by comparing our results between Low-Vig games and High-Vig games. In each season, we divide matches between Low-Vig games (vigs below median) and High-Vig games (vigs above median) and repeat the analysis of Table 7 for each group. Results are shown in Table 8.

In the whole sample (Panel A), all $\gamma$ estimates are significantly positive, and most $\beta$ estimates are significantly positive. Again, both the Favorite-Longshot Bias and Hot-Hand Bias are very consistent and persistent over time and across groups sorted by the vig. When we focus on the Top Four League sample (Panel B), the Favorite-Longshot Bias ($\gamma$) was quite consistent across subsample periods, but was larger and more significant in the first-half sample period (from 2000-01 to 2007-08) than in the latter-half sample period (2008-09 to 2015-16). Interestingly, the Hot-Hand Bias ($\beta$) has been more significant in the Low-Vig group, except the second subsample period (2004-05 to 2007-08). That is, in the Top Four Leagues, the Hot-Hand Bias has been more evident in matches with lower vigs, that is, matches for which bookmakers are more competitive. In our model, bookmakers tend to accommodate bettors’ biases more when the competition among bookmakers increases.
(i.e., $\theta$ increases) and/or when the market gets more sophisticated (i.e. $\eta$ increases). Since the vig level is negatively associated with both $\theta$ and $\eta$, the intriguing result in Panel B is consistent with the implication of our model.

Table 8 about here.

3.3 Consistency of Evidence: A Portfolio Analysis

To examine the significance and consistency of our multinomial logit regression results during the period of increased competition, we conduct a simple experiment by considering a hypothetical long-short portfolio of bets on football matches in the whole sample (all 22 leagues) and in the Top Four Leagues. Since one cannot short bets, the analysis here is purely hypothetical. This exercise does not mean to suggest a profitable betting strategy. Rather, this analysis is meant to complement the multinomial logit analysis with an evaluation of hypothetical portfolio strategies. Multinomial logit analyses inherently in-sample analyses. For example, the subsample analysis in the period between 2012-13 to 2015-16 use all matches during the period to obtain $\beta$ and $\gamma$ estimates. The hypothetical portfolio analysis, on the other hand, conducts an out-of-sample evaluation of hypothetical portfolio strategies. That is, long-short portfolios are constructed before each game using only the information at the time of portfolio formation, and then record hypothetical portfolio returns game by game over the sample period.

Specifically, in the hypothetical portfolio analysis, we form long-short portfolios on Fridays (for weekend games) and on Tuesdays (for midweek games) when odds are offered for more than 25 matches (more than 50 teams) in the whole sample (all 22 leagues). When we conduct the analysis in the Top Four Leagues sample, we form portfolios when odds are offered for more than 10 matches (more than 20 teams). We consider the following long-short portfolio strategies on each betting day (Friday or Tuesday), that are rebalanced 1,060 times (707 times in the Top Four Leagues) between August 2000 and May 2016. Portfolio returns are realized within a few days when the games a played.
**Strategy #1**  In the whole sample, we first sort bets into quintile groups according to their odds ($R_j$). We then form a portfolio consisting of long positions in the lowest-odds group. In the meantime, we form a portfolio consisting of short positions in the highest-odds group. Both long and short legs are equally weighted. This hypothetical long-short portfolio aims to exploit the Favorite-Longshot Bias. In the Top Four Leagues sample, we sort bets into tercile groups rather than quintile groups.

**Strategy #2**  In the whole sample, we first sort bets into quintile groups according to the teams’ recent betting returns ($X_j$). For example, for a bet on the Home team, we look at the Home team’s recent record. We then form a portfolio consisting of long positions in the quintile group with the lowest return records, and a portfolio consisting of short positions in the quintile group with the highest return records. Both long and short legs are equally weighted. This hypothetical long-short portfolio aims to exploit the reversal pattern induced by the Hot-Hand Bias. In the Top Four Leagues sample, we sort bets into tercile groups rather than quintile groups.

**Strategy #3**  In the whole sample, we sort bets into $3 \times 3$ groups according to an independent joint sort by $R_j$ and $X_j$ to create nine betting portfolios: Lowest Odds & Lowest Return ($LL$); Lowest Odds & Medium Return ($LM$); Lowest Odds & Highest Return ($LH$); Medium Odds & Lowest Return ($ML$); Medium Odds & Medium Return ($MM$); Medium Odds & Highest Return ($MH$); Highest Odds & Lowest Return ($HL$); Highest Odds & Medium Return ($HM$); and Highest Odds & Highest Return ($HH$) portfolios. All these portfolios are equally weighted. We then form two long-short (spread) portfolios – the first one taking a long (equally-weighted) position in $LL$, $LM$, and $LH$ portfolios, and a short (equally-weighted) position in $HL$, $HM$, and $HH$ portfolios. The second portfolio takes a long (equally-weighted) position in $LL$, $ML$, and $HL$ portfolios and a short (equally-weighted) position in $LH$, $MH$, and $HH$ portfolios. The former long-short portfolio aims to exploit the Favorite-Longshot Bias controlling for the Hot-Hand Bias. The latter long-
short portfolio aims to exploit the Hot-Hand Bias after controlling for the Favorite-Longshot Bias. In the Top Four Leagues sample, we sort bets into $2 \times 2$ groups rather than $3 \times 3$ groups.

We measure and record returns of these hypothetical long-short portfolios (one long-short portfolio for strategy #1, one long-short portfolio for strategy #2, and two long-short portfolios for strategy #3), and tabulate average returns (average long-short return spreads) per match and corresponding t-values in Table 9.

