

The Favorite-Longshot Bias: Efficiency of Prices in Fixed-Odds Betting Markets

Edvin Ahlander (23826)

Abstract

The statistical weak form efficiency in the fixed-odds betting market for English Premier League football matches is evaluated, applying both a multinomial logit regression model and four different sets of grouped z-tests, using a data set containing 119481 odds from 11 bookmakers over a period of 12 years. The empirical results show that the aggregate betting market is statistically weak form inefficient, implying that prices of odds do not contain all available information on historical prices and therefore are biased predictors of actual match outcomes. Moreover, evidence for the so-called ‘favorite-longshot bias’, i.e. the special case of inefficiency where high probability outcomes are undervalued and low probability outcomes are overvalued, is found only for away win odds while efficiency cannot be rejected for home win and draw odds. The results are explained with that low probability away wins are overestimated to a larger extent as they on average are less likely to occur than other low probability outcomes, resulting in systematic mispricings of away win odds.

Keywords: market efficiency, betting, decision under risk, logit model, behavioral economics

JEL: C25, D81, D91, G14, L83

Supervisor:	Magnus Johannesson
Date submitted:	May 13, 2019
Date examined:	June 5, 2019
Discussants:	Elvira Anderson, Lukas Källén
Examiner:	Andreea Enache

Acknowledgements

I would like to express my sincerest gratitude to my supervisor Magnus Johannesson for extensive support and invaluable guidance throughout the process. I would also like to thank friends and family for insightful comments on earlier drafts. Any remaining shortcomings are my own.

Table of Contents

1. Introduction.....	4
2. Background.....	7
2.1. Characteristics of Betting Markets.....	7
2.2. The Sports Betting Industry.....	7
2.3. How Odds Are Determined.....	9
2.4. Market Efficiency in Sports Betting Markets.....	9
2.5. The Favorite-Longshot Bias and Market Efficiency.....	10
3. Literature Review.....	12
3.1. Early Evidence for the Favorite-Longshot Bias.....	12
3.2. Findings Across Different Betting Markets.....	12
3.3. Market Efficiency and the Favorite-Longshot Bias in Football.....	14
3.4. Explanations for the Favorite-Longshot Bias.....	17
3.5. Contribution to Research.....	18
4. Data.....	19
5. Method.....	20
5.1. Obtaining implicit probabilities from bookmakers' odds.....	20
5.2. Multinomial Logit Model.....	21
5.3. Grouped Z-Tests.....	24
5.4. Hypotheses.....	27
6. Results.....	29
6.1. Multinomial Logit Model.....	29
6.2. Grouped Z-Tests.....	30
7. Discussion.....	39
7.1. Interpretation and Implications of Results.....	39
7.2. Limitations of Study.....	41
8. Concluding Remarks.....	42
9. Suggestions for Further Research.....	44
10. References.....	45

1. Introduction

Ever since Griffith (1949) found that horse-race betting odds seemed to overvalue horses with low probabilities of winning, and at the same time undervalue horses with high probabilities of winning, the pricing of odds in various betting markets has been given lots of consideration in the academic literature. As pointed out by Thaler and Ziemba (1988), betting markets may be better suited for testing rationality and market efficiency than other markets, such as financial markets. The reason is that each asset in a betting market has a well-defined termination point where its value becomes certain, i.e. when the outcome is revealed and the bet is either won or lost. Even though it could be argued that this repeated feedback procedure will result in that betting markets are more efficient than other markets, research has repeatedly uncovered systematic mispricings of odds, indicating inefficiencies. The anomaly seen most consistently, initially found by Griffith (1949), is the so-called *favorite-longshot bias* which is defined as bettors' tendency to overvalue low probability events and undervalue high probability events, resulting in a corresponding mispricing of the odds representing these events.

Horse racing betting in general, and pari-mutuel horse racing betting (i.e. that odds are determined by the fraction of bets made on a particular event relative to the entire pool of money, see e.g. Dowie, 1976; Sauer, 1998) in particular, is the betting market that has been studied most extensively in relation to market efficiency. Although some evidences for deviations have been found, most conclusions point to the fact that these betting markets suffer from a favorite-longshot bias (see e.g. Griffith, 1949; McGlothlin, 1956; Ali, 1977; Snyder, 1978; Asch *et al.*, 1982). More recently, studies of betting market efficiency and the favorite-longshot bias have been conducted for other sports such as American Football, Baseball, Ice Hockey, Basketball, Tennis, Australian Rules Football, Golf, Greyhound Racing, Boxing, Cricket and Association Football (henceforth referred to as 'football'). As regards to the betting market for football match outcomes, market efficiency tests have come to rather ambiguous conclusions as evidence both for the favorite-longshot bias (see e.g. Cain *et al.*, 2000; Vlastakis *et al.*, 2009; Koning, 2012; Direr, 2013; Nyberg, 2014), other forms of biases (such as the 'reverse' favorite-longshot bias, see e.g. Dixon and Pope, 2004; Deschamps and Gergaud, 2007) and efficient markets (see e.g. Kuypers, 2000; Cain *et al.*, 2003; Forrest *et al.*, 2005) has been found.

Betting market efficiency and the favorite-longshot bias can be tested in different ways. Applying the efficient market hypothesis (Fama, 1970) to the betting market setting, efficiency is evaluated as whether odds fully reflect all available information. More specifically, three separate forms of efficiency can be evaluated by distinguishing between different types of information, i.e. weak form efficiency (information on historical prices), semi-strong form efficiency (publicly

available information) and strong form efficiency (all available information including insider information). In the previous literature, evaluations of all efficiency forms in sports betting markets have been made, with various results. Following the terminology adopted for instance by Gandar *et al.* (1988) and Gray and Gray (1997), betting market efficiency tests can furthermore be categorized as either statistical or economic. Statistical tests refer to statistical models which aim to test whether prices of odds contain all information available in the market. Economic tests, on the other hand, refer to sophisticated betting algorithms which aim to expose inefficiencies through development of profitable betting strategies.

The purpose of this study is to investigate the statistical weak form market efficiency as well as the favorite-longshot bias in the fixed-odds (i.e. that odds are determined by bookmakers and are fixed once a bet is placed, see e.g. Nyberg, 2014) betting market for football matches. In other words, the aim is to examine whether odds prices in football betting markets fully reflect all information on historical prices and thus are unbiased predictors of actual match outcomes. If it appears to be a systematic difference between odds and the observed match outcomes, the aim is furthermore to investigate whether it corresponds to the favorite-longshot bias, i.e. that odds underestimate the probability of favorite outcomes and overestimate the probability of longshot outcomes. This is done by performing (1) a multinomial logit regression which aim to test the aggregate efficiency in this particular betting market and (2) several grouped z-tests which allow tests across specific odds intervals.

The data used in this study consists of odds for match outcomes in 4560 English Premier League matches between 2006 and 2018 from 11 different bookmakers. The English Premier League is chosen as it is the world's largest football competition by revenue as well as it constitutes one of the largest sports betting markets. Furthermore, considering the recent changes in sports betting markets worldwide, which has mainly been driven by the emergence and increasing popularity of online betting, resulting in growing betting volumes and possibly reshaped bettor behavior patterns, it is particularly favorable to use data including matches played recently. Regarding the tests, it should be noted that both the multinomial logit regression and the z-tests fall under the 'statistical' tests category as they do not concern profitable betting strategies. The reason that these tests are chosen in favor of economic tests is that systematic and robust profitable betting strategies not depending on unrealistic assumptions in football betting markets rarely have been found (Pope and Peel, 1989; Cain *et al.*, 2000; Kuypers, 2000; Goddard and Asimakopoulous, 2004; Deschamps and Gergaud, 2007, see discussion in Direr, 2013). This is due both to bookmaker margins and increased convergence of odds across bookmakers, where the latter is a result of intensified competition and the use of more sophisticated statistical models to predict

match outcomes (Deschamps and Gergaud, 2007). Furthermore, weak form market efficiency in particular is evaluated because of two main reasons. Firstly, it has been shown that bookmakers are more accurate, or at least as accurate, in predicting match outcomes based on publicly available information than the typical bettor (Levitt, 2004). These tendencies have additionally become even more apparent as intensified competition has increased the costs of inaccurate forecasting (Forrest *et al.*, 2005). Thus, even if a betting market would be semi-strong form inefficient, bookmakers would still make better use of publicly available information than most other market participants and, consequently, such inefficiencies are rather unlikely to expose. Secondly, insider betting based on private information seem to be of little or no value for insider bettors as bookmakers adjust odds in the presence of insiders so that bookmaker margins correlate with the degree of insider betting (Shin, 1992, 1993; Cain *et al.*, 2003) Furthermore, it is unlikely that there is a substantial amount of insider information in football anyway (Kuypers, 2000). This makes tests of strong form market efficiency less relevant to perform.

Based on above, two separate research questions are formulated,

1. *Is the fixed-odds betting market for English Premier League football matches statistically weak form efficient?*
2. *Does this betting market suffer from a favorite-longshot bias, i.e. are low probability outcomes overvalued and high probability outcomes undervalued?,*

where a demonstrated favorite-longshot bias (Q2) implies a statistically weak form inefficient market (Q1), but a statistically weak form inefficient market does not imply a favorite-longshot bias (i.e. other biases may be present). The favorite-longshot bias can therefore be seen as a special case of an inefficient market.

The remainder of the study is structured as follows. Section 2 provides relevant background information on betting markets, the sports betting industry and how market efficiency can be approached in these markets. Section 3 reviews existing literature concerning market efficiency and the favorite-longshot bias in sports betting markets. Section 4 describes the data set used in this study. Section 5 lays out the methodological framework that is applied to answer the research questions. Section 6 presents the results obtained from the tests. Section 7 provides interpretation and discussion of the results while section 8 concludes. Lastly, section 9 provides suggestions for further research.

2. Background

2.1. Characteristics of Betting Markets

Betting markets involve participants whom try to make predictions about the future. To do so, they invest in assets, also known as odds, which are priced according to the estimated likelihood that the corresponding event will occur. If the event occurs, the bettor will earn the invested money as well as some profit. If the event does not occur, the invested money will typically be lost. The return will be determined by the estimated likelihood, i.e. if participants place bets on an unlikely event which happens to occur the return will be larger than if participants predict more likely events correctly.

Betting markets have similarities with other markets, particularly financial markets. For instance, both markets involve a large number of investors/bettors with heterogenous beliefs whom have access to widely available information. Betting as well as trading is a zero-sum game with one agent on each side of the transaction, with potentially large amounts of money at stake (Levitt, 2004). Furthermore, returns on investments cannot be guaranteed and therefore imply uncertainty for investors. The uncertainty is typically related to potential payoffs, thus, both markets involve the universal risk-return trade-off (Gabriel and Marsden, 1990). In addition, it has been showed that bets and specific financial assets, such as derivatives, have direct similarities (see e.g. Ruhm, 2003; Hodges *et al.*, 2003; Vecer *et al.*, 2006).

