

# MARKET EFFICIENCY AND BEHAVIORAL BIASES IN THE SPORTS BETTING MARKET

Angie Andrikogiannopoulou\*  
HEC Geneva

Filippos Papakonstantinou†  
Imperial College London

May 2011

## Abstract

We use both aggregate- and individual-level data from the soccer wagering market to examine whether the prices set by bookmakers on a large number of soccer events are efficient, as well as the extent to which individuals are affected by inefficiencies or contribute to their formation. We find evidence of a mild systematic bias in this market, in particular the favorite-longshot bias (FLB), i.e. bettors' tendency to underbet (overbet) outcomes with short (long) odds. We find that the bias is mainly driven by an underbetting of favorite home and away teams, and an overbetting of longshot draw outcomes. The analysis of individual-level returns indicates that, contrary to the common perception that the majority (or totality) of bettors suffer from the FLB, only 6% of the bettors in the sample systematically bet on biased longshot odds, while 2% of bettors possibly exploit the bias and earn significantly positive returns from betting on favorites. Finally, we find that about 4% of the bettors may suffer from a "home bias" which manifests itself as an overbetting of their home-area team.

*Keywords: Risk Preferences, Behavioral Biases, Market Efficiency, Favorite-longshot, Sports Betting, Gambling, Individuals*

*JEL Classification: D11, D12, D81, G00*

---

\*HEC, University of Geneva and Swiss Finance Institute, 102 Bd Carl Vogt, CH-1211 Geneva 4, Switzerland, e-mail: Angeliki.Andrikogiannopoulou@unige.ch.

†Imperial College Business School, Tanaka Building, South Kensington Campus, London SW7 2AZ, UK, e-mail: fpapakon@imperial.ac.uk, <http://www.imperial.ac.uk/people/f.papakonstantinou>

# 1 Introduction

Since Fama first introduced the concept of market efficiency in 1970, there has been a lot of interest in examining whether market prices reflect available information in various financial markets. While several deviations from market efficiency have been observed in aggregate price data, little is known about the extent to which individual market participants are affected by them or contribute to their formation. Furthermore, little is known about the origin of these inefficiencies: behavioral theories suggest that market anomalies arise because people suffer from systematic biases when forming beliefs, while other studies suggest that inefficiencies can also arise in markets with rational agents. Clearly, an interesting set of questions awaits further empirical investigation: What drives possible inefficiencies? What is the extent to which people exhibit biased behavior? What is the percentage of people that exhibit such behavior? What is the percentage of people who earn positive returns by exploiting, consciously or not, the existing inefficiencies? In this study, we examine these questions using a unique dataset that contains both aggregate- and individual-level data from a less typical financial market that is nevertheless particularly suited for giving us useful insights: the sports wagering market.

As Thaler and Ziemba (1988) point out, sports betting markets provide an idealized laboratory setting for applying efficiency tests because the true value of assets (bets) is revealed with certainty in a short period of time, i.e., when the underlying sporting event takes place. At the same time, the structure of sports betting markets resembles to a large extent that of conventional financial markets: both markets are populated with a large number of participants with varying degrees of sophistication who risk their capital on the uncertain outcome of future events; information about sports teams and events is widely publicly available, as is information about companies and stock prices; the sports bookmaker is the analog of the market maker in financial markets while sports handicappers play the role of financial analysts. Fur-

thermore, it has often been suggested that many stock market traders share a similar psychology with casino gamblers, e.g., both are governed by the thrill of chasing higher returns, they both over-invest in domestic assets/teams (“home bias”), etc.

In this study, we observe the prices set by an online bookmaker on a large number of soccer matches along with the individual betting choices over time for a randomly selected sample of the bookmaker’s customers. We first use the aggregate price data to perform various tests of the weak-form market efficiency hypothesis, which posits that current prices reflect all information contained in historical prices. We find evidence of a mild systematic bias in this market and, in particular, in a way that is consistent with the favorite-longshot bias that has been widely documented in race-track gambling.<sup>1</sup> The favorite-longshot bias refers to the tendency of bettors to consistently overbet outcomes with long odds (longshots) and underbet outcomes with short odds (favorites) relative to their observed frequency of winning. As a result, market prices, i.e., betting odds, over-predict (under-predict) the probability of the favorite (longshot) outcomes occurring. Furthermore, we investigate whether this inefficiency pertains to specific outcomes of soccer matches, i.e., the home team win, the draw or the away team win. For each outcome, we compare the objective win probabilities for bets whose odds fall within a specific range with the market probabilities implied by the bookmaker’s odds and find that the former are significantly higher than the latter mainly for home and away favorites while the opposite holds mainly for draw longshots. Regression analysis yields statistically strong evidence that the favorite-longshot bias holds for all outcomes, and statistically weak evidence that there is possibly an additional compounding effect associated with draw outcomes being overbet. We finally explore the presence of profitable betting opportunities and find that simple strategies of betting against the public do not yield substantial

---

<sup>1</sup>Some of the most influential studies in this literature include Griffith (1949), Ali (1977), Asch et al. (1982), Busche and Hall (1988), Hausch et al. (1981), Snyder (1978), Figlewski (1979), Swidler and Shaw (1995).

positive returns and could only be marginally profitable in the best case, depending on the level of the various costs involved.

Subsequently, we use the individual-level data to study the extent to which bettors in our sample exhibit the favorite-longshot bias. The individual-level analysis indicates that, contrary to the common perception that the majority (or even the totality) of bettors contributes to the generation of the bias, i) in our sample only 38% of the bettors have bet more than a handful of times on strong longshot outcomes, and ii) only 6% of the bettors earn significant negative returns relative to the bookmaker’s commission when betting on strong longshots (i.e., have systematically bet on wrong odds), indicating that few longshot outcomes are inefficient and few bettors suffer from the bias. These results are particularly interesting in view of the representative agent framework that is usually employed in the literature, which implies that the favorite-longshot bias governs the behavior of all bettors and is present in all matches that involve longshot outcomes.

Apart from the favorite-longshot bias, the individual-level information available in our dataset is useful in exploring several other biases that have a natural analog in typical financial markets. For instance, we examine the extent to which individuals exhibit the “home bias” that is often observed in the stock market, i.e., the over-investing (over-betting) in home-area stocks (teams). We identify the favorite team for each individual based on his/her zip code and find that around 4% of the bettors might exhibit a bias related to their home-area team, i.e., earn significantly negative returns relative to the bookmaker’s commission from betting on soccer matches in which their home team is participating.

The remainder of the paper is organized as follows. Section 2 reviews the related literature on the efficiency of sports wagering markets and the biases that have been identified in several financial markets. Section 3 describes the data. Section 4 presents the results of our empirical analysis of the favorite-longshot bias, both at the aggregate

level and at the individual-bettor level. Section 5 presents the results of the individual-level analysis of the “home bias” and Section 6 concludes.

## 2 Literature Review

The vast majority of earlier studies that examine market efficiency in wagering markets has focused on parimutuel racetrack betting, where the money bet on all outcomes is pooled, and after the house’s commission is deduced, the remaining pool is shared proportionally among the winning bets. In the parimutuel system, the odds associated with each outcome are not set by a bookmaker and solely reflect the total betting volume in this outcome. Almost all of these studies find that the weak-form of market efficiency is violated since the expected return from betting on favorites is higher than the expected return from betting on longshots. The favorite-longshot bias, i.e., the systematic underbetting (overbetting) of favorite (longshot) horses, has been termed as the “most robust anomalous empirical regularity” in the racetrack. Some studies (e.g., Ali (1977), Hausch et al. (1981)) also detect profitable betting opportunities from following particular wagering rules, while others find that deviations from efficiency are not sufficient to allow for positive returns (e.g., Asch et al. (1982), Snyder (1978)).

Contrary to the parimutuel system, in fixed-odds betting markets the odds are set by a bookmaker, and bettors’ final payout is determined by the odds prevailing at the time they place their bets rather than at the end of the betting period. A lot of studies that focus on these markets reject the hypothesis of weak-form market efficiency, although the evidence is not as unanimous as it is in the racetrack. Inefficiencies, when detected, are driven mainly by the overbetting of longshot and underbetting of favorite outcomes. When betting with bookmakers, however, it is not clear whether the favorite-longshot bias is a result of the behavior of the demand-side agents (bettors)

or the supply-side agents (bookmakers). Traditional models of sportsbook pricing suggest that the bookmaker sets the odds in order to balance the amount of money wagered on the various outcomes of events and therefore the odds reflect the behavior of the bettors. However, other studies (e.g., Kuypers (2000), Levitt (2004)) suggest that bookmakers are more skilled than bettors at predicting the outcomes of matches and optimally set inefficient odds in order to exploit bettors' biases. Therefore, a test of market efficiency in this market is a joint test of what is the price-setting behavior of the bookmaker (or more accurately, what is the bookmaker behavior assumed by the bettors) and whether there is a bias.