In the full sample, the average return spread on the Favorite-Longshot Bias is highly significant at 14.0% ($t = 9.1$) per match. The average return spread on the Hot-Hand Bias is also highly significant at 5.4% ($t = 4.2$). Once we control for each other, the average return spreads on the Favorite-Longshot Bias and the Hot-Hand Bias are highly significant at 11.7% ($t = 9.4$) and 4.5% ($t = 4.5$).

The portfolio analysis demonstrates that both the Favorite-Longshot Bias and the Hot-Hand Bias generate significant predictability of betting returns, though effects of the former are much larger in magnitude and significance than the latter. Of course, this analysis is purely hypothetical: we cannot short bookmakers’ odds and hence we cannot implement these long-short portfolio positions as actual portfolio betting strategies. All portfolio components considered in this exercise, such as the quintile portfolios in Strategies #1 and #2 and nine portfolios in Strategy #3, all have negative average returns after accounting for the vig. Thus, it would be very difficult to create a portfolio betting strategy that earns positive returns consistently without being able to short the bets.

Nevertheless, the portfolio analysis in this section is useful because the analysis allows us to inspect the consistency of the effects of two biases over time. Figure 1 plots the cumulative sums of the return spreads generated from the two long-short portfolio strategies in Strategy #3. Both lines exhibit upward sloping lines, meaning that both the Favorite-Longshot Bias
and the Hot-Hand Bias have been very consistent over time, especially in the whole sample (Panel A). We can also see that the Favorite-Longshot Bias has had stronger effects than the Hot-Hand Bias. It is tempting to associate the significant Favorite-Longshot Bias with the “negative skewness premium” in equity markets. However, the Favorite-Longshot Bias in sports betting markets may have little to do with financial market anomalies, if the observed bias is specific to the organizational structure of fixed odds betting markets, as our model and Shin’s (1991, 1992, 1993) theory suggest (also corroborated by the evidence of Bruce and Johnson (2000)). On the other hand, the moderate but consistent and significant effect of the Hot-Hand Bias is interesting as it generates a return reversal pattern in European football betting markets, which Moskowitz (2015) dubs as “value” effects in US sports betting markets.

Figure 1 about here.

4 Conclusion

Viewing Fixed-Odds Betting markets as a contingent claims market, one may argue that odds-implied probabilities coincide with the market participants’ subjective probabilities because risks associated with game outcomes are not sources of risk premiums. If so, observed biases, such as the Favorite-Longshot Bias and the Hot-Hand Bias, should reflect the market’s irrationality. This argument, however, overlooks the fact that betting markets are organized very differently from financial markets. In fact, observed prices in betting markets often deviate from market-clearing prices. Bookmakers in betting markets use their superior skills in assessing probabilities of betting outcomes to set prices in a manner that maximizes their expected profits (Levitt, 2004). Consequently, bookmakers often set “wrong prices” (non-market-clearing prices) deliberately, and hence observed inefficiencies in betting markets may simply reflect the bookmakers’ rational price setting behavior.

With this view, we consider a simple model that aims to embody Levitt’s (2004) empiri-
cal insights. Bookmakers are monopolistically competitive. They can use their information advantage in exerting market power in their own specialties, but they face competition from other bookmakers and potential entrants. In this situation, our model shows that the Favorite-Longshot Bias arises from the bookmakers’ rational price setting behavior. Bettors’ misperceptions of probabilities, as suggested by Tversky and Kahneman (1992), can exacerbate the Favorite-Longshot Bias. Furthermore, bookmakers accommodate bettors’ biases to exploit bettors’ wagering demands, which can generate mispricing originated from bettors’ irrationality. We set up our model to incorporate these features with the minimal set of assumptions and parameters, and then implement it in a tractable multinomial logit framework.

We apply the model to a sample of about 120,000 European football matches on FootballData.co.uk in 16 seasons between 2000-01 and 2015-16. We find a significant and consistent evidence for the Favorite-Longshot Bias. We also find a smaller but significant evidence for the Hot-Hand Bias beyond the Favorite-Longshot Bias, which reflects biases in bettors’ subjective probabilities. By overstating the probability of winning by a team with strong recent performance, the Hot-Hand Bias generates a significant return reversal pattern in European football betting markets, which may also suggest a source behind the reversal pattern in financial asset returns (Rabin and Vayanos, 2010).

Effects of the Favorite-Longshot Bias are much larger in magnitude and significance than those of the Hot-Hand Bias. Moreover, consistent with the rational responses of bookmakers to declining vigs, both biases persist throughout the sample period, despite the rapid growth and increased competition (e.g. on the internet) in European football betting markets. Although this evidence is difficult to reconcile with models that rely solely on bettors’ biases, it is consistent with the predicted price-setting behavior of rational bookmakers we describe in this paper.

A useful extension of this study would consider a richer model that explicitly incorporates pricing decisions of bookmakers and feedback effects of their odds through other
bookmakers’ strategies. We may also different bettors with signals of varying precision to learn from the odds and actual betting outcomes. Bookmakers may also adjust odds as demand for various bets is revealed over time. These are some of the possible ways to capture the richness of the interactions that take place in actual betting markets. We leave these model extensions for future research.