Obviously, there are several differences between these markets as well, such as that prices in financial markets change frequently to equilibrate supply and demand while prices in certain betting markets are determined by a bookmaker after which adjustments normally are small and where the bookmaker faces substantial risk if that price is not the equilibrium price (Levitt, 2004). Furthermore, odds in betting markets have a well-defined termination point where its value becomes certain. In contrast, financial assets such as stocks are infinitely lived where their value depends both on the present value of upcoming cash flows and on the price the investor will pay for the stock in the future. Consequently, it is difficult to test for market efficiency and rationality in financial markets and, instead, betting markets may be better suited for such tests (Thaler and Ziemba, 1988).

2.2. The Sports Betting Industry

During recent years, the overall betting industry has undergone significant changes. This has been led by several fundamental industry transformations mainly related to the emergence of online betting. Firstly, cheap access to fast internet and various mobile devices have introduced a wide range of new gambling opportunities, often allowing faster and more convenient ways of placing

bets (Wood and Williams, 2011; Gainsbury *et al.*, 2012). Furthermore, the internet has increased the amount of information available to consumers, for instance providing instructions, tutorials and recommendations, allowing gamblers to feel informed (Gainsbury *et al.*, 2012). Lastly, the legislative reality of betting markets obstructs individual states' ability to gain control of the online betting activity. The main reason is that states cannot prevent offshore firms from offering their services onshore, allowing online betting providers to operate from all over the world (Beem and Mikler, 2011; Gainsbury, 2015). The emergence of online betting is likely to have influenced the behavior of bettors in several different ways. For instance, it has been showed that online bettors tend to gamble more frequently, spend more money (Kairouz *et al.*, 2012) as well as commit more errors and thus potentially engage in a more harmful gambling behavior than non-online bettors (Montes and Weatherly, 2017). Furthermore, incentives for bookmakers to set odds that reflect accurate predictions of actual match outcomes are likely to have increased in this new betting market reality, as higher levels of competition and more accessible information has increased potential losses from inaccurate predictions (Forrest *et al.*, 2005).

As a result of the emergence of online betting, the total gross gambling yield (GGY, i.e. the amount wagered minus return to bettors) has steadily increased during the last couple of years, reaching \$110.67 billion for Europe and \$480 billion globally in 2018. Furthermore, the overall online betting market grew at a pace of around 10% between 2015 and 2018 and is expected to reach a market value of \$94.4 billion in 2024, in contrast to a market value of \$45.8 billion in 2017. Sports betting accounted for 48% of all online gambling market activity worldwide in 2018 (Statista, 2019a). In the U.K., the number of active online gambling customer accounts has increased by roughly 48% since 2013 and reached 35.59 million in 2018. The same year, the number of new online gambling player registrations in the U.K. was 36.59 million. In comparison, this number averaged around 4 million between 2008 and 2013. Lastly, the value of funds held in online gambling customer accounts increased from £281.02 million in 2008 to £784.78 million in 2018 (Statista, 2019b).

Football is the largest sports betting market in the U.K., accounting for 45.9% of total betting GGY in 2018 (U.K. Gambling Commission, 2018). Furthermore, the English Premier League, the top tier of the English football league system contested by in total 20 teams from England and Wales, is the largest football competition worldwide accounting for around 20% of all football market revenues in 2018. Premier League revenues have increased from around €1000 million during the 1998/1999 season to nearly €6000 million for the 2018/ 2019 season (Statista, 2019c). As the largest football competition, the betting market for Premier League matches is as

well one of the world's most considerable betting markets. On average, £2.7 million is bet on each Premier League match (Deutscher *et al.*, 2019).

2.3. How Odds Are Determined

Broadly speaking, there are two distinct forms of betting, (1) pari-mutuel betting and (2) fixed-odds betting, which apply different odds-setting procedures.

In pari-mutuel betting, which is exclusively used in racetrack sports (e.g. horse- and greyhound racing) betting markets in North America, France, Hong Kong and Japan, a pool of all bets on a given race is collected. Then, a predetermined percentage of the betting pool is withdrawn by the bookmaker to cover for costs and profits. The remainder is paid out to the winning bettors in proportion to their individual bets. This implies that odds are fluctuating right until the start of the race and, consequently, payoffs will be heavily reduced when large amounts of money are placed on a given outcome (Sauer, 1998).

In contrast, odds in fixed-odds betting markets are determined by a bookmaker at which they accept bets from the public (Ottaviani and Sørensen, 2005). Odds are initially set several days before the event takes place and once a bet is placed, i.e. when the bookmaker and the bettor enter a contract, the odds is fixed regardless of volume of bets and any new information that may be received by the market (Pope and Peel, 1989; Nyberg, 2014). Thus, individuals who want to place bets large enough to affect the odds in a pari-mutuel betting market will normally prefer fixed-odds betting (Sauer, 1998). Fixed-odds betting is the most common form of football betting in the U.K. (Kuypers, 2000).

Bookmakers' odds-setting procedure is based mainly on sophisticated predictions about outcomes of sports events considering as much relevant information as possible, but also on the amount betted on each outcome in order to manage risk and avoid heavy losses (Cortis, 2015). Furthermore, prices of odds include bookmakers' margins and, consequently, the odds for all outcomes in an event will correspond to a probability larger than one (Kuypers, 2000). Levitt (2004) further points out that bookmakers intentionally may set odds which deviate from the estimated probabilities as they in general are better predictors of outcomes than individual bettors. Thus, bookmakers can use this advantage to make greater profits. In the presence of bettors who are particularly skillful, there are however limits to how large these deviations can be.

2.4. Market Efficiency in Sports Betting Markets

Market efficiency is preferably approached applying the widely adopted efficient market hypothesis. Within this framework, market efficiency implies that prices reflect all available information in the

market. More specifically, the market efficiency definition is divided into three separate levels depending on different types of information. Firstly, weak form market efficiency is defined as that prices reflect all available information solely on historical prices. Secondly, the definition of semi-strong market efficiency extends to that prices reflect all publicly available information. Thirdly, the strong market efficiency definition includes as well insider information to be reflected in prices (Fama, 1970).

In the context of betting markets, the efficient market hypothesis suggests that odds reflect all available information. Applying the three different levels of market efficiency, weak form efficiency in betting markets is defined as that prices of odds reflect all information on historical odds. Furthermore, semi-strong market efficiency is defined as that prices of odds reflect all publicly available information relating to the concerned event, while strong market efficiency implies that prices additionally reflect information held by inside bettors (see e.g. Dowie, 1976; Gabriel and Marsden, 1990; Kuypers, 2000).

Following the terminology adopted by e.g. Gandar *et al.* (1988) and Gray and Gray (1997), market efficiency in betting markets can further be divided up into two categories. Firstly, statistical tests of market efficiency refer to statistical modelling aiming to evaluate the information reflected in odds. Thus, statistical efficiency is defined as that odds contain all information available in the market, and can more specifically be categorized as weak, semi-strong and strong form efficiency depending on the type of information considered. Secondly, economic tests of market efficiency evaluate market efficiency from the perspective of the bettor. In other words, such tests make use of sophisticated betting algorithms which aim to expose inefficiencies through development of profitable betting strategies. In this setting, market efficiency implies that it is impossible to formulate profitable betting strategies based on available information. Thus, economic weak form efficiency, semi-strong form efficiency and strong form efficiency is defined as that information on historical prices, publicly available information and all available information (including insider information), respectively, cannot be used to formulate such strategies.

2.5. The Favorite-Longshot Bias and Market Efficiency

In previous research on market efficiency in betting markets (see section 3), the anomaly found most consistently is the favorite-longshot bias, which is defined as that high probability events (i.e. “favorites”) are systematically undervalued while low probability events (i.e. “longshots”) are systematically overvalued in betting markets (see e.g. Griffith, 1949). To give an example, let us consider a football match from the data set used in this study. On May 6th 2018, Manchester City F.C. faced Huddersfield Town A.F.C. at home. During the 2017/2018 season, Manchester City

would go all the way to win the English Premier League while Huddersfield ended up on place 16, being only four points from relegation to the second tier in the English league system. The odds offered by bookmaker Bet365 for a Manchester City (home) win, a draw and a Huddersfield Town (away) win, respectively, are presented in Table 1.

Table 1: *Manchester City vs Huddersfield Town Odds*

Manchester City	Draw	Huddersfield Town
1.08	13	34

Note: Odds offered by bookmaker Bet365 for outcomes of the game Manchester City F.C. vs Huddersfield Town A.F.C., played on May 6th 2018. Final result: 0-0.

It is clear that the market considers Manchester City heavy favorites in this game while betting on a Huddersfield win is a true “longshot”, as a correctly predicted Manchester City win only yields a revenue of 1.08 times the invested money and a correctly predicted Huddersfield win yields a revenue of 34 times the invested money. If the betting market was perfectly efficient, these odds would fully reflect all information available. Conversely, in an inefficient market, there would be a deviation between the odds (which is the inverse of a probability) and the “true” probability for at least one of the outcomes. If a favorite-longshot bias was present, the probability given by the odds for a Manchester City win (i.e. a favorite win) would be lower than the “true” probability for that outcome (i.e. the odds is “too high”), while the probability given by the odds for a Huddersfield win (i.e. a longshot win) would be higher than the “true” probability of that outcome (i.e. the odds is “too low”). Playing this game infinitely many times, this means that Manchester City would win a larger proportion of the games, and Huddersfield would win a smaller proportion of the games, than what is implied by the odds. Thus, the favorite-longshot bias can be seen as a particular case of an inefficient market, as an inefficient market does not necessarily imply that a favorite-longshot bias is present. For instance, if odds for all outcomes in a market would overestimate the “true” probability of those events occurring, that would imply an inefficiency, however, it would not be a case of the favorite-longshot bias.

3. Literature Review

3.1. Early Evidence for the Favorite-Longshot Bias

Studying the pari-mutuel betting market for horse races in the U.S., using a data set containing winners and numbers of horses at each odds in 1386 horse races, Griffith (1949) was first to discover evidences for the favorite-longshot bias. By simply comparing the total number of entries with the number of winners times the odds adjusted for racetrack margins at each odds level, it was found that this product consistently exceeded the number of entries for low odds groups and conversely was smaller than the number of entries for high odds groups. Thus, bettors at these racetracks seemed to undervalue high probability outcomes (i.e. low odds groups) as well as overvalue low probability outcomes (i.e. high odds groups), indicating that this betting market suffered from a favorite-longshot bias.

McGlothlin (1956) likewise studies pari-mutuel betting markets for U.S. horse races by calculating an expected return of bets in different odds levels. If the market was efficient, the expected return would be the same regardless of the odds, however, it is found to be positive for high probability outcomes and negative for low probability outcomes, indicating a favorite-longshot bias. Furthermore, it is shown that bettors are particularly likely to place bets on longshots towards the end of the day. Lastly, using the same data set including roughly 12000 U.S. horse races between 1954 and 1964, the findings of Weitzman (1965) and Rosett (1965) further support the existence of a favorite-longshot bias.