Another bookmaker wagering market that has received a lot of attention in the literature is point spread betting, i.e., betting on the actual score difference between the participating teams rather than the outcome of the game. Deviations from market efficiency are found in most of the studies that examine this market but there is a considerable amount of variation in both the extent and the direction of the documented bias. In many cases (e.g., Gandar (1988), Woodland and Woodland (1994, 2001, 2003)), the opposite of the favorite-longshot bias has been observed, with favorites (longshots) winning less (more) frequently than implied by efficiency. To a smaller extent, a bias towards visiting teams has been reported, with bettors systematically underestimating the impact of home-field advantage. Finally, a few studies (e.g., Avery and Chevalier (1999)) find that sentiment-caused biases (e.g., hot hand bias, experts' opinion bias, etc.) affect the formation of the point spread resulting in market inefficiencies. Table 1 provides an overview of the results of some of the existing studies that examine market efficiency in various sports betting markets.<sup>2</sup>

Many of the biases documented in the aforementioned studies have a natural analogue observed in typical financial markets.<sup>3</sup> For instance, a regular favorite-longshot

---

<sup>2</sup>See Thaler and Ziemba (1988), Hausch and Ziemba (1995), Sauer (1998), and Jullien and Salanie (2008) for comprehensive surveys of this literature.

<sup>3</sup>See De Bondt and Thaler (1989) for a summary of the literature on stock market efficiency.

bias has been found in options pricing, with deep in-the-money (out-of-the-money) options being significantly underpriced (overpriced) (e.g., Rubinstein (1985, 2001), Shastri and Wethyavivorn (1987)). Furthermore, two of the most prominent financial market irregularities are i) the tendency of assets with good (poor) recent performance to continue overperforming (underperforming) in the near future (momentum) and ii) the tendency of assets that performed well (poorly) over a long period to subsequently underperform (overperform) in the future (reversal).<sup>4</sup> Behavioral finance (e.g., Barberis et al. (1998)) explains these phenomena assuming that agents react incorrectly to information signals, i.e., they consistently rely on patterns of past performance when assessing future prospects. The presence of sentimental investing (betting) can also be tested in sports betting markets (e.g., Durham et al. (2005)) by examining whether bettors significantly overbet teams based on their prior performance or react to specific patterns of past team performance, e.g., short versus long winning/losing streaks. Another phenomenon that has been observed in financial markets is the “home bias” (e.g., French and Poterba (1991)), i.e., investors tend to hold a disproportionately large share of their equity portfolio in stocks they are more familiar with, e.g., home-area stocks. A similar bias in sports gambling would manifest itself as an overbetting of home-area teams or overall popular teams.

With the favorite-longshot bias being the most widely studied inefficiency in sports betting markets, several theories have been proposed to explain it.<sup>5</sup> First, neoclassical theory suggests that local convexities in people’s utility functions may create a preference for risk, making people willing to accept a lower expected return when betting on longshots. This explanation was first proposed by Weitzman (1965), while Golec and Tamarkin (1998) later suggested that the observed bias could better be explained by skewness-loving (and risk-averse) betting behavior. The second expla-

---

<sup>4</sup>See, for example, De Bondt and Thaler (1985), Jegadeesh and Titman (1993, 2001), Schwert (2003).

<sup>5</sup>See Ottaviani and Sorensen (2008) for an overview of the main explanations that have been proposed in the literature.

nation is based on behavioral theories (e.g., prospect theory) which suggest that there is a psychological bias that leads people to overestimate (underestimate) small (large) probabilities and therefore overbet (underbet) longshot (favorite) outcomes. In fixed-odds betting markets, Shin (1991, 1992) proposes that the bias could also arise as an optimal response of an uninformed bookmaker who wants to insure himself against insiders. In the parimutuel system, Ali (1977) suggests that the bias could also be explained by bettors having heterogeneous beliefs with respect to the win probabilities. Unfortunately, with the aggregate price data on standard bets that are usually available to researchers, it is impossible to disentangle between the above explanations. But Snowberg and Wolfers (2010) use aggregate data on exotic bets and conclude that the probabilities misperception model better fits the behavior of a representative racetrack gambler than the risk-loving model.

### 3 Data Description

This study utilizes a unique dataset of individual betting activity in an online sportsbook provided to us by a large European bookmaker. The analysis includes 10,852 unique soccer matches of all major and most minor soccer leagues bet by 100 randomly selected online gamblers over a period of around 3.5 years (October 2005 - May 2009). We restrict our attention to the most popular soccer betting market in which bettors predict the final result of matches, i.e., bets are placed on either a home team win, a draw, or an away team win. Each of these outcomes is associated with a price that is quoted by the bookmaker and determines the payoff of a unit wager on that outcome. For example, odds 2 imply that the bettor will make a profit of \$1 for each \$1 staked. We have both the final odds that prevailed at the end of the betting period, and the odds at which each individual placed his bet; we use the former in the aggregate-level analysis, and the latter in the individual-level analysis.



The traditional model of sportsbook pricing suggests that bookmakers set the market-clearing prices, i.e., they attract wagers on the various outcomes of events in such proportions as to guarantee themselves a fixed profit (commission) regardless of the match outcome. The quoted odds on all outcomes of an event imply a probability with which each of these outcomes is expected to occur. For example, if the commission is zero, odds of 2 for a given outcome imply that the probability of this outcome occurring is simply the inverse of the odds, i.e.,  $1/2$  or 50%. Of course, the commission is always positive, hence under the market-clearing model, the sum of the inverses of the odds over all outcomes of an event invariably exceeds one, so clearly the implied probability for a given outcome is not simply the inverse of its odds. Instead, the “sum to one” implied probabilities are obtained by dividing the inverse odds of each outcome by their sum over all outcomes of the event. Formally, letting the bookmaker’s posted odds for the home win, the draw and the away win be denoted by  $o_H$ ,  $o_D$  and  $o_A$  respectively, the inverse odds of the respective outcomes are  $\frac{1}{o_H}$ ,  $\frac{1}{o_D}$  and  $\frac{1}{o_A}$ . If the book is balanced, the bookmaker maintains a commission of  $c = \left( \frac{1}{o_H} + \frac{1}{o_D} + \frac{1}{o_A} \right) - 1 > 0$  regardless of the match outcome. The odds-implied probabilities which sum to one are  $p_H = \frac{\frac{1}{o_H}}{\frac{1}{o_H} + \frac{1}{o_D} + \frac{1}{o_A}} = \frac{1}{o_H(1-c)}$  and likewise for  $p_A$  and  $p_D$ . Thus, in a balanced book, the expected value of each wager should be equal to the negative of the bookmaker’s commission. The average commission in the soccer betting market under study is 9% with a standard deviation of 2%.

In this study, we observe the following information for each bet placed by each of the bettors in our sample: i) the date of the bet, ii) the soccer match on which the bet was made (e.g., Premier League match between Manchester United and Liverpool that will take place on dd/mm/yy), iii) the outcome chosen (i.e., home win, draw, away win), iv) the bet amount, v) the odds at the time the bet was placed and vi) the bet result.<sup>6</sup> In addition to these, our dataset includes information about the gender, age,

---

<sup>6</sup>Note that various combinations of bets are allowed in the sportsbook and bettors often combine bets on various sports and events under various bet types (e.g., double, treble, etc.). In this study,

country of origin and zip code of the bettors as provided by them upon registration in the sportsbook. Our initial dataset did not contain the match outcomes. We therefore matched all soccer games included in our sample with historical match statistics available online.

Table 2 presents summary statistics of the final odds associated with the three outcomes of soccer events. The quoted odds range from 1.01 to 50 with the home team win having on average the most favorite odds (2.41) followed by the draw outcome (3.53) and the away team win (4.59). Table 3 presents summary statistics of the characteristics of the unit bets placed by the bettors in our sample.