The lesson we learned in this paper is that studies of market efficiency in betting markets do not necessarily get around the joint hypothesis problem. A test of market efficiency in sports betting markets involves its own joint tests of the market efficiency and the model of rational bookmaker-determined prices. Observed biases in fixed-odds betting markets, such as the Favorite-Longshot Bias, may be attributable to the optimal price-setting behavior of bookmakers given a specific structure of betting markets, and hence may not be readily applicable to the discussion of the (in)efficiency of financial asset prices. We hope this study brings further attention and discussions on the similarity and dissimilarity between betting markets and financial markets.
References


Table 1: Descriptive Statistics
(16 seasons from 2000-01 to 2015-16)

Panel A: Game Outcomes

<table>
<thead>
<tr>
<th>Samples (Leagues)</th>
<th># of Home Team Wins</th>
<th>Draws</th>
<th>Away Team Wins</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td># of games</td>
<td>(j = 1)</td>
<td>(j = 0)</td>
</tr>
<tr>
<td>All (22 Leagues)</td>
<td>120,661</td>
<td>54,684</td>
<td>45.3%</td>
</tr>
<tr>
<td>Top Four Leagues</td>
<td>22,830</td>
<td>10,688</td>
<td>46.8%</td>
</tr>
</tbody>
</table>

Panel B: Betting Prices $Q_j$

<table>
<thead>
<tr>
<th>Summary Stats</th>
<th>All 22 Leagues</th>
<th>Top Four Leagues</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Home Team Wins $Q_1$</td>
<td>Draws $Q_0$</td>
</tr>
<tr>
<td>Average</td>
<td>48.9%</td>
<td>30.0%</td>
</tr>
<tr>
<td>Std.Dev.</td>
<td>14.5%</td>
<td>6.8%</td>
</tr>
<tr>
<td>Min.</td>
<td>3.6%</td>
<td>5.8%</td>
</tr>
<tr>
<td>5%</td>
<td>24.7%</td>
<td>22.0%</td>
</tr>
<tr>
<td>25%</td>
<td>40.3%</td>
<td>28.8%</td>
</tr>
<tr>
<td>Median</td>
<td>47.9%</td>
<td>30.5%</td>
</tr>
<tr>
<td>75%</td>
<td>57.5%</td>
<td>31.7%</td>
</tr>
<tr>
<td>95%</td>
<td>74.8%</td>
<td>35.7%</td>
</tr>
<tr>
<td>Max</td>
<td>99.0%</td>
<td>80.0%</td>
</tr>
</tbody>
</table>

Panel C: Team-Level Trailing Average Betting Returns

<table>
<thead>
<tr>
<th>Summary Stats</th>
<th>All 22 Leagues</th>
<th>Top Four Leagues</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average</td>
<td>-2.2%</td>
<td>-1.9%</td>
</tr>
<tr>
<td>Std.Dev.</td>
<td>51.2%</td>
<td>54.6%</td>
</tr>
<tr>
<td>Min</td>
<td>-100%</td>
<td>-100%</td>
</tr>
<tr>
<td>5%</td>
<td>-100%</td>
<td>-100%</td>
</tr>
<tr>
<td>25%</td>
<td>-30.3%</td>
<td>-31.1%</td>
</tr>
<tr>
<td>Median</td>
<td>-3.2%</td>
<td>-3.7%</td>
</tr>
<tr>
<td>75%</td>
<td>23.2%</td>
<td>22.8%</td>
</tr>
<tr>
<td>95%</td>
<td>71.8%</td>
<td>76.7%</td>
</tr>
<tr>
<td>99% - Max (winsorized)</td>
<td>153.5%</td>
<td>165.1%</td>
</tr>
</tbody>
</table>

Notes to Table 1: Panel A reports the number of matches in the sample of All 22 Leagues (on Football-Data.co.uk) and that of Top Four Leagues (England, Germany, Italy, Spain) by match outcomes between 2000-01 and 2015-16 seasons (16 seasons in total). Panel B tabulates descriptive statistics for betting prices, $Q_j = \frac{R_j}{j}$ for $j = 1$ (Home Team Wins), $j = 0$ (Draws), and $j = -1$ (Away Team Wins). Panel C reports descriptive statistics of trailing average betting returns on a team. Trailing average returns is a sum of up to (and including) 20 recent realized betting returns of the team in the current season, winsorized at 99%.
Table 2: Average Vigs by Seasons and Leagues (Top Four Leagues)

<table>
<thead>
<tr>
<th>Year (Season)</th>
<th>England Premier League</th>
<th>Germany Bundesliga</th>
<th>Italy Serie A</th>
<th>Spain La Liga</th>
</tr>
</thead>
<tbody>
<tr>
<td>2000-01</td>
<td>13.3%</td>
<td>13.5%</td>
<td>13.5%</td>
<td>13.5%</td>
</tr>
<tr>
<td>2001-02</td>
<td>12.4%</td>
<td>12.7%</td>
<td>12.6%</td>
<td>12.7%</td>
</tr>
<tr>
<td>2002-03</td>
<td>11.7%</td>
<td>11.9%</td>
<td>12.0%</td>
<td>11.9%</td>
</tr>
<tr>
<td>2003-04</td>
<td>11.1%</td>
<td>11.9%</td>
<td>11.8%</td>
<td>11.8%</td>
</tr>
<tr>
<td>2004-05</td>
<td>10.5%</td>
<td>11.2%</td>
<td>11.5%</td>
<td>11.1%</td>
</tr>
<tr>
<td>2005-06</td>
<td>10.0%</td>
<td>10.7%</td>
<td>10.8%</td>
<td>10.7%</td>
</tr>
<tr>
<td>2006-07</td>
<td>10.1%</td>
<td>10.5%</td>
<td>10.8%</td>
<td>10.6%</td>
</tr>
<tr>
<td>2007-08</td>
<td>9.6%</td>
<td>10.2%</td>
<td>10.4%</td>
<td>10.2%</td>
</tr>
<tr>
<td>2008-09</td>
<td>7.9%</td>
<td>9.3%</td>
<td>9.5%</td>
<td>9.3%</td>
</tr>
<tr>
<td>2009-10</td>
<td>7.5%</td>
<td>8.7%</td>
<td>8.8%</td>
<td>8.6%</td>
</tr>
<tr>
<td>2010-11</td>
<td>7.1%</td>
<td>7.3%</td>
<td>7.8%</td>
<td>7.4%</td>
</tr>
<tr>
<td>2011-12</td>
<td>6.6%</td>
<td>6.7%</td>
<td>6.9%</td>
<td>6.9%</td>
</tr>
<tr>
<td>2012-13</td>
<td>5.5%</td>
<td>6.0%</td>
<td>6.1%</td>
<td>6.0%</td>
</tr>
<tr>
<td>2013-14</td>
<td>5.0%</td>
<td>5.9%</td>
<td>5.9%</td>
<td>5.5%</td>
</tr>
<tr>
<td>2014-15</td>
<td>4.6%</td>
<td>5.4%</td>
<td>5.3%</td>
<td>5.3%</td>
</tr>
<tr>
<td>2015-16</td>
<td>4.7%</td>
<td>5.4%</td>
<td>5.4%</td>
<td>5.2%</td>
</tr>
</tbody>
</table>