3.2. Findings Across Different Betting Markets

Following the early findings of a favorite-longshot bias in U.S. horse racing pari-mutuel betting markets, a significant body of literature has investigated the efficiency of betting markets as well as the favorite-longshot bias, both for horse races and for several other sports.

Ali (1977) points out limitations in Griffith (1949), McGlothlin (1956) and Weitzman (1965) (e.g. small data sets, estimation errors) and uses a data set containing 20247 U.S. harness horse races. By distinguishing between objective probabilities, which is defined as the proportion of wins for a given horse, and subjective probabilities, which is defined as the proportion of the win pool that is bet on a given horse (i.e. the basis for the odds), and estimating the difference between these, it is shown that the market suffers from a favorite-longshot bias. Snyder (1978) once again finds evidence for a favorite-longshot in U.S. horse racing betting markets and furthermore tests its overall efficiency. Even though both weak, semi-strong and strong form tests reveal statistically significant differences between objective and subjective probabilities, indicating an inefficient market, it is proven to be difficult to make positive profits from horse race betting. The reason is

that profits are not large enough to exceed the racetrack stake of the betting pool. Similar results are found by Asch *et al.* (1982, 1984). Hausch *et al.* (1981) find the favorite-longshot bias pattern in both U.S. and Canadian horse racing betting markets and as well develop a technical betting system that manages to generate substantial positive profits. The profits are demonstrated not to be due to chance, and the market is thereby concluded to be economically weak form inefficient. Figlewski (1979), on the other hand, shows that bets on U.S. horse races do well in discounting subjective information contained in published predictions and thus concludes that these betting markets may be seen as rather efficient.

Evaluations of market efficiency in horse racing betting markets outside of the U.S. have been performed as well. In the U.K., where the horse racing betting markets mainly but not exclusively involve bookmakers (in contrast to the U.S. pari-mutuel betting system), Dowie (1976) shows that betting markets are (1) fairly economically weak form efficient in the sense that there is no systematic deviation in expected return across odds groups and (2) strong form efficient as outsider bettors have access to equally good information as insider bettors. Crafts (1985) argues that the method adopted by Dowie (1976) is inappropriate for testing strong form efficiency, instead uses a redesigned test of the same data set and concludes that insider bettors have the opportunity to exploit private information to gain profits. Gabriel and Marsden (1990) additionally claim that the U.K. horse racing betting market suffers both from semi-strong and strong form market inefficiencies, while Vaughan Williams and Paton (1997) particularly show evidence of a favorite-longshot bias in the same market. Tuckwell (1983) finds that the Australian horse racing betting market is both weak, semi-strong and strong form inefficient. The strong form inefficiency is additionally found possible to exploit in order to earn above average returns, and is thought to be a consequence of either imperfect communication channels or deliberate interference in the outcomes of races. Gandar, Zuber and Johnson (2001) show indications of a favorite-longshot bias in New Zealand horse racing betting markets. The evidence for the bias is particularly strong by the opening of betting, but slightly decreases towards closing. Lastly, Asian pari-mutuel horse racing betting markets seem to differ from equivalent markets on other continents, as no evidences for neither the favorite-longshot bias nor any market inefficiencies have been found. This has been concluded when studying both Hong Kong and Japanese racetrack odds (see e.g. Busche and Hall, 1988; Busche, 1994; Walls and Busche, 2003).

For other sports betting markets, research show widely varying results regarding efficiency and the favorite-longshot bias. For American football, Vergin and Scriabin (1978) show that the National Football League (NFL, the top tier of the American Football league system) betting market suffer from economic inefficiencies as it is possible to formulate profitable betting strategies

despite the conventional 5% betting fee. Golec and Tamarkin (1991) and Gray and Gray (1997) report similar conclusions and additionally show that bets on underdogs or home teams yield higher profits than bets on favorites or away teams, which could be seen as an indication of a “reverse” favorite-longshot bias. Zuber *et al.* (1985) and Gandar *et al.* (1988) cannot reject statistical market efficiencies in NFL betting markets, but find slight evidences for the possibility to make positive profits. In baseball, Woodland and Woodland (1994, 2003) show that the betting market for Major League Baseball (MLB, the top tier of the American baseball league system) matches is remarkably efficient as there are no significant differences between objective and subjective probabilities. Even though it is found difficult to formulate profitable betting strategies, bets on underdogs yield significantly less negative profits than other bets, which is interpreted as evidence for a reverse favorite-longshot bias. Gandar *et al.* (2002) as well find that the MLB betting market is efficient but disagree with the claim that it would be characterized by a reverse favorite-longshot bias, while Paul *et al.* (2008) similarly cannot reject the efficient market hypothesis in this betting market. The evidences supporting market efficiency in baseball betting markets corresponds to the notion that bettors engaged in such markets are among the most sophisticated and knowledgeable of all sports bettors (Woodland and Woodland, 1994). Further evidence on a reverse favorite-longshot bias in U.S. sports betting markets are provided by Woodland and Woodland (2001) as well as Paul and Weinbach (2005), examining odds for ice hockey and basketball matches respectively. Forrest and McHale (2007) find indications of a favorite-longshot bias in odds for men’s single tennis matches, looking at in total 8500 matches in international tournaments between 2002 and 2005. Schnytzer and Weinberg (2008) find no evidence of neither a favorite-longshot bias nor any profitable betting strategies in the betting market for Australian Rules Football matches. Studying odds for outcomes on the PGA Tour, Shmanske (2005) concludes that the efficient market hypothesis for golf betting markets overall holds up. Terrell and Farmer (1996) find that U.S. greyhound racing betting markets, which are similar to horse racing markets in the sense that they as well apply the pari-mutuel odds-setting procedure, are characterized by a favorite-longshot bias. Lastly, Cain *et al.* (2003) investigate market efficiency in a wide range of sports betting markets, including baseball, boxing, cricket, horse racing, greyhound racing, football, snooker and tennis. The favorite-longshot bias is discovered in odds for greyhound- and horse racing, snooker and boxing competitions, while the other betting markets are concluded to be efficient.

3.3. Market Efficiency and the Favorite-Longshot Bias in Football

In the previous literature, mainly two types of models have been used to evaluate statistical market efficiency and the favorite-longshot bias in fixed-odds football betting markets. The first type of

model involves forecasting of football match outcomes where the forecasts are used as objective (“true”) probabilities, which are compared to the implicit (subjective) probabilities given by the odds to evaluate market efficiency. The forecasting has been done either by predicting goals scored and conceded through various forms of Poisson regressions or by predicting match outcomes mainly through ordered probit regressions (Forrest *et al.*, 2005; Goddard, 2005). The second type of model involves direct testing of efficiency by including relevant information as predictors of match outcomes in for instance linear probability (ordinary least square, weighted least square), binary logit or multinomial logit regressions and then analyzing the estimated parameters. Lastly, economic market efficiency has been evaluated by testing if betting strategies (1) yield positive profits and (2) systematically generate varying profits across different odds levels.

Dixon and Coles (1997) apply a bivariate Poisson distribution model for English league and cup match odds between 1992 and 1995 and find that the estimated outcomes in some cases deviate from bookmakers’ odds to such an extent that slightly positive returns can be yielded. Cain *et al.* (2000) use Poisson as well as Negative Binomial distributions and show that the betting market for English league matches is characterized by a favorite-longshot bias which, however, rarely can be exploited to make profits. Dixon and Pope (2004) as well perform Poisson regressions but conclude that the English football betting market instead is characterized by a reverse favorite-longshot bias.

Kuypers (2000) uses an ordered probit model to test semi-strong form efficiency in the betting market for English football league matches in seasons 1993/1994 and 1994/1995, where the latent variable y^* is determined by including information on teams’ recent performances and overall season performances in terms of number of points gained as well as number of goals scored and conceded. Comparing the forecasts with offered odds, it is shown that the market is economically semi-strong inefficient as it is possible for sophisticated bettors to gain positive profits. Goddard and Asimakopoulous (2004) adopt a similar model and include information on recent performance, the importance of the particular match for each team, domestic cup elimination as well as travel distance for the away team for English league matches. Evidence for semi-strong form inefficiency¹ is found, however, only rather modest returns can be made. Forrest *et al.* (2005) extend the model to include as well an explanatory variable representing home teams’ attendance relative to league position and find that British bookmakers seem to make good use of

¹ Goddard and Asimakopoulous (2004) claim that they perform a weak form efficiency test but as Vlastakis *et al.* (2009) point out it is in fact a semi-strong form efficiency test. The reason is that they include publicly available information not only on historical odds prices, but as well on match significance, recent domestic cup results and travel distance for away teams to predict match outcomes in their model.

available information and thus conclude that the same betting market can be seen as semi-strong form efficient. Lastly, Goddard (2005) compares the forecasting performance of the goal-scoring bivariate Poisson regression approach with the match-outcome ordered probit regression approach. Even though the difference in performance between the models is rather small, it is shown that the most accurate forecasts are obtained when applying an ordered probit regression with number of goals scored and conceded included as covariates, thus a “hybrid” model using elements from both approaches.

Pope and Peel (1989) use both ordinary and weighted least square regressions to directly test the relationship between odds and actual match outcomes and thus investigate weak form efficiency. Using a data set including English league matches and corresponding odds from the 1981/1982 season, it is found that there are indications pointing toward statistical inefficiencies, particularly for draw outcomes. However, this does not allow for implementation of profitable betting strategies. Kuypers (2000) similarly adopts an ordinary least square regression but finds no significant differences between odds and actual outcomes in English 1993-1995 league matches. Consequently, the market is considered statistically weak form efficient. Studying football matches from the ten highest ranked European leagues (Belgium, England, France, Germany, Greece, Italy, Netherlands, Portugal, Spain and Turkey) between 2002 and 2010, applying a binary logit regression to model the relationship between odds and actual outcomes, Koning (2012) finds evidence for a favorite-longshot bias in several countries both for home and away outcomes. Nyberg (2014) essentially tests the same relationship, however, instead using a multinomial logit model. The data set consists of matches in the top four tiers of the English league system (i.e. the Premier League, Championship, League One and League Two), and it is concluded that only the Premier League betting market can be considered weak form inefficient. Additionally, a favorite-longshot bias as well as an away-favorite bias (i.e. that away teams with a high probability of winning are undervalued by the betting market to a larger extent than home teams with a high probability of winning) are discovered for these matches. Vlastakis *et al.* (2009) likewise find evidences for the favorite-longshot- and away-favorite bias, assessing the efficiency of the overall European betting market by applying both a Poisson regression model and a multinomial logit model. Furthermore, indications of highly profitable arbitrage opportunities are found. Even though such opportunities appear to be rare, the market is considered to be statistically and economically inefficient.