## 4 Empirical Tests of Favorite-Longshot Bias

### 4.1 Aggregate-level Analysis

Market efficiency implies that the odds quoted by the bookmaker reflect all publicly available information related to the assessment of the match outcome probabilities. If the market is efficient and the bookmaker balances his book, then all wagering strategies would yield expected losses equal to the bookmaker’s commission. In this section we perform various tests of the weak-form efficiency of the soccer betting market under study.

The first test of market efficiency compares the objective win probabilities of observations whose odds fall within a specific range, with the market probabilities implied by the bookmaker’s odds. We first divide the observations pertaining to each of the three outcomes (home, draw, away) in  $K = 10$  odds groups such that we maximize between-group variation and minimize within-group variation subject to the constraint that all groups contain at least one “winning” observation. The groups

---

however, we focus on a specific submarket of the sportsbook, so we have extracted all bets on final outcomes of soccer matches from combination bets.

have an equal number of observations unless there are tied values at their boundaries which are all assigned to the same group. Letting  $N_k$  be the number of observations with odds that fall in group  $k$ , we compute the average subjective probability implied by the odds for odds group  $k$ ,  $p_k = (\frac{1}{N_k}) \sum_{j=1}^{N_k} p_{kj}$ , and the corresponding proportion of wins in group  $k$ ,  $\hat{\pi}_k = (\frac{1}{N_k}) \sum_{j=1}^{N_k} W_{kj}$ , where  $W_{kj} \in \{0, 1\}$  is a dummy indicating whether observation  $j$  in odds group  $k$  “won”. For sufficiently large  $N_k$ , the observed win proportion  $\hat{\pi}_k$ , approaches a normal distribution with mean equal to the objective win probability,  $\pi_k$ , and variance equal to  $\frac{\pi_k(1-\pi_k)}{N_k}$ . We therefore calculate the standardized test statistic  $z_k = \frac{\hat{\pi}_k - p_k}{\sqrt{\frac{p_k(1-p_k)}{N_k}}}$  and perform a number of  $z$ -tests to examine the significance of the difference between the subjective and the objective probabilities per odds group. Table 4 reports the results of the  $z$ -tests. The results provide a first indication that there are some inefficiencies in the quoted odds. In particular, we find that for favorites (with odds lower than 1.6) the objective win probability is significantly higher than the subjective win probability, while the reverse is true for longshots (with odds higher than 3.3). This indicates that favorites (longshots) have been significantly underbet (overbet), a phenomenon in line with the favorite-longshot bias that has been widely documented in racetrack betting.

The second test of the weak-form efficiency hypothesis that is often employed in the literature is based on regressions of match outcomes on the probabilities implied by the bookmaker’s odds, i.e.,

$$Result_j = \beta_0 + \beta_1 p_j + \varepsilon_j,$$

where  $Result_j = 1$  if observation  $j$  “won” and  $p_j$  is the win probability implied by the odds associated with it. Market efficiency is the null hypothesis  $H_0 : \{\beta_0, \beta_1\} = \{0, 1\}$ . Table 6 presents the estimates for this linear probability model, along with the results of the  $F$ -test of  $H_0$ . Similarly to the results of the first test, the null hypothesis of

market efficiency is rejected ( $F$ -statistic = 55.85 and  $p$ -value < .0001).<sup>7</sup>

As a final test, we take a look at the returns of betting on all observations of each odds group. Let  $R_k = \left(\frac{1}{N_k}\right) \sum_{j=1}^{N_k} [W_{kj}(o_{kj} - 1) + (1 - W_{kj})(-1)]$  be the average realized return from betting in odds group  $k$ , where  $o_{kj}$  are the odds of observation  $j$  in group  $k$ , and  $W_{kj} = 1$  if observation  $j$  in odds group  $k$  “won”. Figure 1 plots  $R_k$  for the 10 odds groups together with the 95% confidence intervals calculated using the bootstrap method. We also plot the negative of the average commission per odds group, which corresponds to the average expected return under the assumption that the bookmaker balances the book. The average return from betting on strong favorites (with odds less than 1.6) is  $-3\%$ , on mid-range odds (between 1.6 and 3.3) is between  $-6\%$  and  $-10\%$ , while betting on strong longshots (with odds higher than 3.3) yields a much lower rate of return of around  $-17\%$ . Since the expected return from betting on favorites exceeds the expected return from betting on longshots, the weak form of market efficiency fails.

We have therefore concluded that prices do not accurately reflect the true probabilities of match outcomes. It is important to note, however, that the true extent of the favorite-longshot bias also depends on how the bookmaker sets the odds. If the bookmaker sets the market-clearing odds, then these odds reflect solely bettors’ biased perceptions of the win probabilities, and thus the results above capture the full extent of the bias. If, however, as Levitt (2004) suggests, the bookmaker sets the profit-maximizing odds in order to exploit the bias himself (i.e., the final odds lie between the correct and the balancing odds), the true extent of the bias will be greater than reported in our previous tests.

Now, we perform the three tests of market efficiency for each outcome separately, i.e., for the home team win, the draw and the away team win. This is important in order to disentangle whether the bias is related to the odds associated with outcomes,

---

<sup>7</sup>The same conclusion is reached if we use a probit/logit or ordered probit/logit regression to examine the relationship between true probabilities and those implied by the odds.

the outcomes themselves or some interaction between the two. Table 5 presents the results of the  $z$ -tests, Table 6 presents the results of the regression analysis and Figure 2 plots the average realized returns per odds group and outcome. The joint hypothesis tests indicate that the null hypothesis of efficiency is rejected for all three outcomes. But a closer look at the  $z$ -tests shows that the objective probabilities are significantly higher than the corresponding subjective probabilities only for home and away favorites, while the opposite is true mainly for draw longshots. This observation suggests that the underbetting of favorites is particularly pronounced in home and away favorites, while the overbetting of longshots is very strong in draw longshots. To further examine what drives the inefficiency observed in the pooled data, we perform some additional regression tests.

In particular, we first estimate a model that explains the mean difference between true and implied win probabilities with outcome (i.e., home, draw, away) dummies, implied win probabilities, and interactions thereof. Table 7 presents the results of the various regression specifications employed. Specifications (1) through (5) present Ordinary Least Squares (OLS) estimates of several model variations of the form

$$Result_j - p_j = \mathbf{x}_j' \beta + \varepsilon_j,$$

where,  $Result_j$  and  $p_j$  are as before, and  $\mathbf{x}_j$  is the vector of the other aforementioned explanatory variables. In particular, specification (1) is a baseline that assumes that outcome dummies do but implied probabilities do not enter the model; specifications (2) and (4) assume that the relationship between the dependent variable and the implied probabilities is linear; while specifications (3) and (5) assume that this relationship is non-linear and is captured by a set of dummy variables indicating whether the observation belongs to a specific odds group. Specifications (4) and (5) also include interaction terms between the odds and the outcome indicator variables

to capture possible complementarities between the two. On the other hand, specifications (6)-(8) show the results of the corresponding semi-parametric regressions of the form

$$Result_j - p_j = f(p_j) + \tilde{\mathbf{x}}'_j \tilde{\beta} + \varepsilon_j,$$

where  $\tilde{\mathbf{x}}_j$  is like  $\mathbf{x}_j$  but obviously excludes the implied win probability,  $p_j$ , which is assumed to enter the regression non-linearly according to some arbitrary function  $f$ . These specifications are estimated using the difference estimator by Yatchew (1998), where the data is first sorted by  $p$ , then the nonparametric part is removed by differencing (of some order), and the resulting equation is estimated by OLS to find consistent estimates of  $\tilde{\beta}$ ; subsequently,  $f(\cdot)$  is estimated using the OLS estimates of  $\tilde{\beta}$  in place of  $\beta$  itself.

In all specifications, we find strong evidence of the favorite-longshot bias, with strong favorites being significantly underbet and all longshots being significantly overbet. In addition to this, there is some evidence of a bias towards draws and some evidence of a possible bias towards draw longshots. These results suggest that an additional bias towards draws might be amplifying the effect of the usual favorite-longshot bias, making the draw longshots significantly overbet, which is consistent with the evidence from the  $z$ -tests that we described earlier. A possible source of the bias towards the draw might be the difficulty to predict draws with any degree of reliability which creates greater disagreement on whether the observed odds deviate from the true probabilities on these outcomes, and results in an overbetting of the draw relative to the home win and the away win.