Notes to Table 2: This table reports average vigs (margins, commissions) charged by bookmakers for bets on matches in the Top Four Leagues by seasons from 2000-01 to 2015-16. Vigs are calculated as $v = \sum_{j=-1}^{1} R_j^{-1} - 1$, where $R_j$ is the odds on the $j$-th outcome ($j = 1, 0, -1$).
Table 3: Multinomial Logit Model Estimation Results

(16 seasons from 2000-01 to 2015-16)

<table>
<thead>
<tr>
<th>Leagues</th>
<th># of games</th>
<th>γ: coef.</th>
<th>t-val.</th>
<th>β: coef.</th>
<th>t-val.</th>
<th>pseudo-R²</th>
</tr>
</thead>
<tbody>
<tr>
<td>All (22 Leagues)</td>
<td>117,467</td>
<td>0.19</td>
<td>16.6***</td>
<td>0.09</td>
<td>8.35***</td>
<td>5.6%</td>
</tr>
<tr>
<td>Top Four Leagues</td>
<td>22,208</td>
<td>0.15</td>
<td>6.86***</td>
<td>0.09</td>
<td>4.13***</td>
<td>7.7%</td>
</tr>
</tbody>
</table>

*p < 0.10, **p < 0.05, ***p < 0.01

Notes to Table 3: We estimate the multinomial logit model in the sample of All 22 leagues and that of Top Four Leagues (England, Germany, Italy, Spain) for the full sample period (16 seasons from 2000-01 to 2015-16). This table reports maximum likelihood coefficient estimates of the multinomial logit model,

\[
\ln \frac{p_1^o}{p_0^o} = (1 + \gamma) \ln \frac{q_1}{q_0} - X_1\beta + a_1, \\
\ln \frac{p_{-1}^o}{p_0^o} = (1 + \gamma) \ln \frac{q_{-1}}{q_0} - X_{-1}\beta + a_{-1},
\]

where the subscript 1 and -1 denote "Home team win" and "Away team win" outcomes respectively for each sample. "Draw" outcomes (indicated by subscript 0) are used as the pivot/base category in estimating the multinomial logit model. \(X_j (j = 1, -1)\) is the trailing average returns on betting on the \(j\)-th team \((J = 1\) for the Home team and \(J = -1\) for the Away team) in the past matches (maximum 20 matches) where each betting return on the \(j\)-th team is

\[
r_j = \begin{cases} 
R_j - 1 & \text{if the team } j \text{ wins} \\
-1 & \text{otherwise}
\end{cases}
\]

where \(R_j\) denotes the odds. Suppose we consider a match between Barcelona and Real Madrid in Spanish La Liga Primera Division. When Barcelona is the Home team \((j = 1)\) and Real Madrid is the Away team \((j = -1)\), we calculate \(X_1\) as the average return of bets on Barcelona in the previous (up to) 20 matches Barcelona played (either as the Home team or as the Away team) and \(X_{-1}\) as the average return of bets on Real Madrid in the previous (up to) 20 matches Real Madrid played (either as the Home team or as the Away team). The table reports coefficient estimates of \(\gamma\) and \(\beta\), along with their corresponding t-values. Intercept terms \((a_1\) and \(a_{-1}\)) are omitted from the table. The table also reports McFadden’s (1974) pseudo-\(R^2\),

\[
\text{pseudo-}R^2 = 1 - \frac{\ln L(\text{with predictors})}{\ln L(\text{without predictors})}
\]

where \(\ln L(\text{with predictors})\) and \(\ln L(\text{without predictors})\) are log likelihoods with \(q_j\) and \(X_j\) and without them (intercept only), respectively. We conduct the analysis with the “mlogit” package (Croissant, 2013) on https://cran.r-project.org.
Table 4: Multinomial Logit Model Estimation Results (by Record Calc. Lengths)

(16 seasons from 2000-01 to 2015-16)