Deschamps and Gergaud (2007) perform an economic weak form efficiency test for the English betting market by splitting implicit probabilities into 20 separate groups so that nearby probabilities are assigned to the same group and then computing the average return from betting £1 on odds in each group. Despite that this strategy does not yield positive returns, not even when

only the best (i.e. highest) odds across several bookmakers is chosen, tendencies pointing towards a favorite-longshot bias for home and away wins are found as betting on longshots generate substantially lower returns than betting on favorites. However, for draw outcomes, a reverse favorite-longshot is instead identified. Direr (2013) develops a betting strategy where European football odds less than or equal to various thresholds are chosen. As this strategy yields profitable returns for certain thresholds, it is concluded that the market is economically weak form inefficient. Furthermore, evidence for the favorite-longshot bias is found as returns fall as odds levels increase.

3.4. Explanations for the Favorite-Longshot Bias

The existence of a favorite-longshot bias in sports betting markets has generally been explained by two distinct set of theories. Neoclassical theories claim that the bias is a result of that rational but risk-loving bettors, whom gain utility from increased levels of risk, overbet longshots. Behavioral theories instead suggest that cognitive errors and misperceptions about probabilities are the main causes for inefficient betting markets (Snowberg and Wolfers, 2010). Both risk-loving and cognitive misperceptions may affect bettors' behavior in a way that results in overvaluation of longshots and undervaluation of favorites.

Rosett (1965) concludes that bettors are rational but seem to display strong preferences for longshot bets, resulting in the emergence of the favorite-longshot bias. Weitzman (1965) similarly assumes rationality and explains bettors' preferences for longshots with convex utility functions implying that they are at least locally risk-loving. Ali (1977) as well estimates the representative bettor to be risk-loving, while Quandt (1986) shows that the equilibrium in betting markets with risk-loving bettors by definition results in the favorite-longshot bias. Explanations for why bettors are risk-loving include that it is more "fun" to bet on a longshot, as it potentially comes with the reward from outperforming everyone else (Snyder, 1978) and considerable bragging rights; thus, bettors may gain utility solely from holding a longshot bet (Thaler and Ziemba, 1988).

Alternatively, behavioral theories explain the existence of a favorite-longshot bias with cognitive errors and systematic misperceptions of probabilities. Prospect theory suggests, among other things, that (1) small probabilities tend to be overweighted and (2) that certainty is overweighted relative to uncertainty (Kahneman and Tversky, 1979). In other words, (1) means that the subjective estimation of a probability is larger than the actual probability for very small probabilities, which could explain bettors' tendency to overbet longshots (Griffith, 1949; Snowberg and Wolfers, 2010). At the same time, (2) implies that sure outcomes are preferred over better outcomes that are extremely likely. Thus, bettors avoid favorites as they are disproportionately afraid of losing such bets (Jullien and Salanié, 2000). Furthermore, pure irrationalities could to some

extent explain why longshots are overbet and favorites are underbet. For instance, if some bettors would choose to bet on a horse (or a football team) only because of its name, it could drive the odds down for unlikely outcomes (Thaler and Ziemba, 1988). Moreover, horses (or football teams) that has lost many races (games) in a row are generally longshots, thus, many bettors' fallacy belief that such horses (teams) now are due for a win could lead to a general pattern of overbetting on low probability outcomes (Camerer, 1998).

It should also be noted that an additional set of explanations emphasis bookmakers' potential role in contributing to the favorite-longshot bias in fixed-odds betting markets. For instance, Henery (1985) claims that betting on favorites is much more competitive and therefore bookmakers raise favorite odds to attract bettors, while they reduce longshot odds to make up for lost profit. Additionally, Shin (1991, 1992) suggests that the favorite-longshot bias is a result of uninformed bookmakers responding to the existence of insider information in the betting market.

3.5. Contribution to Research

By investigating the research questions stated in section 1, this study aims to extend the current body of knowledge in two different ways. Firstly, the study uses a data set including odds for Premier League matches up until 2018. Considering the significant growth of betting markets due to the emergence of online betting, in particular during the last couple of years, which may have had an impact on both bettors' behavior and bookmakers' incentives, it is relevant to re-examine market efficiency and the favorite-longshot bias in the light of this rapidly changing market. Secondly, as shown in section 3.3., the previous literature provides rather ambiguous results concerning whether the favorite-longshot bias actually is present in betting markets for football matches or not. Thus, by analyzing a new and large data set through various methods of sophisticated statistical modelling, the study aims to give further insights about the present characteristics of this particular betting market.

4. Data

The data set includes odds for home wins, draws and away wins from 11 different bookmakers, as well as final match outcomes, for all English Premier League matches played from the 2006/2007 season up until the 2017/2018 season. The total number of matches is 4560 and the 11 bookmakers are Bet365, Bet&Win, Gamebookers, Interwetten, Ladbrokes, Pinnacle, Sportingbet, William Hill, Stan James, VC Bet and Blue Square. Odds were collected on Friday afternoons for matches played on weekends and on Tuesday afternoons for matches played during midweeks. Unfortunately, odds from all bookmakers are not available for all seasons, thus, the total number of odds is 119481. All data was provided by www.football-data.co.uk. Descriptive statistics of the data set are presented in Table 2.

Table 2: *Descriptive Statistics of the Data Set*

	06/07	07/08	08/09	09/10	10/11	11/12	12/13	13/14	14/15	15/16	16/17	17/18
B365	1140	1140	1140	1140	1140	1140	1140	1140	1140	1140	1140	1140
BW	1140	1140	1140	1140	1140	1140	1140	1140	1140	1137	1140	1140
GB	1140	1140	1140	1140	1140	1140	1137	0	0	0	0	0
IW	1140	1140	1137	1140	1140	1140	1140	1140	1137	1140	1140	1140
LB	1140	1140	1140	1140	1140	1140	1137	1140	1140	1140	1140	1140
PS	0	0	0	0	0	0	1140	1140	1140	1140	1140	1140
SB	1140	1140	1140	1140	1140	1140	0	0	0	0	0	0
WH	1140	819	1140	1140	1140	1140	1140	1140	1140	1140	1140	1140
SJ	1137	1140	1140	1140	1140	1140	1140	1140	120	0	0	0
VC	1140	1140	1140	1140	1140	1140	1140	1140	1140	1140	1140	1140
BS	0	1140	1140	1140	1140	1140	1140	0	0	0	0	0
Total	10257	11079	11397	11400	11400	11400	11394	9120	8097	7977	7980	7980

Note: The table illustrates number of odds from each bookmaker for each season. The bookmakers are Bet365 (B365), Bet&Win (BW), Gamebookers (GB), Interwetten (IW), Ladbrokes (LB), Pinnacle (PS), Sportingbet (SB), William Hill (WH), Stan James (SJ), VC Bet (VC) and Blue Square (BS).

5. Method

5.1. Obtaining implicit probabilities from bookmakers' odds

This study aims to test the statistical weak form efficiency in general, as well as the favorite-longshot bias in particular, in the fixed-odds betting market for English Premier League football matches. If all information on historical prices are reflected in odds, implying statistical weak form efficiency, there should be no systematic differences between probabilities given by odds and probabilities estimated from actual match outcomes, which essentially is what the tests will evaluate. For football matches there are three possible outcomes – home win (i.e. goals scored by the home team > goals scored by the away team), away win (goals scored by the home team < goals scored by the away team) and draw (goals scored by the home team = goals scored by the away team). In order to perform the tests, odds are transformed into implicit probabilities. As odds and probabilities are inversely related, i.e. a high odds implies a low probability and vice versa, the implicit probabilities are calculated by taking the inverse of the odds.

Furthermore, following the standard procedure in the previous literature (see e.g. Pope and Peel, 1989; Goddard and Asimakopoulou, 2004; Levitt, 2004; Forrest *et al.*, 2005; Deschamps and Gergaud, 2007; Vlastakis *et al.*, 2009 and Nyberg, 2014), these probabilities are adjusted for bookmakers' margins. As data neither on actual margins nor distribution of bets (which could be used to obtain actual margins) is available, implicit margins are calculated. To do so, it is assumed that margins are equally distributed across match outcomes (i.e. that the bookmaker margin is equal for a home win, away win and draw in a given match) and that odds are determined on the basis of true probabilities (which are estimated by bookmakers). As discussed in section 2.3., Levitt (2004) argues that bookmakers intentionally may set odds which deviate from true probabilities in order to gain larger profits, which would make the second assumption somewhat problematic. On the other hand, due to the increased competition in sports betting markets and the emergence of online betting (Forrest *et al.*, 2005), overall margins have decreased and incentives for bookmakers to set odds that reflect true probabilities in order to avoid substantial losses have increased. This limits the effect that these assumptions have on the results.

To obtain the implicit probabilities, the average odds for each outcome in a given match is calculated across all bookmakers as,

$$\bar{o}_i^H = \frac{1}{K} \sum_{k=1}^K o_{i,k}^H$$

$$\begin{aligned}\bar{O}_i^D &= \frac{1}{K} \sum_{k=1}^K O_{i,k}^D \\ \bar{O}_i^A &= \frac{1}{K} \sum_{k=1}^K O_{i,k}^A\end{aligned}\tag{1}$$

where $O_{i,k}^H$, $O_{i,k}^D$ and $O_{i,k}^A$ represent the k th bookmaker's odds in match i for a home win, draw and away win respectively. The implicit probabilities for each outcome in a given match are then calculated as,

$$\begin{aligned}\phi_i^H &= \frac{1}{\bar{O}_i^H(1 + \xi_i)} \\ \phi_i^D &= \frac{1}{\bar{O}_i^D(1 + \xi_i)} \\ \phi_i^A &= \frac{1}{\bar{O}_i^A(1 + \xi_i)},\end{aligned}\tag{2}$$

where the bookmaker margin is given by,

$$\xi_i = \frac{1}{\bar{O}_i^H} + \frac{1}{\bar{O}_i^D} + \frac{1}{\bar{O}_i^A} - 1,\tag{3}$$

which is assumed to be positive, $\xi_i > 0$, to cover for costs and profit. Since adjusted for margins, the implicit probabilities for all outcomes will sum to one for each match. As the data set contains 4560 matches, the total number of implicit probabilities will be 13680.

5.2. Multinomial Logit Model

To test for statistical weak form efficiency as well as the favorite-longshot bias in the aggregate fixed-odds betting market for Premier League football matches, a multinomial logit regression² is applied. As discussed by Nyberg (2014), the multinomial logit model has two main advantages compared to other models that have been applied in the previous literature. Firstly, it allows the three possible match outcomes to be tested simultaneously within the same model, which is not possible with neither binary regressions (see e.g. the binary logit model used by Koning, 2012) nor

² The multinomial logit regression was executed in R using the mlogit package. Code is available upon request.

linear regressions (see e.g. Zuber *et al.*, 1985; Gandar *et al.*, 1988; Sauer *et al.*, 1988; Pope and Peel, 1989; Goddard and Asimakopoulous, 2004 and Franck *et al.*, 2010) where each outcome must be evaluated separately. Thus, these approaches fail to take into account the restriction that one of the possible outcomes actually will occur for each match, resulting in that the predicted probabilities may not sum up to one.