In view of the results above, we then turn our attention to whether profitable betting opportunities can arise by following some simple wagering rules. If the bias is large enough and the bookmaker does not (fully) exploit it himself, then a contrarian strategy of betting against public opinion could yield positive expected returns. Therefore, a natural starting point for our tests is to focus on matches in which the

draw outcome is a longshot and thus on average overbet, while either the home win or the away win have favorite odds which are on average underbet. We then calculate the return from betting i) on the home win in matches where the home win is favorite while the away win and the draw outcome have long odds (2,715 matches) and ii) on the away win in matches where the away win is favorite while the home win and the draw outcome have long odds (656 matches). The first strategy yields an average return of  $-2.65\%$ , while the second strategy yields a positive average return of  $0.13\%$ , which is significantly different from zero at the 5% level ( $p\text{-value} = 0.044$ ). In both cases, the average return is substantially higher than the returns usually observed in the sportsbook (see Table 5 for the average realized returns per odds groups and outcomes), indicating that bettors can significantly reduce their losses by following simple strategies designed to exploit the favorite-longshot bias. These strategies could also be marginally profitable, although the small positive return found here will likely vanish once taxes and other transaction costs are taken into account. We therefore conclude that simple strategies of betting against the public do not yield substantial positive returns and could only be marginally profitable in the best case, depending on the level of the various costs involved.

## 4.2 Individual-level Analysis

An important constraint of the datasets typically available in sports gambling is that they only contain information on prices and event attributes but not individual-level data on bets and bettors' characteristics. The lack of individual-level data has led most researchers to focus on the behavior of a representative agent. The common assumptions employed by these studies are i) that all bettors are identical, they place equally-sized bets, and are represented by a single agent, and ii) that the odds are in equilibrium in the sense that the representative bettor is indifferent among the various event outcomes. Under these assumptions, the documented favorite-longshot

phenomenon should appear in all events containing favorite and longshot outcomes and under the reasonable assumption that the bookmaker balances the book, everyone who bets on the favorites (longshots) should earn positive (negative) returns in excess of the bookmaker’s commission.

In this section, we use the individual-level data available to us to examine the extent to which bettors in our sample gain or suffer from the favorite-longshot bias. We first take a look at some summary statistics of our dataset. Figure 3 presents the distribution of the number of unit bets placed on strong longshot outcomes by the bettors in our sample.<sup>8</sup> We observe that the number of bettors who have placed more than 5 bets on strong longshot outcomes of soccer events is only 38. This first observation indicates that, contrary to the common perception that the majority of bettors exhibits the favorite-longshot bias, the majority of bettors in our sample have not even bet on these outcomes more than a handful of times. Thus, the percentage of bettors who exhibit the bias in our sample should be no more than 38.

If all matches that involve longshot outcomes exhibit the favorite-longshot phenomenon, as it is the case in a representative agent setting, then all bettors should exhibit the bias and earn negative excess returns from betting on longshots. We test this hypothesis by first calculating the average excess return from betting on strong longshots for each of the 38 bettors and then performing a series of individual-level hypothesis tests to find the number of bettors for whom the average excess return is significantly negative. We find that for only 6 of these bettors can the null hypothesis of zero excess returns be rejected.<sup>9</sup> We therefore conclude that only 6% of the bettors in our sample exhibit the favorite-longshot bias by overbetting the longshot outcomes. This finding also indicates that not all longshot odds in the sportsbook are inefficient, and the majority of bettors has indeed bet on the efficient longshots odds.

---

<sup>8</sup>The cutoff odds for defining strong longshots is the same as that used in the aggregate analysis.

<sup>9</sup>Note that there are no bettors that earn positive excess returns from betting on strong longshot outcomes.



On the other side of the wager, 74% of the bettors have placed more than 5 bets on strong favorite outcomes, while 16% of the bettors earn significant positive excess return from doing so. However, we should be careful when interpreting the latter percentage since people can choose to bet on favorite outcomes for different reasons. Some people might bet on favorites in order to consciously exploit the favorite-longshot bias, while others might do so solely for preference reasons but they also earn positive excess returns due to the existing bias in the quoted prices. It would be interesting to attempt to disentangle between these two groups of bettors in future research.

Having said that, we can more comfortably interpret as bettors who consciously exploit the bias, those who have earned significant positive average returns rather than all bettors who significantly beat the bookmaker’s commission. Hence, we repeat the multiple hypothesis testing analysis to find the number of bettors for whom the average realized return from all bets is significantly positive. Our results indicate that 2% of the bettors are likely to consciously exploit the bias.

## 5 Empirical Tests of Home Bias

It has been commonly observed in typical financial markets that investors tend to concentrate their equity investments at their domestic market, a phenomenon known as the “home bias”. A similar bias could appear in sports betting if, for example, bettors place a disproportionately large share of their bets on the teams that they are more familiar with, e.g., their home-area team. In this section, we test whether this bias appears in the sports betting market under study by calculating i) the proportion of wagers that bettors have placed on their home-area team and ii) the number of bettors who have overbet their home-area teams and are therefore earning significantly negative returns relative to the bookmaker’s commission.

Having information on bettors' zip codes allows us to identify the team(s) that are based on each bettor's home location. For international matches, we define the home-area team as the national team of the bettor's country of origin, and for domestic matches the team that is based in the bettor's area of residence (according to the bettor's zip code). Figure 4 shows the histogram of the percentage of total wagers placed by each individual on his/her home-area team. We find that for the vast majority of bettors this percentage is very close to zero (the median is equal to 0.64%), which provides a first indication that, under the odds quoted by the bookmaker, most bettors in our sample do not exhibit the "home bias". In particular, only 57 of the bettors have bet at least once on their home-area team.

We then examine whether the expected return that people get from betting on their home team is systematically lower than the bookmaker's commission, which would indicate that they have consistently bet on it at unfavorable odds. For each bettor  $i$ , we separately estimate the following regression:

$$ExcessReturn_{ij} = \beta_{0i} + \beta_{1i}Favorite_{ij} + \beta_{2i}HomeTeam_{ij} + \varepsilon_{ij},$$

where  $ExcessReturn_{ij}$  is the difference between the realized return and the expected return of bet  $j$  for bettor  $i$ ,  $Favorite_{ij}$  is a dummy that is equal to 1 if bet  $j$  of bettor  $i$  was placed at favorite odds, and  $HomeTeam_{ij}$  is a dummy that is equal to 1 if bet  $j$  is on the bettor  $i$ 's home-area team. We then carry out multiple hypothesis tests to find the number of bettors for whom  $\beta_{2i}$  is significantly negative. We find that  $\beta_{2i}$  is significantly negative for 3% of the bettors, indicating that there is a small percentage of bettors in our sample who have overbet their home-area team.

In addition to the above, we also examine whether people have placed a disproportionately large fraction of their bets on matches in which their home-area team is participating but not necessarily in favor of their home-area team. This could be the

case, for example, if bettors bet against their home-area team in order to hedge their utility loss in case this team loses, or if bettors have (or think they have) superior information about the specific event since it involves a team that they are familiar with. The histogram presented in Figure 5 shows that the percentage of unit wagers placed on events involving bettors' home-area team is still relatively small (equal to 3.85% for the median bettor) with 68 bettors having placed at least one of these bets. Furthermore, 4% of the bettors seem to have earned significantly negative excess return by consistently placing such bets at unfavorable odds. We therefore conclude that around 4% of the bettors may exhibit some bias related to their favorite team.

## 6 Conclusion

In this study we use both aggregate- and individual-level data from an online soccer wagering market to examine whether the prices set by bookmakers on a large number of soccer events are efficient, as well as the extent to which individual bettors are affected by possible inefficiencies or contribute to their formation. The findings suggest that the financial market under study violates the weak form of market efficiency and, in particular, in a way that is consistent with the favorite-longshot bias, i.e., the tendency of bettors to consistently underbet outcomes with short odds and overbet outcomes with long odds relative to their observed frequency of winning. Focusing on the prices associated with specific outcomes of soccer events, we find that the inefficiency is mainly driven by the underbetting of favorite odds on the home win and the away win, and an overbetting of longshot odds on the draw outcome. The analysis of individual-level returns indicates that contrary to the common perception that the majority (or totality) of bettors contributes to the generation of the favorite-longshot bias, only 6% of the bettors under study have systematically bet on biased odds associated with longshot outcomes, while 2% of the bettors earns significant

positive returns from betting on favorite outcomes. Finally, we find that around 4% of the bettors may suffer from a “home bias” which manifests itself as an overbetting of the home-area team for each individual. In future research, we plan to disentangle between the different explanations of the favorite-longshot bias suggested in the literature as well as examine whether other sentiment-caused biases (e.g., the tendency of bettors to overbet teams based on patterns in their past performance) appear in both the aggregate- and the individual-level data.