<table>
<thead>
<tr>
<th>$T$: # of games</th>
<th>All 22 Leagues</th>
<th></th>
<th>Top Four Leagues</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\gamma$</td>
<td>$\beta$</td>
<td>$\gamma$</td>
<td>$\beta$</td>
</tr>
<tr>
<td></td>
<td>coeff. t-val.</td>
<td>coeff. t-val.</td>
<td>coeff. t-val.</td>
<td>coeff. t-val.</td>
</tr>
<tr>
<td>1</td>
<td>0.16 15.2***</td>
<td>0.01 3.05***</td>
<td>0.13 6.33***</td>
<td>0.02 2.08***</td>
</tr>
<tr>
<td>$\leq 5$</td>
<td>0.17 15.9***</td>
<td>0.04 6.63***</td>
<td>0.14 6.58***</td>
<td>0.05 3.23***</td>
</tr>
<tr>
<td>$\leq 10$</td>
<td>0.18 16.4***</td>
<td>0.07 8.15***</td>
<td>0.14 6.68***</td>
<td>0.07 3.41***</td>
</tr>
<tr>
<td>$\leq 15$</td>
<td>0.18 16.4***</td>
<td>0.08 7.91***</td>
<td>0.15 6.73***</td>
<td>0.08 3.54***</td>
</tr>
<tr>
<td>$\leq 20$</td>
<td>0.19 16.6***</td>
<td>0.09 8.35***</td>
<td>0.15 6.86***</td>
<td>0.09 4.13***</td>
</tr>
<tr>
<td>$\leq 25$</td>
<td>0.19 16.7***</td>
<td>0.09 8.53***</td>
<td>0.15 6.88***</td>
<td>0.10 4.22***</td>
</tr>
<tr>
<td>$\leq 30$</td>
<td>0.19 16.7***</td>
<td>0.09 8.37***</td>
<td>0.15 6.84***</td>
<td>0.09 3.97***</td>
</tr>
<tr>
<td>$\leq 35$</td>
<td>0.19 16.7***</td>
<td>0.09 8.39***</td>
<td>0.15 6.85***</td>
<td>0.09 4.02***</td>
</tr>
<tr>
<td>$\leq 40$</td>
<td>0.19 16.7***</td>
<td>0.09 8.51***</td>
<td>0.15 6.85***</td>
<td>0.10 4.05***</td>
</tr>
</tbody>
</table>

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes to Table 4: We estimate the multinomial logit model in the sample of All 22 Leagues and that of Top Four Leagues (England, Germany, Spain, Italy) for the full sample period (2000-01 to 2015-16) by varying the maximum lengths of periods used in the calculation of recent winning records $X_j$, $j = 1,...,16$. $T = 1$ means that only the most recent betting return on the team is used in the calculation of $X_j$. $T \leq 5$ means that average returns of up to (and including) five recent bets on the team is used in the calculation of $X_j$, and so forth. Pseudo $R^2$s are 5.6% in all regressions using the sample of All 22 Leagues, and 7.7% in all regressions using a sample of the Top Four Leagues. (Varying $T$ has very small effects on pseudo $R^2$s in these regressions.) Please see Notes to Table 3 for more details of our model and methodology.
Table 5: Multinomial Logit Model Estimation Results (By Leagues)

(16 seasons from 2000-01 to 2015-16)

| Leagues                          | \# of games | \(\gamma\)   | t-val. | \(\beta\)   | t-val. | pseudo-
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Matches Played in the Top Four Leagues</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(R^2)</td>
</tr>
<tr>
<td>England Premier League</td>
<td>5,919</td>
<td>0.15</td>
<td>3.82***</td>
<td>0.14</td>
<td>3.17***</td>
<td>8.4%</td>
</tr>
<tr>
<td>German Bundesliga</td>
<td>4,751</td>
<td>0.05</td>
<td>1.05</td>
<td>0.04</td>
<td>0.90</td>
<td>5.8%</td>
</tr>
<tr>
<td>Italian Serie A</td>
<td>5,620</td>
<td>0.29</td>
<td>6.43***</td>
<td>0.10</td>
<td>2.27**</td>
<td>9.0%</td>
</tr>
<tr>
<td>Spanish La Liga</td>
<td>5,918</td>
<td>0.08</td>
<td>2.04**</td>
<td>0.11</td>
<td>2.31**</td>
<td>7.3%</td>
</tr>
<tr>
<td>Matches Played in England, Germany, Italy, and Spain outside the Top Four Leagues</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>England Championship</td>
<td>8,636</td>
<td>0.19</td>
<td>3.39***</td>
<td>0.15</td>
<td>3.58***</td>
<td>2.8%</td>
</tr>
<tr>
<td>England League 1</td>
<td>8,639</td>
<td>0.20</td>
<td>3.69***</td>
<td>0.14</td>
<td>3.39***</td>
<td>3.0%</td>
</tr>
<tr>
<td>England League 2</td>
<td>8,640</td>
<td>0.10</td>
<td>1.61</td>
<td>0.05</td>
<td>1.35</td>
<td>2.2%</td>
</tr>
<tr>
<td>England Conference</td>
<td>5,769</td>
<td>0.16</td>
<td>2.76***</td>
<td>0.04</td>
<td>0.84</td>
<td>4.3%</td>
</tr>
<tr>
<td>German 2.Bundesliga</td>
<td>4,748</td>
<td>0.11</td>
<td>1.63</td>
<td>0.01</td>
<td>0.25</td>
<td>3.2%</td>
</tr>
<tr>
<td>Italian Serie B</td>
<td>6,890</td>
<td>0.23</td>
<td>4.18***</td>
<td>0.04</td>
<td>0.97</td>
<td>4.0%</td>
</tr>
<tr>
<td>Spanish La Liga 2</td>
<td>7,103</td>
<td>0.12</td>
<td>1.76*</td>
<td>0.10</td>
<td>2.24**</td>
<td>7.1%</td>
</tr>
<tr>
<td>Matches Played in France, Scotland, Belgium, Greece, Netherlands, Portugal, and Turkey</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>French Ligue 1</td>
<td>5,702</td>
<td>0.15</td>
<td>2.80***</td>
<td>0.06</td>
<td>1.30</td>
<td>4.4%</td>
</tr>
<tr>
<td>French Ligue 2</td>
<td>5,893</td>
<td>0.21</td>
<td>2.64***</td>
<td>0.06</td>
<td>1.39</td>
<td>2.2%</td>
</tr>
<tr>
<td>Scottish Premiership</td>
<td>3,505</td>
<td>0.19</td>
<td>3.73***</td>
<td>0.13</td>
<td>2.11**</td>
<td>10.3%</td>
</tr>
<tr>
<td>Scottish Championship</td>
<td>2,793</td>
<td>0.13</td>
<td>1.75*</td>
<td>0.02</td>
<td>0.32</td>
<td>5.0%</td>
</tr>
<tr>
<td>Scottish League 1</td>
<td>2,787</td>
<td>0.22</td>
<td>2.58**</td>
<td>0.19</td>
<td>2.63***</td>
<td>4.6%</td>
</tr>
<tr>
<td>Scottish League 2</td>
<td>2,784</td>
<td>0.09</td>
<td>1.23</td>
<td>0.07</td>
<td>1.10</td>
<td>5.4%</td>
</tr>
<tr>
<td>Belgian First Division A</td>
<td>4,213</td>
<td>0.19</td>
<td>3.60***</td>
<td>−0.02</td>
<td>−0.31</td>
<td>7.7%</td>
</tr>
<tr>
<td>Superleague Greece</td>
<td>3,536</td>
<td>0.31</td>
<td>6.16***</td>
<td>0.22</td>
<td>3.72***</td>
<td>12.8%</td>
</tr>
<tr>
<td>Dutch Eredivisie</td>
<td>4,695</td>
<td>0.18</td>
<td>3.92***</td>
<td>0.06</td>
<td>1.17</td>
<td>9.8%</td>
</tr>
<tr>
<td>Portuguese Primeira Liga</td>
<td>4,217</td>
<td>0.28</td>
<td>5.48***</td>
<td>0.05</td>
<td>0.99</td>
<td>10.1%</td>
</tr>
<tr>
<td>Turkish Süper Lig</td>
<td>4,709</td>
<td>0.16</td>
<td>3.26***</td>
<td>0.08</td>
<td>1.63**</td>
<td>7.1%</td>
</tr>
</tbody>
</table>