Secondly, the multinomial logit model allows the efficiency of the implicit probabilities to be straightforwardly tested. For a large proportion of the statistical modeling that has been used to evaluate fixed-odds betting market efficiency for football matches, such as independent Poission distributions, modelling of the number of goals scored, and ordered probit regressions, estimating the probability of match outcomes (see section 3.3. for details about these models), this is not the case. The reason is that these models are based upon a single latent variable which implies that the parameter vector will be the same for all outcomes. In the multinomial logit regression each outcome will instead have its own parameter vector, which makes it possible to construct an hypothesis which can test the weak form efficiency of the aggregated market directly.

Multinomial logit regressions have been used to test market efficiency in sports betting markets in several previous studies. For football betting markets, Vlastakis *et al.* (2009) and Nyberg (2014) both use multinomial logit regressions, including implicit probabilities obtained from odds as independent variables, to test for market efficiency. Furthermore, Figlewski (1979) applies a similar model to evaluate efficiency with respect to subjective information in the U.S. horse race betting market. Lastly, Akhtar and Scarf (2012) use a multinomial logit model to predict outcomes in test cricket matches.

Following Nyberg (2014), the outcome of a given match i is denoted by $y_i = j$ where j is the final result and takes the value $y_i = 1$ for a home win (denoted by H in the previous section), $y_i = 0$ for a draw (D) and $y_i = -1$ for an away win (A). To be able to perform the multinomial logit regression, three binary indicator variables $y_{i,j}$ are introduced for $j = 1, 0, -1$ of which one of them takes the value 1 depending of the outcome of the match, i.e. if the outcome is a home win, then $y_{i,1} = 1$, $y_{i,0} = 0$ and $y_{i,-1} = 0$, if the outcome is a draw then $y_{i,1} = 0$, $y_{i,0} = 1$ and $y_{i,-1} = 0$ and if the outcome is an away win then $y_{i,1} = 0$, $y_{i,0} = 0$ and $y_{i,-1} = 1$. Furthermore, the implicit probabilities for the three possible outcomes are contained in the vector,

$$\phi_i = [\phi_i^1 \quad \phi_i^0 \quad \phi_i^{-1}]'. \quad (4)$$

To determine the probabilities $\rho_{i,j}$ ($j = 1, 0, -1$) for the match outcomes $y_i = j$, conditional on the implicit probabilities, the multinomial logit model is written using log-odds ratios,

$$\log\left(\frac{\rho_{i,j}}{\rho_{i,0}}\right) = \pi_{i,j}, \quad (5)$$

where $\pi_{i,j}$ is a linear function that should be determined to complete the model. This is achieved by including the log-odds ratios consisting of the implicit probabilities,

$$\log\left(\frac{\phi_i^j}{\phi_i^0}\right) = \hat{\pi}_{i,j}, \quad (6)$$

as predictors in the linear function which is specified as,

$$\pi_{i,j} = \alpha_j + \hat{\pi}_{i,j}\beta_j, \quad j = 1, -1. \quad (7)$$

Since the three conditional probabilities in a given match will sum to one, it is only necessary to estimate the parameters α_j and β_j for two out of three outcomes in order to obtain the probabilities for all outcomes. Following Vlastakis *et al.* (2009) and Nyberg (2014), draws are used as this benchmark category, implying that the linear function is not determined for those outcomes.

Thus, solving for the conditional probabilities from eq. (5) gives,

$$\begin{aligned} \rho_{i,1} &= P(y_i = 1|\boldsymbol{\phi}_i) = \frac{\exp(\pi_{i,1})}{1 + \exp(\pi_{i,1}) + \exp(\pi_{i,-1})} \\ \rho_{i,0} &= P(y_i = 0|\boldsymbol{\phi}_i) = \frac{1}{1 + \exp(\pi_{i,1}) + \exp(\pi_{i,-1})} \\ \rho_{i,-1} &= P(y_i = -1|\boldsymbol{\phi}_i) = \frac{\exp(\pi_{i,-1})}{1 + \exp(\pi_{i,1}) + \exp(\pi_{i,-1})} \end{aligned} \quad (8)$$

where $\sum_{j=-1}^1 \rho_{i,j} = 1$. The parameters in the multinomial logit regression are estimated by the maximum likelihood method. Using the binary indicator variables, the likelihood function is written as,

$$L(\boldsymbol{\theta}) = \prod_{i=1}^N \prod_{j=-1}^1 \rho_{i,j}^{y_{i,j}}, \quad (9)$$

where N is the sample size and the parameters are included in the vector,

$$\boldsymbol{\theta} = [\alpha_1 \quad \alpha_{-1} \quad \beta_1 \quad \beta_{-1}]'. \quad (10)$$

The log-likelihood function is thus given by,

$$l(\boldsymbol{\theta}) = \log L(\boldsymbol{\theta}) = \sum_{i=1}^N l_i(\boldsymbol{\theta}) = \sum_{i=1}^N \sum_{j=-1}^1 y_{i,j} \log(\rho_{i,j}), \quad (11)$$

which is estimated using numerical methods performed in R.

5.3. Grouped Z-Tests

In addition to the multinomial logit regression, grouped z-tests are executed. In this setting, “grouped” refers to that the implicit probabilities obtained in eq. (2) will be divided into groups so that high implicit probabilities are assigned to the same group and vice versa. For each group is then the proportion of “winning” bets (i.e. the number of odds representing the actual outcome of the match divided by the total number of odds) as well as the average implicit probability calculated, which allows for a z-test that tests if the average implicit probability significantly differs from the proportion of winning bets and thus if the market is statistically weak form inefficient. There are two main reasons why these tests are executed. Firstly, the grouping allows for a more detailed plotting of the relationship between implicit and objective probabilities (given by the predicted conditional probabilities in the multinomial logit regression and by the proportion of winning bets in the grouped z-tests). Since the multinomial logit regression tests the aggregated efficiency, this relationship will be determined parametrically across the full range of probabilities. In contrast, the grouped z-tests allow for a non-parametric determination of the relationship between implicit and objective probabilities across the full range of probabilities, even though each z-test in itself is parametric. Thus, the grouped z-test procedure can be seen as a somewhat more flexible test than the multinomial logit regression, however, with less ability to evaluate the aggregated market efficiency. Secondly, the grouped z-tests are used to evaluate robustness in the results obtained from the multinomial logit regression, which is considered the main test of this study. This is possible even though the two tests are designed somewhat differently, - as discussed above, the z-tests test different probability levels separately while the multinomial logit regression tests the full range of probabilities aggregately – as they both fundamentally evaluate whether there

are significant differences between probabilities given by odds and probabilities estimated from actual match outcomes.

In order to perform the z-tests, the first thing to consider is how the implicit probabilities should be grouped as well as how many groups they should be divided into. As discussed by Direr (2013), this kind of grouping involves an inevitable trade-off between statistical significance (i.e. having enough observations in each group) and homogeneity (i.e. having implicit probabilities with similar characteristics in the same group) which has to be dealt with. In the previous literature, a set of different grouping approaches based on odds levels have been adopted. The first approach implies dividing the implicit probabilities into equally sized groups (see e.g. Deschamps and Gergaud, 2007), which allows for statistical significance as the group size can be controlled by adjusting number of groups, however, may result in heterogeneous groups if there are few observations across a wide probability range. Conversely, a second approach implies equal probability intervals for all groups (see e.g. Cain *et al.*, 2003). This reduces at least to some extent the homogeneity problem in the first approach, however, may result in groups with few observations. Lastly, Direr (2013) combines the two approaches through an algorithm which aims to balance statistical significance and homogeneous groups by adding the implicit probabilities to a given group until a threshold for number of observations is reached.

The procedure for the grouped z-tests is as follows. The implicit probabilities are classified into groups (see details about the grouping procedure below) so that nearby probabilities are included in the same group. Then, the average implicit probability for each group is calculated as,

$$\bar{\phi}_m = \frac{\sum_{i=1}^n \phi_{i,m}^j}{n_m}, \quad (12)$$

where $\bar{\phi}_m$ is the average implicit probability in the m th group, $\phi_{i,m}^j$ is the implicit probability for a given outcome in match i now assigned to group m , and n is the number of observations in each group. Furthermore, the proportion of winning bets in each group is calculated as,

$$\gamma_m = \frac{\sum_{i=1}^n y_{i,j,m}}{n_m}, \quad (13)$$

where γ_m is the proportion of winning bets in the m th group and $y_{i,j}$ is the binary indicator variable introduced in section 5.2. for the corresponding outcome in each group. Following Woodland and

Woodland (1994), the binomial distribution is approximated by a normal distribution, which gives the following test statistic for group m ,

$$z_m = \frac{\gamma_m - \bar{\phi}_m}{\sqrt{\frac{\bar{\phi}_m(1 - \bar{\phi}_m)}{n_m}}} \quad (14)$$

As the proportion of winning bets approaches a normal distribution only for sufficiently large n_m , groups are excluded if $n_m \bar{\phi}_m < 5$ or $n_m(1 - \bar{\phi}_m) < 5$.³

In order to manage the significance-homogeneity trade-off, the grouping procedure is done in four separate ways. Firstly, implicit probabilities for home wins, draws and away wins are separately divided into 10 groups with equal probability intervals. Secondly, implicit probabilities for home wins, draws and away wins are separately divided into 10 groups with equal number of observations. In order to perform z-tests for all outcomes simultaneously and thus as well increase the number of observations in each group, implicit probabilities for home wins, draws and away wins are pooled together and then divided into 10 groups both with equal probability intervals and equal number of observations in each group. The pooling gives rise to a dependency issue as implicit probabilities for different outcomes in the same match (which are highly dependent of each other, i.e. if an outcome in a given match to a large extent is likely, the others are by definition unlikely as the probabilities for all outcomes sum to one) risk to be assigned to the same group. In order to avoid this, in the case where implicit probabilities for more than one outcome in the same match are assigned to the same group, probabilities are randomly withdrawn from the group so that only a probability for one outcome in a given match can be found in a group. This approach reduces the total number of implicit probabilities from 13680 to 11587 for grouping with equal probability intervals and to 12677 for grouping with equal number of observations, however, the number of observations in each group is still significantly larger than when home wins, draws and away wins are considered separately. Based on the grouping procedures described above, four different categories of z-tests are executed in order to test the difference between the implicit probabilities and actual match outcomes.

³ In total, 2 groups were excluded due to insufficiently large n_m .