## References

- Ali, M. M., 1977. Probability and utility estimates for racetrack bettors. *The Journal of Political Economy* 85 (4), 803–815.
- Asch, P., Malkiel, B. G., Quandt, R. E., July 1982. Racetrack betting and informed behavior. *Journal of Financial Economics* 10 (2), 187–194.
- Avery, C., Chevalier, J., October 1999. Identifying investor sentiment from price paths: The case of football betting. *Journal of Business* 72 (4), 493–521.
- Barberis, N., Shleifer, A., Vishny, R., September 1998. A model of investor sentiment<sup>1</sup>. *Journal of Financial Economics* 49 (3), 307–343.
- Busche, K., Hall, C. D., July 1988. An exception to the risk preference anomaly. *Journal of Business* 61 (3), 337–46.
- De Bondt, W. F. M., Thaler, R., July 1985. Does the stock market overreact? *Journal of Finance* 40 (3), 793–805.
- De Bondt, W. F. M., Thaler, R. H., Winter 1989. A mean-reverting walk down wall street. *Journal of Economic Perspectives* 3 (1), 189–202.
- Durham, G. R., Hertzel, M. G., Martin, J. S., 2005. The market impact of trends and sequences in performance: New evidence. *The Journal of Finance* 60 (5), pp. 2551–2569.
- Fama, E. F., 1970. Efficient capital markets: A review of theory and empirical work. *The Journal of Finance* 25 (2), pp. 383–417.
- Figlewski, S., 1979. Subjective information and market efficiency in a betting market. *The Journal of Political Economy* 87 (1), 75–88.

- French, K. R., Poterba, J. M., May 1991. Investor diversification and international equity markets. *American Economic Review* 81 (2), 222–26.
- Gandar, John, e. a., September 1988. Testing rationality in the point spread betting market. *Journal of Finance* 43 (4), 995–1008.
- Golec, J., Tamarkin, M., 1998. Bettors love skewness, not risk, at the horse track. *The Journal of Political Economy* 106 (1), 205–225.
- Griffith, R. M., 1949. Odds adjustments by american horse-race bettors. *The American Journal of Psychology* 62 (2), pp. 290–294.
- Hausch, D. B., Ziemba, W. T., 1995. Efficiency of sports and lottery betting markets. In: Robert A. Jarrow, V. M., Ziemba, W. T. (Eds.), *Handbook of Operations Research and Management Science*. Vol. 9. North-Holland, Amsterdam, pp. 545–580.
- Hausch, D. B., Ziemba, W. T., Rubinstein, M., 1981. Efficiency of the market for racetrack betting. *Management Science* 27 (12), pp. 1435–1452.
- Jegadeesh, N., Titman, S., March 1993. Returns to buying winners and selling losers: Implications for stock market efficiency. *Journal of Finance* 48 (1), 65–91.
- Jegadeesh, N., Titman, S., 2001. Profitability of momentum strategies: An evaluation of alternative explanations. *Journal of Finance* 56 (2), 699–720.
- Jullien, B., Salanie, B., 2008. Empirical evidence on the preferences of racetrack bettors. In: Ziemba, W. T., Hausch, D. B. (Eds.), *Handbook of Sports and Lottery Markets*. North-Holland, Amsterdam, pp. 27–49.
- Kuypers, T., September 2000. Information and efficiency: An empirical study of a fixed odds betting market. *Applied Economics* 32 (11), 1353–63.

- Levitt, S. D., 2004. Why are gambling markets organised so differently from financial markets? *The Economic Journal* 114 (495), 223–246.
- Ottaviani, M., Sorensen, P., 2008. The favorite-longshot bias: An overview of the main explanations. In: Ziemba, W. T., Hausch, D. B. (Eds.), *Handbook of Sports and Lottery Markets*. North-Holland, Amsterdam, pp. 83–101.
- Rubinstein, A., 2001. Comments on the risk and time preferences in economics, mimeo, Tel Aviv University.
- Rubinstein, M., June 1985. Nonparametric tests of alternative option pricing models using all reported trades and quotes on the 30 most active cboe option classes from august 23, 1976 through august 31, 1978. *Journal of Finance* 40 (2), 455–80.
- Sauer, R. D., December 1998. The economics of wagering markets. *Journal of Economic Literature* 36 (4), 2021–2064.
- Schwert, G. W., 2003. Anomalies and market efficiency. In: Constantinides, G., Harris, M., Stulz, R. M. (Eds.), *Handbook of the Economics of Finance*. Vol. 1 of *Handbook of the Economics of Finance*. Elsevier, Ch. 15, pp. 939–974.
- Shastri, K., Wethyavivorn, K., 1987. The valuation of currency options for alternate stochastic processes. *Journal of Financial Research* 10, 283–293.
- Shin, H. S., September 1991. Optimal betting odds against insider traders. *Economic Journal* 101 (408), 1179–85.
- Shin, H. S., March 1992. Prices of state contingent claims with insider traders, and the favourite-longshot bias. *Economic Journal* 102 (411), 426–35.
- Snowberg, E., Wolfers, J., 08 2010. Explaining the favorite-long shot bias: Is it risk-love or misperceptions? *Journal of Political Economy* 118 (4), 723–746.

- Snyder, W. W., September 1978. Horse racing: Testing the efficient markets model. *Journal of Finance* 33 (4), 1109–18.
- Swidler, S., Shaw, R., 1995. Racetrack wagering and the “uninformed” bettor: A study of market efficiency. *The Quarterly Review of Economics and Finance* 35 (3), 305–314.
- Thaler, R. H., Ziemba, W. T., 1988. Anomalies: Parimutuel betting markets: Race-tracks and lotteries. *The Journal of Economic Perspectives* 2 (2), 161–174.
- Weitzman, M., 1965. Utility analysis and group behavior: An empirical study. *The Journal of Political Economy* 73 (1), 18–26.
- Woodland, L. M., Woodland, B. M., March 1994. Market efficiency and the favorite-longshot bias: The baseball betting market. *Journal of Finance* 49 (1), 269–79.
- Woodland, L. M., Woodland, B. M., April 2001. Market efficiency and profitable wagering in the national hockey league: Can bettors score on longshots? *Southern Economic Journal* 67 (4), 983–995.
- Woodland, L. M., Woodland, B. M., 04 2003. The reverse favourite-longshot bias and market efficiency in major league baseball: An update. *Bulletin of Economic Research* 55 (2), 113–123.
- Yatchew, A., 1998. Nonparametric regression techniques in economics. *Journal of Economic Literature* 57, 135–143.



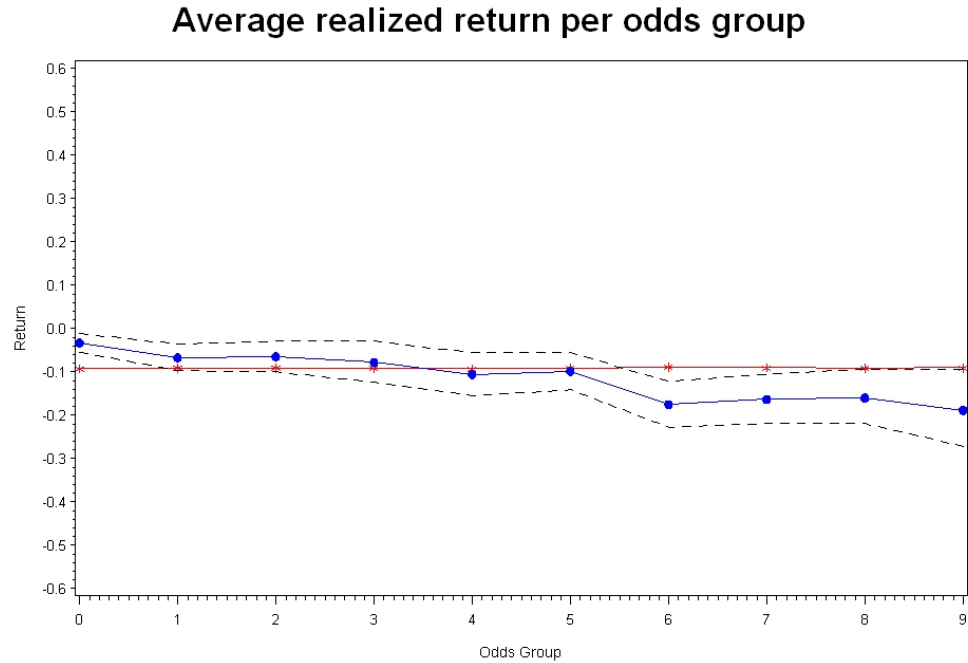


Figure 1: The figure shows the average realized return in each of the ten odds groups. The dashed lines delimit the 95% confidence intervals calculated using the bootstrap method. The (almost) horizontal line is the negative of the average commission of the bookmaker per odds group.