\(p < 0.10, \ast p < 0.05, \ast\ast p < 0.01\)

Notes to Table 5: We estimate the multinomial logit model in the sample of all matches on FootballData.co.uk and report regression results by leagues. The sample period covers 16 seasons from 2000-01 to 2015-16. Please see Notes to Table 3 for more details of our model and methodology.
Table 6: Multinomial Logit Model Estimation Results (by Calendar Months)
(16 seasons from 2000-01 to 2015-16)

Panel A: All 22 Leagues

<table>
<thead>
<tr>
<th>Periods</th>
<th># of games</th>
<th>(\gamma) coeff.</th>
<th>t-val.</th>
<th>(\beta) coeff.</th>
<th>t-val.</th>
<th>Pseudo-(R^2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full (July-June)</td>
<td>117,467</td>
<td>0.19</td>
<td>16.6***</td>
<td>0.09</td>
<td>8.35***</td>
<td>5.61%</td>
</tr>
<tr>
<td>July-September</td>
<td>20,027</td>
<td>0.19</td>
<td>6.45***</td>
<td>0.02</td>
<td>1.85*</td>
<td>4.75%</td>
</tr>
<tr>
<td>October-December</td>
<td>36,711</td>
<td>0.21</td>
<td>10.1***</td>
<td>0.13</td>
<td>6.31***</td>
<td>5.43%</td>
</tr>
<tr>
<td>January-March</td>
<td>38,527</td>
<td>0.16</td>
<td>8.19***</td>
<td>0.16</td>
<td>6.11***</td>
<td>5.41%</td>
</tr>
<tr>
<td>April-June</td>
<td>22,202</td>
<td>0.24</td>
<td>9.81***</td>
<td>0.18</td>
<td>5.17***</td>
<td>7.12%</td>
</tr>
</tbody>
</table>

\(^* p < 0.10, ^{**} p < 0.05, ^{***} p < 0.01\)

Panel B: Top Four Leagues

<table>
<thead>
<tr>
<th>Periods</th>
<th># of games</th>
<th>(\gamma) coeff.</th>
<th>t-val.</th>
<th>(\beta) coeff.</th>
<th>t-val.</th>
<th>Pseudo-(R^2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full (July-June)</td>
<td>22,208</td>
<td>0.15</td>
<td>6.86***</td>
<td>0.09</td>
<td>4.13***</td>
<td>7.70%</td>
</tr>
<tr>
<td>July-September</td>
<td>2,963</td>
<td>0.10</td>
<td>1.80*</td>
<td>0.05</td>
<td>1.67</td>
<td>7.09%</td>
</tr>
<tr>
<td>October-December</td>
<td>7,168</td>
<td>0.15</td>
<td>3.76***</td>
<td>0.06</td>
<td>1.46</td>
<td>7.39%</td>
</tr>
<tr>
<td>January-March</td>
<td>7,195</td>
<td>0.14</td>
<td>3.52***</td>
<td>0.23</td>
<td>3.88***</td>
<td>7.39%</td>
</tr>
<tr>
<td>April-June</td>
<td>4,882</td>
<td>0.22</td>
<td>4.88***</td>
<td>0.19</td>
<td>2.51***</td>
<td>9.09%</td>
</tr>
</tbody>
</table>

\(^* p < 0.10, ^{**} p < 0.05, ^{***} p < 0.01\)

Notes to Table 6: We estimate the multinomial logit model in the sample of All 22 Leagues (Panel A) and that of Top Four Leagues (England, Germany, Spain, Italy) (Panel B) for the full sample period (2000-01 to 2015-16) by splitting each season into four periods by calendar months: July-September, October-December, January-March, and April-June. Please see Notes to Table 3 for more details of our model and methodology.
Table 7: Multinomial Logit Model Estimation Results (by Years)