5.4. Hypotheses

For the multinomial logit regression, it is evaluated whether odds are unbiased predictors of actual match outcomes on an aggregate level. In other words, statistical weak form efficiency is given by,

$$\pi_{i,1} = \hat{\pi}_{i,1}, \pi_{i,-1} = \hat{\pi}_{i,-1}. \quad (15)$$

Thus, if the null hypothesis,

$$\begin{aligned} H_0: \alpha_j &= 0, \beta_j = 1, j = 1, -1 \\ H_1: \alpha_j &\neq 0, \beta_j \neq 1, j = 1, -1 \end{aligned} \quad (16)$$

is rejected, the conditional probabilities predicted by the multinomial logit regression systematically deviate from the implicit probabilities given by the odds, implying that this fixed-odds betting market is statistically weak form inefficient. Furthermore, if the null hypothesis is rejected, each of the four parameters will be tested separately in order to determine whether the market suffers from a favorite-longshot bias. In other words, if $\alpha_j < 0$ and $\beta_j > 1$ for $j = 1, -1$ a favorite-longshot bias is present, as these conditions imply that longshots (i.e. low probability outcomes) will occur less often than what is implied by their odds and that favorites (i.e. high probability outcomes) will occur more often than what is implied by their odds.

The null hypothesis given by eq. (16) is evaluated performing a Wald test. In order to do this, the null hypothesis is rewritten as,

$$A\theta = c, \quad (17)$$

where θ is the parameter vector introduced in section 5.2., A represents an identity matrix, i.e. $A = I_4$, and $c = [0 \ 0 \ 1 \ 1]'$. The Wald test statistic is then given by,

$$W = (A\hat{\theta} - c)' [AJ(\hat{\theta})^{-1}A']^{-1} (A\hat{\theta} - c), \quad (18)$$

where $\hat{\theta}$ is the maximum likelihood estimator and $J(\hat{\theta})$ is a consistent estimator of $I(\theta_0)$ in which θ_0 represents the true value of the parameter vector given by the data generating process. The Wald test statistic follows an asymptotic χ^2_4 -distribution under the null hypothesis (Nyberg, 2014).

For the grouped z-tests, it is evaluated whether the observed proportion of winning bets differs from the average implicit probability in each group. Rejection of the null hypothesis,

$$\begin{aligned} H_0: \gamma_m - \bar{\phi}_m &= 0 \\ H_1: \gamma_m - \bar{\phi}_m &\neq 0, \end{aligned} \tag{19}$$

for a given group suggests that a significant difference is present and that odds in that particular probability interval therefore not are priced in accordance with the observed probability of the corresponding outcome and thus are weak form inefficient. Furthermore, if the null hypothesis is rejected for all groups, or at least the vast majority of the groups, it would indicate that odds systematically differs from the observed probabilities of actual outcomes, which is interpreted as an overall weak form inefficient market. In order to evaluate the favorite-longshot bias, the groups with the highest and lowest implicit probabilities respectively are considered. If the observed proportion of winning bets is larger than the average implicit probability in high probability groups, and smaller in low probability groups, “favorite” outcomes occur more often and “longshot” outcomes occur less often than what is suggested by their odds, indicating a favorite-longshot bias.⁴

⁴ All hypotheses are tested at a 5% significance level and all tests were fully determined prior to conducting any analysis on the data.

6. Results

6.1. Multinomial Logit Model

The estimates of the multinomial logit regression are presented in Table 3.

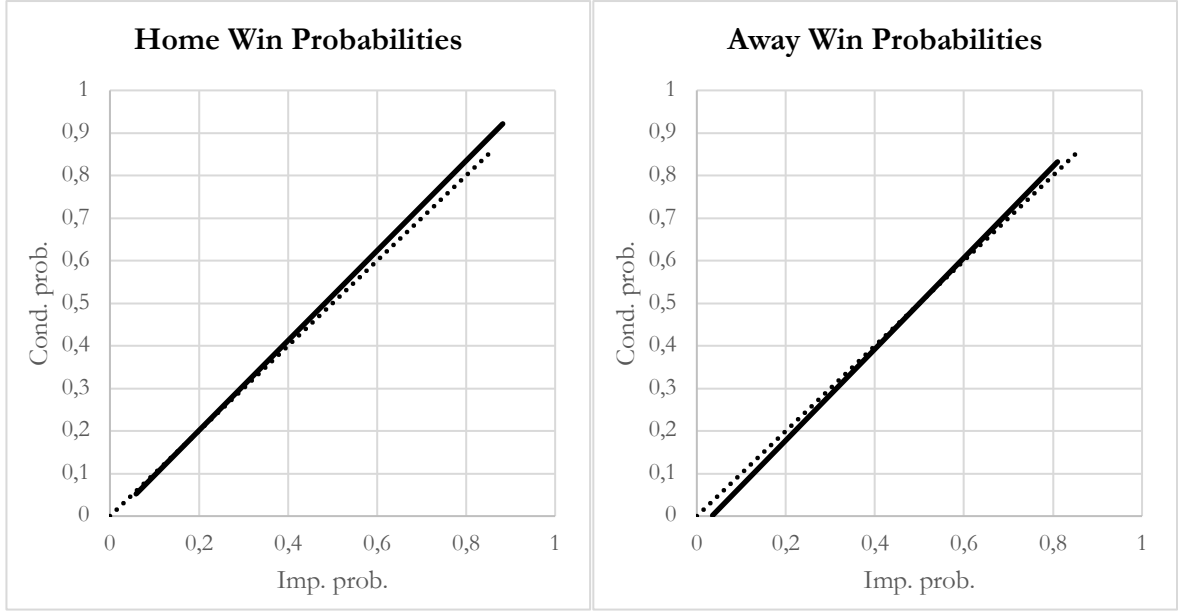
Table 3: *Multinomial Logit Regression Estimates*

α_{-1}	-0.0969644* (0.0441751)
α_1	0.0056965 (0.0527148)
β_{-1}	1.1709930*** (0.0837186)
β_1	1.0459386*** (0.0708711)
<i>Log-likelihood</i>	-4377.8
$\chi^2_4 (\alpha_j = 0, \beta_j = 1)$	13.5**

Note: Estimates for $j = 1, -1$ as draws ($j = 0$) is used as the benchmark category. Standard errors are reported in parentheses. $\chi^2_4 (\alpha_j = 0, \beta_j = 1)$ refers to Wald's test statistic for testing the null hypothesis (eq. 16) presented in section 5.4. Significance codes: * = 0.05, ** = 0.01 and *** = 0.001.

It is shown that three out of the four estimated parameters are significant at a 5% significance level. Furthermore, obtaining the results from the Wald test, the null hypothesis for statistical weak form efficiency (eq. 16) is rejected at a 1% significance level, indicating that the aggregate market is statistically weak form inefficient. Consequently, Wald tests are performed for each parameter separately. The results show that α_{-1} is smaller than 0 and that β_{-1} is larger than 1 on a 5% significance level, while α_1 is statistically equal to 0 and β_1 is statistically equal to 1. Thus, the betting market for away win odds is considered to be statistically weak form inefficient and is furthermore characterized by a favorite-longshot bias, while the null hypothesis for weak form efficiency cannot be rejected for home win odds. This is further illustrated in Figure 1, which plots the relationship between the conditional probabilities estimated by the multinomial logit regression $\rho_{i,j}$ (y-axis) and the implicit probabilities obtained from the odds ϕ_i^j (x-axis) for home and away wins separately ($j = 1, -1$). In a perfectly efficient market, the conditional probability should equal the implicit probability for each probability level, which is represented by the dashed 45-degree line in both graphs.

Figure 1: *Graphical Illustrations of the Multinomial Logit Regression Estimates*



Note: Conditional probabilities estimated by the multinomial logit model $\rho_{i,j}$ plotted against implicit probabilities obtained from odds ϕ_i^j , along with a dashed 45-degree line.

For home wins, the estimated line follows the 45-degree line up until high probability levels where slight deviations are found. However, as shown by the results from the Wald test, these deviations are too small to reject statistical weak form efficiency. For away wins, on the other hand, the estimated line deviates from the 45-degree line across a larger range of probability levels. As shown by the Wald test, this is concluded to be evidence for weak form inefficiencies. Furthermore, the deviations are not consistent across all probability levels – for low probabilities the estimated line is below the 45-degree line and for high probabilities the estimated line is above the 45-degree line. The implication of this is that for low probabilities (i.e. longshots), implicit probabilities obtained from odds are higher than the conditional probabilities while for high probabilities (i.e. favorites), implicit probabilities are lower than the conditional probabilities. In other words, considering the inverse relationship between odds and probabilities, longshot outcomes seem to be overvalued as they occur more often than what is implied by their odds, while favorite outcomes seem to be undervalued as they occur less often than what is implied by their odds, indicating a favorite-longshot bias for away wins.

6.2. Grouped Z-Tests

The results from the first category of grouped z-tests, where implicit probabilities are divided into 10 groups with equal probability intervals, and home wins, away wins and draws are considered separately, are shown in Table 4.