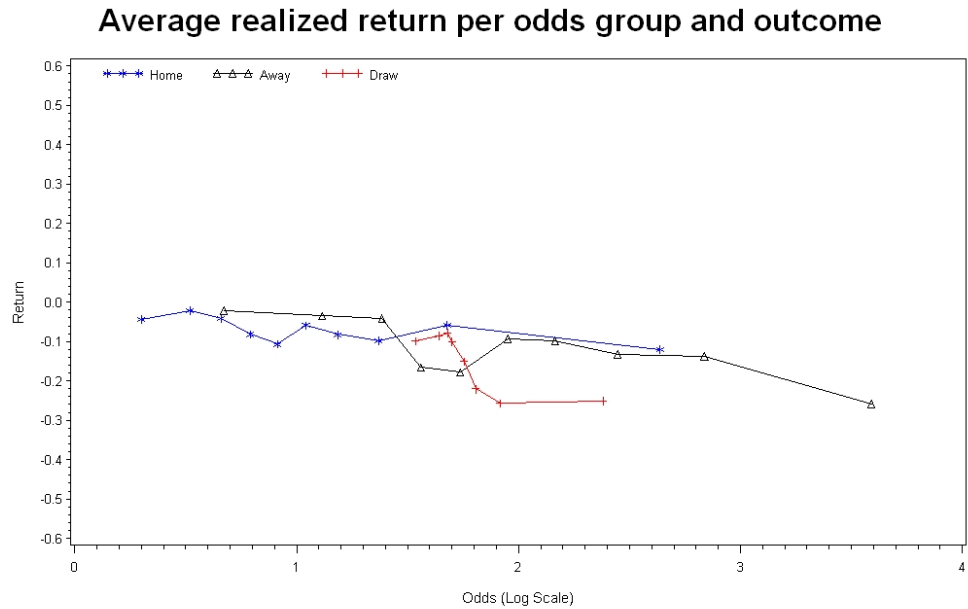


Figure 2: The figure shows the average realized return per odds group and outcome, i.e., home team win, draw and away team win.

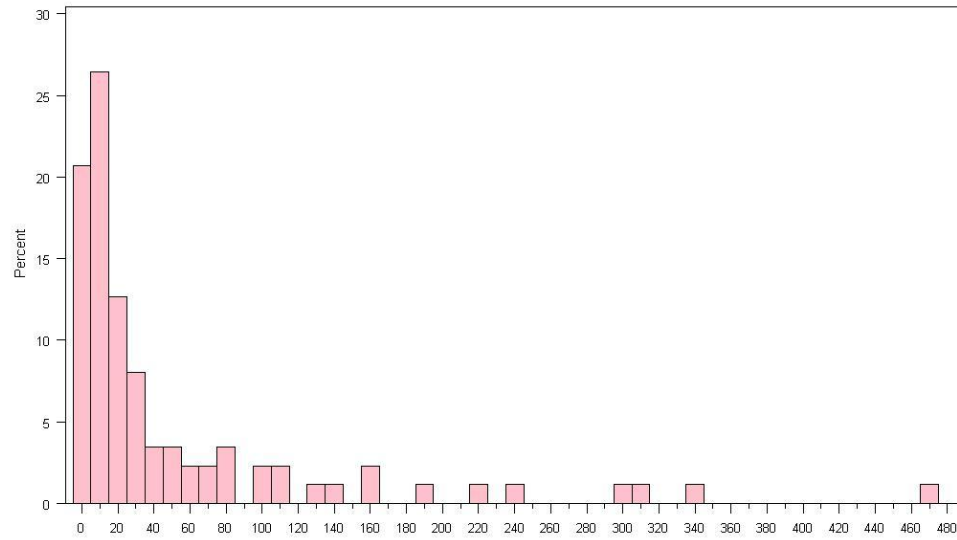


Figure 3: The figure shows the histogram of the number of bets placed by bettors on strong longshot outcomes. Strong longshot outcomes are defined as the outcomes with associated win probabilities smaller than 0.3.

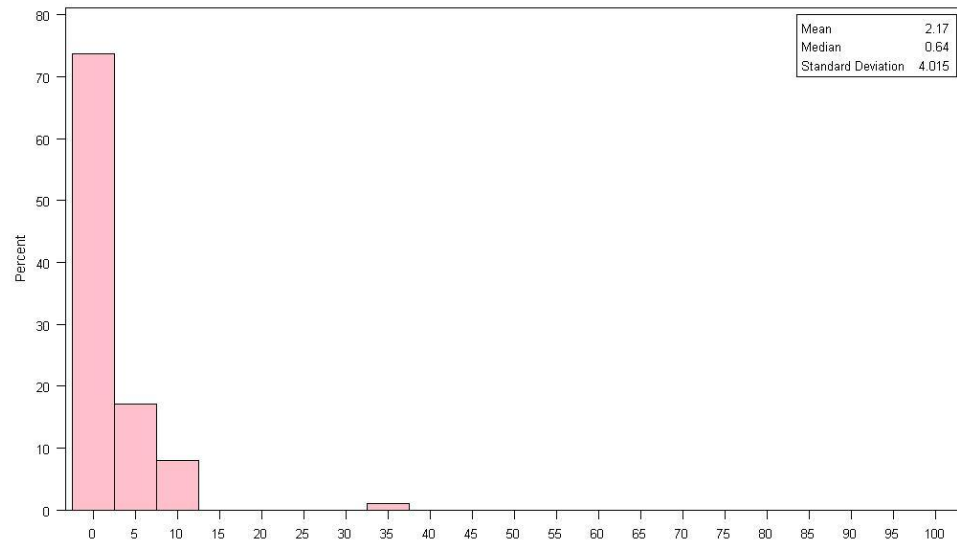


Figure 4: The figure shows the percentage of wagers that bettors have placed on their home-area team. The vertical axis represents the percentage of bettors associated with each bar.

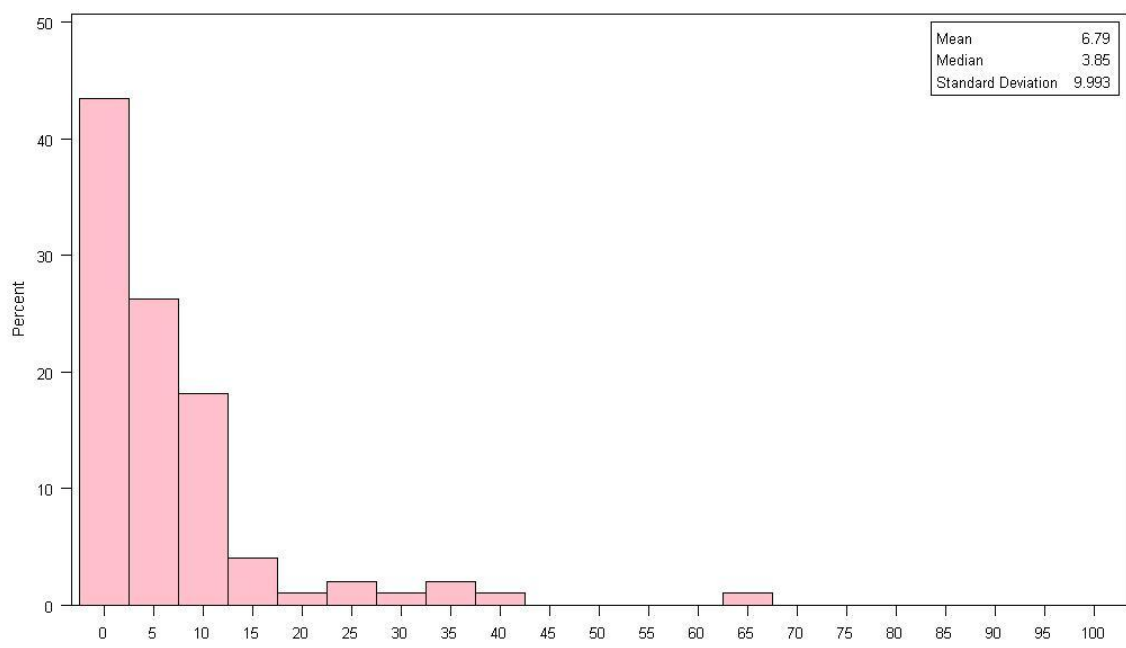


Figure 5: The figure shows the percentage of wagers that bettors have placed on matches that involve their home-area team. The vertical axis represents the percentage of bettors associated with each bar.