Panel A: All 22 Leagues

<table>
<thead>
<tr>
<th>Years (Seasons)</th>
<th># of games</th>
<th>γ</th>
<th>t-val.</th>
<th>β</th>
<th>t-val.</th>
<th>pseudo-R²</th>
</tr>
</thead>
<tbody>
<tr>
<td>2000-01 to 2015-16</td>
<td>117,467</td>
<td>0.19</td>
<td>16.6***</td>
<td>0.09</td>
<td>8.35***</td>
<td>5.61%</td>
</tr>
<tr>
<td>2000-01 to 2003-04</td>
<td>27,299</td>
<td>0.28</td>
<td>10.9***</td>
<td>0.11</td>
<td>4.94***</td>
<td>5.42%</td>
</tr>
<tr>
<td>2004-05 to 2007-08</td>
<td>29,921</td>
<td>0.19</td>
<td>8.20***</td>
<td>0.07</td>
<td>3.24***</td>
<td>5.32%</td>
</tr>
<tr>
<td>2008-09 to 2011-12</td>
<td>30,067</td>
<td>0.18</td>
<td>8.45***</td>
<td>0.11</td>
<td>5.31***</td>
<td>5.73%</td>
</tr>
<tr>
<td>2012-13 to 2015-16</td>
<td>30,180</td>
<td>0.12</td>
<td>6.01***</td>
<td>0.07</td>
<td>3.73***</td>
<td>5.89%</td>
</tr>
</tbody>
</table>

*p < 0.10, **p < 0.05, ***p < 0.01

Panel B: Top Four Leagues

<table>
<thead>
<tr>
<th>Years (Seasons)</th>
<th># of games</th>
<th>γ</th>
<th>t-val.</th>
<th>β</th>
<th>t-val.</th>
<th>pseudo-R²</th>
</tr>
</thead>
<tbody>
<tr>
<td>2000-01 to 2015-16</td>
<td>22,208</td>
<td>0.15</td>
<td>6.86***</td>
<td>0.09</td>
<td>4.13***</td>
<td>7.70%</td>
</tr>
<tr>
<td>2000-01 to 2003-04</td>
<td>5,325</td>
<td>0.20</td>
<td>3.84***</td>
<td>0.15</td>
<td>3.21***</td>
<td>5.75%</td>
</tr>
<tr>
<td>2004-05 to 2007-08</td>
<td>5,628</td>
<td>0.23</td>
<td>4.85***</td>
<td>0.04</td>
<td>0.93</td>
<td>7.62%</td>
</tr>
<tr>
<td>2008-09 to 2011-12</td>
<td>5,627</td>
<td>0.08</td>
<td>2.01*</td>
<td>0.08</td>
<td>1.94*</td>
<td>7.69%</td>
</tr>
<tr>
<td>2012-13 to 2015-16</td>
<td>5,628</td>
<td>0.12</td>
<td>3.19***</td>
<td>0.11</td>
<td>2.50***</td>
<td>9.56%</td>
</tr>
</tbody>
</table>

*p < 0.10, **p < 0.05, ***p < 0.01

Notes to Table 7: We estimate the multinomial logit model in the sample of All 22 Leagues (Panel A) and that of Top Four Leagues (England, Germany, Spain, Italy) (Panel B) in the full sample period (2000-01 to 2015-16 seasons) and in each of the four four-year periods. Please see Notes to Table 3 for more details of our model and methodology.
Table 8: Multinomial Logit Model Estimation Results (by Years and Vig-Groups)

Panel A: All 22 Leagues

<table>
<thead>
<tr>
<th>Years (Seasons)</th>
<th>Groups</th>
<th>avg. vigs</th>
<th>(\gamma)</th>
<th></th>
<th>(\beta)</th>
<th></th>
<th>pseudo- (R^2)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>coeff.</td>
<td>t-val.</td>
<td>coeff.</td>
<td>t-val.</td>
<td></td>
</tr>
<tr>
<td>2000-01 to 2003-04</td>
<td>Low-Vig</td>
<td>11.0%</td>
<td>0.27</td>
<td>7.36***</td>
<td>0.20</td>
<td>5.81***</td>
<td>5.26%</td>
</tr>
<tr>
<td>2004-05 to 2007-08</td>
<td>Low-Vig</td>
<td>9.9%</td>
<td>0.17</td>
<td>5.08***</td>
<td>0.06</td>
<td>1.84*</td>
<td>5.10%</td>
</tr>
<tr>
<td>2008-09 to 2011-12</td>
<td>Low-Vig</td>
<td>7.8%</td>
<td>0.16</td>
<td>5.32***</td>
<td>0.07</td>
<td>4.02***</td>
<td>5.85%</td>
</tr>
<tr>
<td>2012-13 to 2015-16</td>
<td>Low-Vig</td>
<td>5.6%</td>
<td>0.12</td>
<td>4.17***</td>
<td>0.09</td>
<td>3.15***</td>
<td>6.13%</td>
</tr>
</tbody>
</table>

*\(p < 0.10, **p < 0.05, ***p < 0.01\)