Table 4: Results from the First Category of Grouped Z-tests

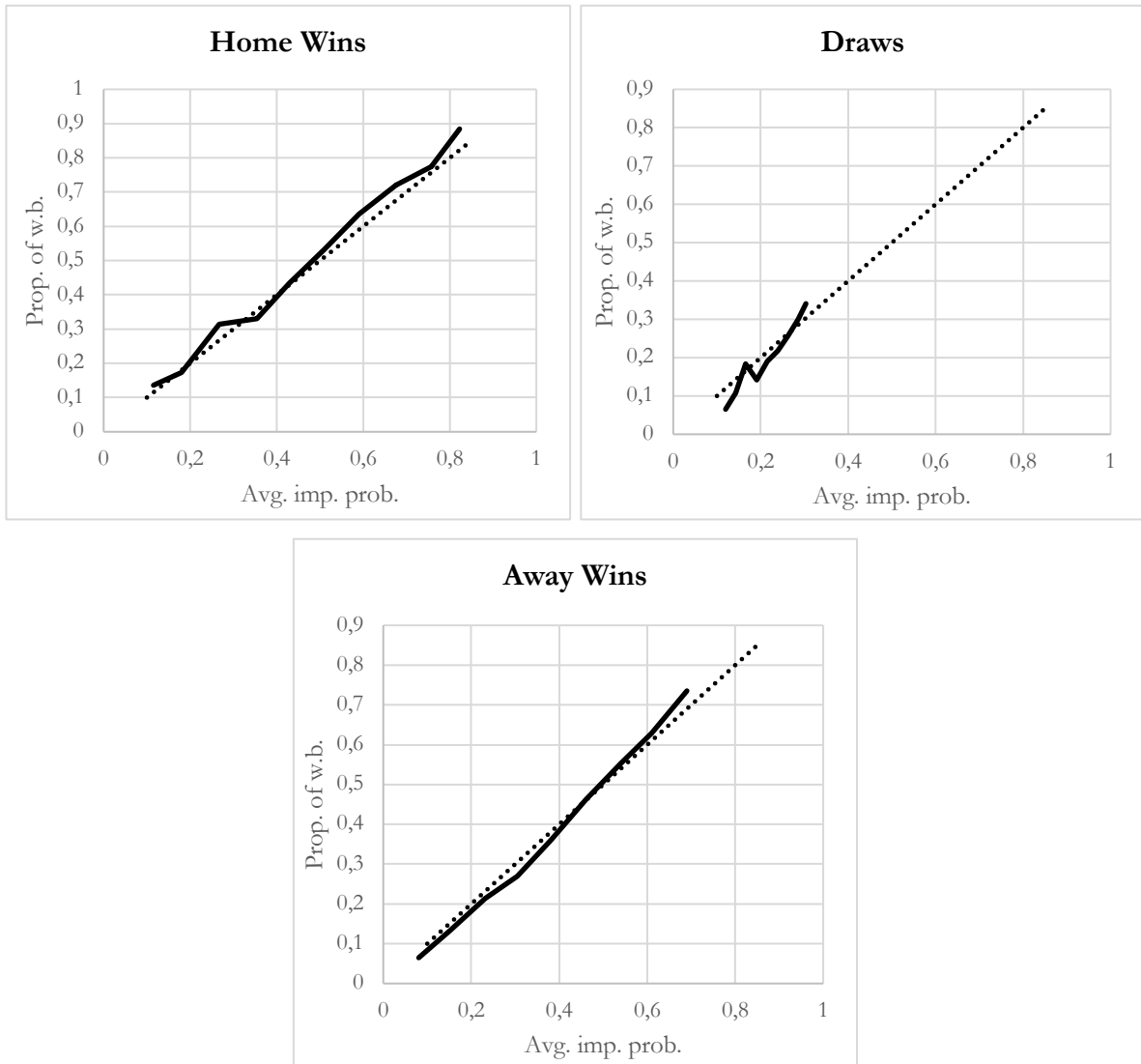
m	n_m	$int.$	$\bar{\phi}_m$	γ_m	z_m
<i>Home Wins</i>					
1	184	$.06 < \bar{\phi}_m \leq .14$	.115	.136	.88
2	469	$.14 < \bar{\phi}_m \leq .22$	.180	.173	-.40
3	353	$.22 < \bar{\phi}_m \leq .31$	.267	.314	2.03*
4	665	$.31 < \bar{\phi}_m \leq .39$	.354	.331	-1.24
5	992	$.39 < \bar{\phi}_m \leq .47$	.429	.433	.30
6	689	$.47 < \bar{\phi}_m \leq .55$	.509	.533	1.22
7	382	$.55 < \bar{\phi}_m \leq .64$	.591	.636	1.81
8	358	$.64 < \bar{\phi}_m \leq .72$	.676	.721	1.80
9	373	$.72 < \bar{\phi}_m \leq .80$	.758	.775	0.75
10	95	$.80 < \bar{\phi}_m \leq .88$	.823	.884	1.57
<i>Draws</i>					
1	8	$.08 < \bar{\phi}_m \leq .11$	.093	.125	n/a
2	46	$.11 < \bar{\phi}_m \leq .13$	.120	.065	-1.14
3	130	$.13 < \bar{\phi}_m \leq .15$	.143	.108	-1.15
4	245	$.15 < \bar{\phi}_m \leq .18$	.166	.184	.75
5	233	$.18 < \bar{\phi}_m \leq .20$	.191	.142	-1.9
6	288	$.20 < \bar{\phi}_m \leq .23$	.215	.191	-0.98
7	461	$.23 < \bar{\phi}_m \leq .25$	.239	.217	-1.12
8	885	$.25 < \bar{\phi}_m \leq .27$	.263	.258	-0.38
9	2085	$.27 < \bar{\phi}_m \leq .30$	.285	.299	1.42
10	179	$.30 < \bar{\phi}_m \leq .32$	.303	.341	1.09
<i>Away Wins</i>					
1	633	$.04 < \bar{\phi}_m \leq .11$	.081	.065	-1.46
2	623	$.11 < \bar{\phi}_m \leq .19$	.152	.133	-1.29
3	917	$.19 < \bar{\phi}_m \leq .27$	.232	.214	-1.28
4	994	$.27 < \bar{\phi}_m \leq .35$	.305	.27	-2.4*
5	464	$.35 < \bar{\phi}_m \leq .42$	.380	.36	-0.91
6	255	$.42 < \bar{\phi}_m \leq .50$	.460	.463	.08
7	288	$.50 < \bar{\phi}_m \leq .58$	.539	.552	.43
8	267	$.58 < \bar{\phi}_m \leq .66$	.61	.629	.64
9	102	$.66 < \bar{\phi}_m \leq .73$	.69	.735	.98
10	17	$.73 < \bar{\phi}_m \leq .81$	.76	.765	n/a

Note: Results from grouped z-tests where implicit probabilities $\phi_{i,m}^j$ for home wins, draws and away wins are separately divided into 10 groups with equal probability intervals. n_m is number of observations and $int.$ is the probability interval for each group. Significance codes: * = 0.05, ** = 0.01 and *** = 0.001.

The null hypothesis (eq. 19) that the proportion of winning bets does not differ from the average implicit probability is rejected only for 2 out of 30 groups. This could be seen as an indication of that the betting market is remarkably efficient – as the null hypothesis is not rejected for most groups, it does not seem to be a significant difference between odds and actual match outcomes. On the other hand, this could also be a result of the small number of observations, as most groups contain only a few hundred implicit probabilities, compared to the total number of 13680 implicit probabilities considered in the multinomial logit regression. Thus, lacking test power could be a contributing factor to the failure of rejecting the null hypothesis.

Furthermore, the relationship between the proportion of winning bets γ_m (y-axis) and the average implicit probability $\bar{\phi}_m$ (x-axis) is presented graphically in Figure 2.

Figure 2: *Graphical Illustrations of the Results from the First Category of Grouped Z-tests*



Note: Proportion of winning bets γ_m plotted against average implicit probability $\bar{\phi}_m$ in each group when implicit probabilities are divided into 10 groups with equal probability intervals, along with a dashed 45-degree line.

Even though the null hypothesis could not be rejected for most groups in this first category of grouped z-tests, it is worth noting that there are obvious similarities between the graphs for home and away wins (Figure 2) and the graphs illustrating the results from the multinomial logit model (Figure 1). For home wins, the proportion of winning bets is close to the average implicit probability up until high probability levels where the estimated relationship deviates from the dashed 45-degree line. Despite this, the betting market for home wins is not considered weak form inefficient as both the Wald test based on the multinomial logit regression and 9 out of 10 groups in the first category of grouped z-tests fail to reject the null hypothesis for market efficiency. For away wins, the proportion of winning bets seem to be consistently smaller than the average implicit probability for low probability levels, and consistently larger than the average implicit probability for high probability levels. This can be interpreted as that away favorites win more often than what is implied by their odds while away longshots win less often than what is implied by their odds, which is consistent with the evidence for a favorite-longshot bias for away wins found when performing the multinomial logit regression. For draws, the estimated relationship between the proportion of winning bets and the average implicit probability is relatively close to the dashed 45-degree line except for a few probability levels where the implicit probability is larger than the proportion of winning bets.

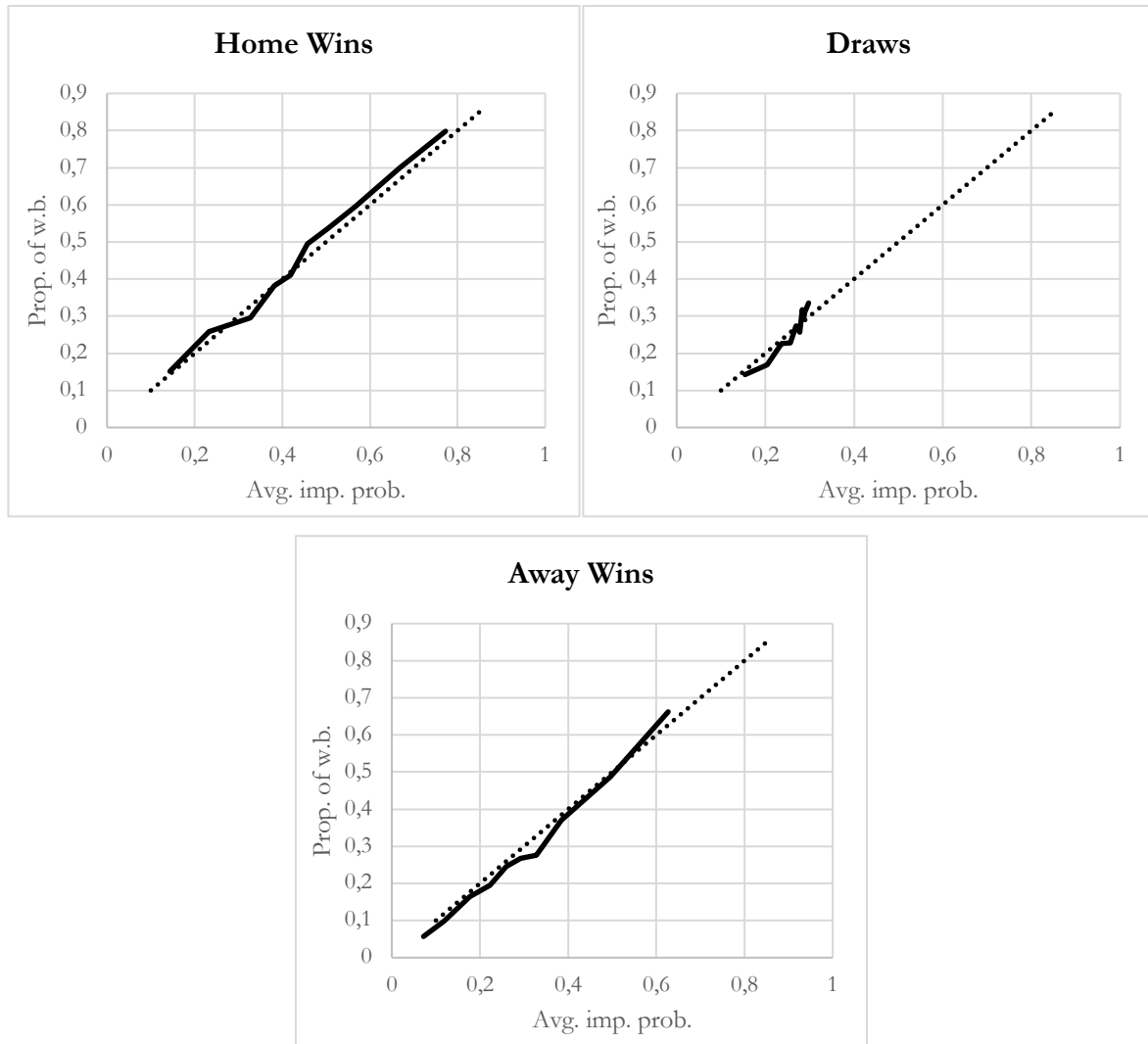
The second category of grouped z-tests, where implicit probabilities are divided into 10 groups with equal number of observations for home wins, draws and away wins separately, show similar results as the first category. More specifically, the null hypothesis cannot be rejected for most group (only the null hypothesis for 1 out of 30 groups is rejected), whereas the estimated relationship between the proportion of winning bets and the average implicit probability for away wins is consistent with the favorite-longshot bias. Detailed results as well as graphical illustrations of the relationship between the proportion of winning bets and the average implicit probability for each group given by this grouping procedure are presented in Table 5 and Figure 3, respectively.