Table 1: Literature on the Efficiency of Sports Betting Markets

| Market                          | Study                        | Dataset       | Bias Type  | Positive Return         |
|---------------------------------|------------------------------|---------------|--|-------------------------|
| <u><i>Favorite-Longshot</i></u> |                              |               |  |                         |
| <b>Parimutuel</b>               | Ali (1977)                   | US racetrack  | FLB  | Yes                     |
|                                 | Asch et al. (1982)           | US racetrack  | FLB  | No                      |
|                                 | Snyder (1978)                | US racetrack  | FLB  | No                      |
|                                 | Hausch et al. (1981)         | US racetrack  | FLB  | Yes                     |
| <b>Fixed-odds</b>               | Dowie (1976)                 | UK racetrack  | No   | No                      |
|                                 | Pope and Peel (1990)         | UK soccer     | Yes  | No                      |
|                                 | Kuypers(2000)                | UK soccer     | FLB  | Yes                     |
|                                 | Cain et al. (2000)           | UK soccer     | FLB  | Yes                     |
|                                 | Jullien and Salanie (2000)   | UK racetrack  | FLB  | N/A                     |
| <b>Point spread</b>             | Zuber et al. (1985)          | NFL           | Yes  | Yes                     |
|                                 | Gandar et al. (1988)         | NFL           | Reverse FLB                                      | Yes                     |
|                                 | Sauer et al. (1988)          | NFL           | FLB  | Yes                     |
|                                 | Golec and Tamarkin (1991)    | NFL           | Reverse FLB                                      | Maybe                   |
|                                 | Woodland and Woodland (1994) | MLB           | Reverse FLB                                      | No                      |
|                                 | Woodland and Woodland (2003) | MLB           | Reverse FLB                                      | No                      |
|                                 | Dare and MacDonald (1995)    | NFL           | No   | No                      |
|                                 |                              | NCAA football | No   | No                      |
|                                 | Woodland and Woodland (2001) | NHL           | Reverse FLB                                      | Yes                     |
|                                 | Gandar et al. (2002)         | MLB           | No   | No                      |
| <u><i>Sentiment-caused</i></u>  |                              |               |  |                         |
|                                 | Camerer (1987)               | NBA           | Hot hand   | Yes                     |
|                                 | Brown and Sauer (1993)       | NBA           | Hot hand   | Yes                     |
|                                 | Avery and Chevalier (1999)   | NFL           | Experts' Opinion<br>Team Familiarity<br>Hot Hand | Maybe<br>Maybe<br>Maybe |
|                                 | Durham et al. (2005)         | NCAA football | Gambler's fallacy                                | N/A                     |
|                                 | Durham and Perry (2008)      | NCAA football | Experts' Opinion<br>Team Familiarity<br>Hot Hand | Maybe<br>Maybe<br>Maybe |

Table 2: Summary Statistics of Bookmaker's Odds

|                     | Mean Odd | St.Dev. | 25%  | Median | 75%  |
|---------------------|----------|---------|------|--------|------|
| <b>All outcomes</b> | 3.51     | 2.53    | 2.20 | 3.20   | 3.75 |
| <b>Home</b>         | 2.41     | 1.93    | 1.59 | 2.00   | 2.52 |
| <b>Draw</b>         | 3.53     | 0.80    | 3.20 | 3.25   | 3.60 |
| <b>Away</b>         | 4.59     | 3.54    | 2.60 | 3.60   | 5.50 |

Table 3: Summary Statistics of Bet Characteristics

This table presents summary statistics for the characteristics of 22,118 bets placed on the final outcomes of soccer matches.

|                              | Variable         | Mean   | St.Dev | Min  | Max  |
|------------------------------|------------------|--------|--------|------|------|
| <b>Match Outcome Dummies</b> | Home             | 0.54   | 0.50   | 0    | 1    |
|                              | Away             | 0.25   | 0.44   | 0    | 1    |
|                              | Draw             | 0.20   | 0.40   | 0    | 1    |
| <b>Odds Category Dummies</b> | Stong Favorites  | 0.21   | 0.41   | 0    | 1    |
|                              | Favorites        | 0.43   | 0.50   | 0    | 1    |
|                              | Longshots        | 0.34   | 0.47   | 0    | 1    |
|                              | Strong Longshots | 0.02   | 0.14   | 0    | 1    |
| <b>Event Characteristics</b> | Commission       | 0.09   | 0.02   | 0.02 | 0.16 |
|                              | Days to event    | 0.33   | 0.75   | 0    | 6    |
| <b>Number of bets</b>        |                  | 235.27 | 424.31 | 2    | 2887 |

Table 4: Objective versus Implied Win Probabilities – One-way

This table presents a comparison of the objective and implied win probabilities, per odds group.

| Group | <i>N</i> | Lo Odd | Hi Odd | Mean Implied<br>Probability | Mean Actual<br>Probability | z-test   | Mean Realized<br>Return |
|-------|----------|--------|--------|-----------------------------|----------------------------|----------|-------------------------|
| 0     | 3252     | 1.010  | 1.610  | 0.661                       | 0.704                      | 5.199**  | -3.323%                 |
| 1     | 3265     | 1.620  | 1.990  | 0.514                       | 0.528                      | 1.620    | -6.580%                 |
| 2     | 3452     | 2.000  | 2.380  | 0.418                       | 0.430                      | 1.400    | -6.570%                 |
| 3     | 2643     | 2.390  | 2.870  | 0.346                       | 0.351                      | 0.584    | -7.824%                 |
| 4     | 3051     | 2.880  | 3.190  | 0.304                       | 0.300                      | -0.580   | -10.742%                |
| 5     | 4281     | 3.200  | 3.250  | 0.282                       | 0.280                      | -0.199   | -9.592%                 |
| 6     | 2993     | 3.260  | 3.500  | 0.266                       | 0.241                      | -3.158** | -17.598%                |
| 7     | 3110     | 3.510  | 4.260  | 0.239                       | 0.220                      | -2.499** | -16.352%                |
| 8     | 3290     | 4.270  | 5.500  | 0.190                       | 0.174                      | -2.273** | -16.259%                |
| 9     | 3219     | 5.510  | 81.000 | 0.115                       | 0.104                      | -1.887*  | -18.778%                |

Table 5: Objective versus Implied Win Probabilities – Two-way

This table presents a comparison of the objective and implied win probabilities, per odds group and outcome.