Panel B: Top Four Leagues

<table>
<thead>
<tr>
<th>Years (Seasons)</th>
<th>Groups</th>
<th>avg. vigs</th>
<th>(\gamma)</th>
<th></th>
<th>(\beta)</th>
<th></th>
<th>pseudo- (R^2)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>coeff.</td>
<td>t-val.</td>
<td>coeff.</td>
<td>t-val.</td>
<td></td>
</tr>
<tr>
<td>2000-01 to 2003-04</td>
<td>Low-Vig</td>
<td>10.8%</td>
<td>0.22</td>
<td>3.02***</td>
<td>0.21</td>
<td>2.98***</td>
<td>5.85%</td>
</tr>
<tr>
<td>2004-05 to 2007-08</td>
<td>Low-Vig</td>
<td>9.2%</td>
<td>0.18</td>
<td>2.79***</td>
<td>0.00</td>
<td>0.04</td>
<td>7.52%</td>
</tr>
<tr>
<td>2008-09 to 2011-12</td>
<td>Low-Vig</td>
<td>6.9%</td>
<td>0.12</td>
<td>2.02**</td>
<td>0.16</td>
<td>2.26**</td>
<td>8.12%</td>
</tr>
<tr>
<td>2012-13 to 2015-16</td>
<td>Low-Vig</td>
<td>4.8%</td>
<td>0.10</td>
<td>1.89*</td>
<td>0.14</td>
<td>2.22*</td>
<td>9.82%</td>
</tr>
</tbody>
</table>

*\(p < 0.10, **p < 0.05, ***p < 0.01\)

Notes to Table 8: In each season (year), we sort all matches into two groups by their vigs in the sample of All 22 Leagues (Panel A) and that of the Top Four Leagues (England, Germany, Spain, Italy) (Panel B). We then estimate the multinomial logit model for each group in each of the four four-year periods for both Panels. Please see Notes to Table 3 for more details of our model and methodology.
Table 9: Portfolio Analysis
(16 seasons from 2000-01 to 2015-16)

Panel A: All 22 Leagues

<table>
<thead>
<tr>
<th>Strategies</th>
<th>Favorite-Longshot Bias</th>
<th>Hot-Hand Bias</th>
</tr>
</thead>
<tbody>
<tr>
<td>#1 (Bottom 1/5 - Top 1/5)</td>
<td>13.98%</td>
<td>9.05***</td>
</tr>
<tr>
<td>#2 (Bottom 1/5 - Top 1/5)</td>
<td>5.39%</td>
<td>4.18***</td>
</tr>
<tr>
<td>#3 (Bottom 1/3 - Top 1/3 for each)</td>
<td>11.65%</td>
<td>9.40***</td>
</tr>
</tbody>
</table>

*p < 0.10, **p < 0.05, ***p < 0.01

Panel B: Top Four Leagues

<table>
<thead>
<tr>
<th>Strategies</th>
<th>Favorite-Longshot Bias</th>
<th>Hot-Hand Bias</th>
</tr>
</thead>
<tbody>
<tr>
<td>#1 (Bottom 1/3 - Top 1/3)</td>
<td>12.93%</td>
<td>5.10***</td>
</tr>
<tr>
<td>#2 (Bottom 1/3 - Top 1/3)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>#3 (Bottom 1/2 - Top 1/2 for each)</td>
<td>11.22%</td>
<td>5.48***</td>
</tr>
</tbody>
</table>

*p < 0.10, **p < 0.05, ***p < 0.01

Notes to Table 9: On each betting day (Friday for weekend games and Tuesday for midweek games), we consider the four long-short portfolios in 3 strategies that aim to exploit the Favorite-Longshot Bias or the Hot-Hand Bias. Portfolios are (hypothetically) formed and rebalanced 1,060 times (in Panel A) or 707 times (in Panel B) between August 2000 and May 2016. This table reports average long-short portfolio returns (average spreads) and corresponding t-values.

- Strategy #1 bets on wagers in the bottom quintile group (or tercile group in Panel B) and against wagers in the top quintile group (or tercile group in Panel B) of odds. This strategy aims to exploit the Favorite-Longshot Bias.
- Strategy #2 bets on wagers in the bottom quintile group (or tercile group in Panel B) and against wagers in the bottom quintile group (or tercile group in Panel B) of recent betting returns. This strategy aims to exploit the Hot-Hand Bias.
- Strategy #3 is based on a $3 \times 3$ (or $2 \times 2$ in Panel B) joint independent sort by odds and recent returns. This strategy consists of two portfolios: The first portfolio bets on low odds & low returns group, low odds & medium returns group, and low odds & high returns group (with equal weights), and against high odds & low returns group, high odds & medium returns group, and high odds & high returns group (with equal weights). The second portfolio bets on low odds & low returns group, medium odds & low returns group, and high odds & low returns group, and against low odds & high returns group, medium odds & high returns group, and high odds & high returns group (with equal weights).
Figure 1: Cumulative Returns on Hypothetical Long-Short Portfolios
(16 seasons from 2000-01 to 2015-16)

Panel A: All 22 Leagues

Panel B: Top Four Leagues
Notes to Figure 1: This figure plots cumulative sum of returns on hypothetical long-short portfolios betting on the Favorite-Longshot Bias and the Hot-Hand Bias in the All 22 Leagues universe (Panel A) and in the Top Four Leagues universe (Panel B). The initial value of the portfolio is set at 0 in July 2000 and cumulative sums of returns are plotted.

The two portfolios are based on a $3 \times 3$ (in Panel A) or $2 \times 2$ (in Panel B) joint independent sort of bets described as Strategy #3 in Table 9. Please see Notes to Table 9 for the methodology. The portfolios are (hypothetically) formed and rebalanced 1,060 times (in Panel A) and 707 times (in Panel B) between August 2000 and May 2016. These cumulative sums of portfolio returns are purely hypothetical and cannot be implemented in practice as one cannot short bookmakers’ odds. This graph is meant to examine the significance, consistency, and relative magnitude of the Favorite-Longshot Bias and the Hot-Hand Bias in European football betting markets.