**HOME WIN**

| Group | <i>N</i> | Lo Odd | Hi Odd | Mean Implied Probability | Mean Actual Probability | z-test  | Mean Realized Return |
|-------|----------|--------|--------|--------------------------|-------------------------|---------|----------------------|
| 0     | 1086     | 1.010  | 1.340  | 0.739                    | 0.780                   | 3.085** | -4.331%              |
| 1     | 1043     | 1.350  | 1.520  | 0.635                    | 0.685                   | 3.354** | -2.105%              |
| 2     | 979      | 1.530  | 1.660  | 0.575                    | 0.608                   | 2.067** | -4.029%              |
| 3     | 1230     | 1.670  | 1.810  | 0.526                    | 0.532                   | 0.423   | -8.159%              |
| 4     | 955      | 1.820  | 1.990  | 0.482                    | 0.473                   | -0.554  | -10.637%             |
| 5     | 1157     | 2.000  | 2.190  | 0.442                    | 0.457                   | 1.073   | -5.900%              |
| 6     | 1226     | 2.200  | 2.380  | 0.400                    | 0.405                   | 0.310   | -8.207%              |
| 7     | 1058     | 2.390  | 2.750  | 0.353                    | 0.350                   | -0.193  | -9.821%              |
| 8     | 1061     | 2.760  | 3.750  | 0.287                    | 0.297                   | 0.737   | -5.772%              |
| 9     | 1057     | 3.770  | 67.000 | 0.169                    | 0.157                   | -1.006  | -11.976%             |

**AWAY WIN**

| Group | <i>N</i> | Lo Odd | Hi Odd | Mean Implied Probability | Mean Actual Probability | z-test  | Mean Realized Return |
|-------|----------|--------|--------|--------------------------|-------------------------|---------|----------------------|
| 0     | 1045     | 1.010  | 1.900  | 0.580                    | 0.624                   | 2.877** | -2.060%              |
| 1     | 1218     | 1.910  | 2.380  | 0.422                    | 0.448                   | 1.862*  | -3.468%              |
| 2     | 1085     | 2.390  | 2.750  | 0.349                    | 0.368                   | 1.309   | -4.128%              |
| 3     | 936      | 2.760  | 3.190  | 0.308                    | 0.283                   | -1.651* | -16.528%             |
| 4     | 1064     | 3.200  | 3.590  | 0.273                    | 0.246                   | -1.938* | -17.805%             |
| 5     | 1162     | 3.600  | 4.310  | 0.236                    | 0.236                   | 0.002   | -9.252%              |
| 6     | 916      | 4.330  | 4.990  | 0.203                    | 0.201                   | -0.145  | -9.781%              |
| 7     | 1336     | 5.000  | 6.000  | 0.168                    | 0.160                   | -0.731  | -13.236%             |
| 8     | 957      | 6.060  | 8.460  | 0.128                    | 0.121                   | -0.602  | -13.768%             |
| 9     | 1133     | 8.500  | 81.000 | 0.083                    | 0.070                   | -1.635  | -25.781%             |

**DRAW**

| Group | <i>N</i> | Lo Odd | Hi Odd | Mean Implied Probability | Mean Actual Probability | z-test   | Mean Realized Return |
|-------|----------|--------|--------|--------------------------|-------------------------|----------|----------------------|
| 0     | 1680     | 1.670  | 3.000  | 0.314                    | 0.312                   | -0.151   | -9.910%              |
| 1     | 470      | 3.010  | 3.190  | 0.289                    | 0.294                   | 0.198    | -8.509%              |
| 2     | 1744     | 3.200  | 3.240  | 0.284                    | 0.288                   | 0.369    | -7.842%              |
| 3     | 1878     | 3.250  | 3.250  | 0.279                    | 0.277                   | -0.242   | -10.011%             |
| 4     | 1380     | 3.260  | 3.400  | 0.270                    | 0.251                   | -1.591   | -15.104%             |
| 5     | 893      | 3.410  | 3.590  | 0.260                    | 0.223                   | -2.526** | -22.037%             |
| 6     | 1451     | 3.600  | 4.170  | 0.242                    | 0.198                   | -3.898** | -25.652%             |
| 7     | 1356     | 4.180  | 15.000 | 0.181                    | 0.152                   | -2.765** | -25.029%             |

Table 6: Linear Probability Model of match outcomes

Estimation of the linear probability model of match outcomes against the implied bookmaker's probabilities. The dependent variables are dummies indicating whether the respective outcome occurred. *Win Probability* is the outcome's win probability implied by its odds. The associated t-statistics are reported below the estimates. The F-statistics test the joint hypothesis that the intercept is equal to 0 and the slope coefficient equal to 1. All standard errors are corrected for heteroskedasticity using White's correction. In the pooled regression, standard errors are clustered at the match level (to allow for correlation of the error within the triple of observations for each match). \* /\*\* /\*\*\* indicate significance at the 10% /5% /1% levels.

|                 | ALL OUTCOMES         | HOME WIN             | AWAY WIN             | DRAW                 |
|-----------------|----------------------|----------------------|----------------------|----------------------|
|                 | Result               | Result               | Result               | Result               |
| Intercept       | -0.036***<br>(-5.73) | -0.031***<br>(-2.57) | -0.029***<br>(-3.86) | -0.091***<br>(-3.88) |
| Win Probability | 1.109***<br>(58.19)  | 1.096***<br>(44.75)  | 1.111***<br>(40.05)  | 1.288***<br>(14.3)   |
| Adj. $R^2$      | 0.133                | 0.125                | 0.128                | 0.015                |
| $F$ -test       | 55.85<br><.0001      | 23.58<br><.0001      | 16.72<br>0.000       | 28<br><.0001         |



Table 7: Actual versus Implies Probability

The  $N = 32,556$  observations correspond to the 3 outcomes (home, draw, away) of the 10,852 matches that were wagered on by at least one individual in our sample. The dependent variable is the result (a dummy equaling 1 if the outcome occurs) less the probability of the outcome implied by its odds. *Home* and *Draw* are dummies indicating the outcome, and *Strong Favorite*, *Longshot*, and *Strong Longshot* are dummies indicating outcomes with odds shorter than 1.5, between 3.5 and 10, and longer than 10, respectively. *Win Probability* is the outcome's win probability implied by its odds. The remaining explanatory variables are interaction terms of the aforementioned variables.  $t$ -statistics using standard errors clustered at the match level (to allow for correlation of the error within the triple of observations for each match) are reported in parentheses. \* / \*\* / \*\*\* indicate significance at the 10% / 5% / 1% levels.

|                        | OLS                   |                       |                       |                       |                      | Partial Linear      |                   |                   |
|------------------------|-----------------------|-----------------------|-----------------------|-----------------------|----------------------|---------------------|-------------------|-------------------|
|                        | (1)                   | (2)                   | (3)                   | (4)                   | (5)                  | (6)                 | (7)               | (8)               |
| Home                   | 0.012 **<br>(2.03)    | -0.008<br>(-1.12)     | 0.002<br>(0.30)       | -0.002<br>(-0.11)     | 0.001<br>(0.17)      | -0.010<br>(-1.10)   | 0.007<br>(0.34)   | -0.012<br>(-1.04) |
| Draw                   | -0.016 ***<br>(-2.80) | -0.015 ***<br>(-2.64) | -0.019 ***<br>(-3.21) | -0.062 **<br>(-2.51)  | -0.011<br>(-1.24)    | -0.020 *<br>(-1.91) | -0.040<br>(-0.93) | -0.008<br>(-0.54) |
| Win Probability        |                       | 0.109 ***<br>(6.07)   |                       | 0.111 ***<br>(4.01)   |                      |                     |                   |                   |
| Home $\times$ Win Prob |                       |                       |                       | -0.015<br>(-0.40)     |                      |                     | -0.047<br>(-0.87) |                   |
| Draw $\times$ Win Prob |                       |                       |                       | 0.177 *<br>(1.87)     |                      |                     | 0.083<br>(0.51)   |                   |
| Strong Favorite        |                       |                       | 0.034 ***<br>(3.27)   |                       | 0.048 *<br>(1.77)    |                     |                   |                   |
| Longshot               |                       |                       | -0.021 ***<br>(-4.00) |                       | -0.015 *<br>(-1.69)  |                     |                   |                   |
| Strong Longshot        |                       |                       | -0.027 ***<br>(-2.61) |                       | -0.027 **<br>(-2.37) |                     |                   |                   |
| Home Strong Favorite   |                       |                       |                       |                       | -0.015<br>(-0.52)    |                     |                   | -0.021<br>(-0.67) |
| Home Longshot          |                       |                       |                       |                       | -0.006<br>(-0.39)    |                     |                   | 0.011<br>(0.53)   |
| Home Strong Longshot   |                       |                       |                       |                       | 0.040<br>(1.01)      |                     |                   | 0.062<br>(0.92)   |
| Draw Longshot          |                       |                       |                       |                       | -0.013<br>(-1.07)    |                     |                   | -0.020<br>(-1.05) |
| Constant               | 0.001<br>(0.32)       | -0.029 ***<br>(-4.97) | 0.013 **<br>(2.33)    | -0.029 ***<br>(-3.86) | 0.009<br>(1.26)      |                     |                   |                   |
| $R^2$                  | 0.0007                | 0.0017                | 0.0016                | 0.0018                | 0.0017               |                     |                   |                   |