Table 5: Results from the Second Category of Grouped Z-tests

m	n_m	$int.$	$\bar{\phi}_m$	γ_m	z_m
<i>Home Wins</i>					
1	456	$.06 < \bar{\phi}_m \leq .19$	.143	.151	.48
2	456	$.19 < \bar{\phi}_m \leq .29$	.233	.259	1.29
3	456	$.29 < \bar{\phi}_m \leq .36$	.327	.296	-1.42
4	456	$.36 < \bar{\phi}_m \leq .40$	.382	.382	-0.02
5	456	$.40 < \bar{\phi}_m \leq .44$	.419	.410	-0.4
6	456	$.44 < \bar{\phi}_m \leq .48$	.457	.496	1.65
7	456	$.48 < \bar{\phi}_m \leq .53$	.504	.535	1.34
8	456	$.53 < \bar{\phi}_m \leq .61$	.567	.596	1.27
9	456	$.61 < \bar{\phi}_m \leq .72$	.667	.700	1.46
10	456	$.72 < \bar{\phi}_m \leq .88$	.773	.798	1.31
<i>Draws</i>					
1	456	$.08 < \bar{\phi}_m \leq .18$	.154	.143	-0.69
2	456	$.18 < \bar{\phi}_m \leq .22$	.204	.169	-1.85
3	456	$.22 < \bar{\phi}_m \leq .25$	.237	.226	-0.56
4	456	$.25 < \bar{\phi}_m \leq .26$	.256	.228	-1.38
5	456	$.26 < \bar{\phi}_m \leq .27$	.269	.274	.26
6	456	$.27 < \bar{\phi}_m \leq .28$	.277	.257	-0.98
7	456	$.280 < \bar{\phi}_m \leq .285$	.283	.318	1.67
8	456	$.285 < \bar{\phi}_m \leq .288$	.286	.292	.26
9	456	$.288 < \bar{\phi}_m \leq .291$	.289	.311	1.05
10	456	$.291 < \bar{\phi}_m \leq .321$	.298	.336	1.77
<i>Away Wins</i>					
1	456	$.04 < \bar{\phi}_m \leq .09$	.072	.057	-1.22
2	456	$.09 < \bar{\phi}_m \leq .15$	.119	.099	-1.34
3	456	$.15 < \bar{\phi}_m \leq .20$	.177	.164	-0.68
4	456	$.20 < \bar{\phi}_m \leq .24$	.222	.195	-1.38
5	456	$.24 < \bar{\phi}_m \leq .28$	.259	.246	-0.65
6	456	$.28 < \bar{\phi}_m \leq .31$	.292	.268	-1.15
7	456	$.31 < \bar{\phi}_m \leq .35$	.328	.276	-2.3*
8	456	$.35 < \bar{\phi}_m \leq .43$	.384	.371	-0.59
9	456	$.43 < \bar{\phi}_m \leq .56$	.495	.487	-0.35
10	456	$.56 < \bar{\phi}_m \leq .81$	.627	.662	1.56

Note: Results from grouped z-tests where implicit probabilities $\phi_{i,m}^j$ for home wins, draws and away wins are separately divided into 10 groups with equal number of observations, n_m . *int.* is the probability interval for each group. Significance codes: * = 0.05, ** = 0.01 and *** = 0.001.

Figure 3: Graphical Illustrations of the Results from the Second Category of Grouped Z-tests



Note: Proportion of winning bets γ_m plotted against average implicit probability $\bar{\phi}_m$ in each group when implicit probabilities are divided into 10 groups with equal number of observations, along with a dashed 45-degree line.

The results from the third category of z-tests, where implicit probabilities for home wins, draws and away wins are pooled together and then divided into 10 groups with equal probability intervals, are presented in Table 6. It is shown that the null hypothesis that the proportion of winning bets does not differ from the average implicit probability can only be rejected for 1 out of 10 groups. Once again, this could be seen as evidence for that the betting market is weak form efficient when home wins, draws and away wins are considered jointly. However, it could also be a result of lacking test power due to the relatively few number of observations – even though the pooled grouping procedure considerably increases the number of implicit probabilities in most groups compared to the z-tests when home wins, draws and away wins are considered separately, each group still includes significantly less implicit probabilities than the total number included in the multinomial logit regression. Furthermore, plotting the relationship between the proportion of winning bets γ_m

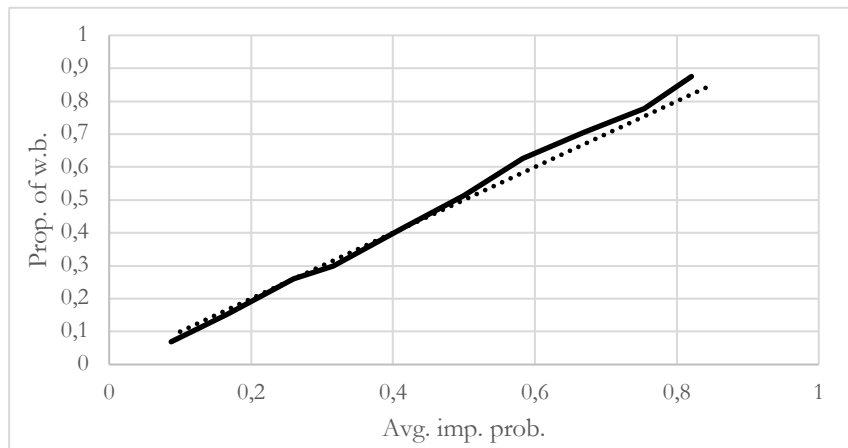
(y-axis) and the average implicit probability $\bar{\phi}_m$ (x-axis) (see Figure 4), it is noted that the estimated line follows the dashed 45-degree line rather closely up until high probability levels where it starts to deviate. Thus, the relationship between proportions of winning bets and average implicit probabilities when home wins, draws and away wins are pooled together have similarities with the same relationship for home wins, when home wins, draws and away wins are considered separately.

Table 6: Results from the Third Category of Grouped Z-tests

m	n_m	$int.$	$\bar{\phi}_m$	γ_m	z_m
1	796	$.04 < \bar{\phi}_m \leq .12$	.087	.07	-1.7
2	1804	$.12 < \bar{\phi}_m \leq .21$	.166	.155	-1.26
3	3397	$.21 < \bar{\phi}_m \leq .29$	.26	.262	.22
4	1330	$.29 < \bar{\phi}_m \leq .37$	.316	.304	-.96
5	1444	$.37 < \bar{\phi}_m \leq .46$	.415	.416	.07
6	1028	$.46 < \bar{\phi}_m \leq .54$	.5	.513	.8
7	758	$.54 < \bar{\phi}_m \leq .63$	.583	.625	2.35*
8	509	$.63 < \bar{\phi}_m \leq .71$	.67	.705	1.71
9	417	$.71 < \bar{\phi}_m \leq .80$	.754	.777	1.1
10	104	$.80 < \bar{\phi}_m \leq .88$	.821	.875	1.45

Note. Results from grouped z-tests where implicit probabilities $\phi_{i,m}^j$ for home wins, draws and away wins are pooled together and divided into 10 groups with equal probability intervals. When two outcomes from the same match are assigned to the same group, outcomes are randomly withdrawn from that group so that only one outcome from one match could be in the same group. n_m is number of observations in each group and $int.$ is the probability interval for each group. Significance codes: * = 0.05, ** = 0.01 and *** = 0.001.

Figure 4: Graphical Illustration of the Results from the Third Category of Grouped Z-tests



Note: Proportion of winning bets γ_m plotted against average implicit probability $\bar{\phi}_m$ in each group when home wins, draws and away wins are pooled together and divided into 10 groups with equal probability intervals, along with a dashed 45-degree line.

The results from the fourth category of z-tests, where implicit probabilities for all outcomes are pooled together and divided into 10 groups with equal number of observations, similarly shows that the null hypothesis can only be rejected for a few number of groups (2 out of 10) and that the proportion of winning bets is larger than the average implicit probability for high probability levels. Interestingly, the null hypothesis is rejected for the 2 groups with highest probability intervals ($m = 9, 10$), which can be seen as evidence for that odds for favorites are consistently mispriced which would further point towards tendencies for statistical weak form inefficiencies in this betting market. Detailed results are presented in Table 7 and Figure 5 below.

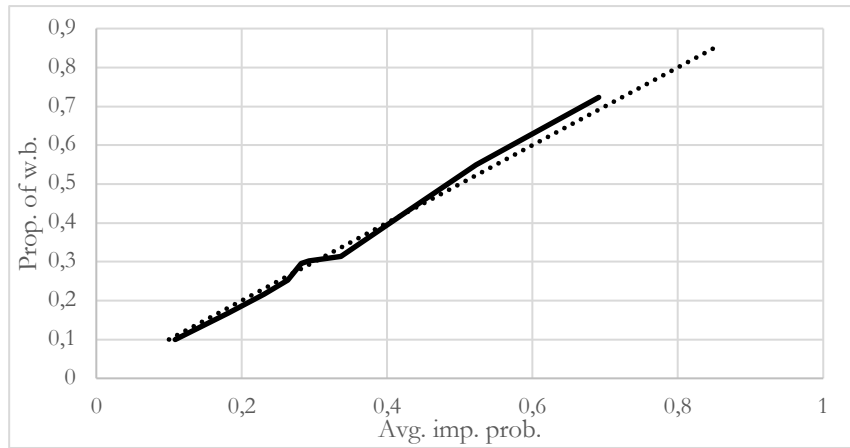
Table 7: *Results from the Fourth Category of Grouped Z-tests*

m	n_m	$int.$	$\bar{\phi}_m$	γ_m	z_m
1	1192	$.04 < \bar{\phi}_m \leq .15$	.108	.095	-1.43
2	1368	$.15 < \bar{\phi}_m \leq .21$	.181	.167	-1.27
3	1358	$.21 < \bar{\phi}_m \leq .25$	.231	.216	-1.23
4	1330	$.25 < \bar{\phi}_m \leq .27$	.263	.256	-0.65
5	1271	$.27 < \bar{\phi}_m \leq .29$	.282	.294	1.01
6	1170	$.29 < \bar{\phi}_m \leq .31$	.292	.308	1.15
7	884	$.31 < \bar{\phi}_m \leq .38$	.336	.308	-1.8
8	1368	$.38 < \bar{\phi}_m \leq .47$	.422	.423	.02
9	1368	$.47 < \bar{\phi}_m \leq .59$	.522	.549	1.98*
10	1368	$.59 < \bar{\phi}_m \leq .88$	.691	.723	2.55*

Note: Results from grouped z-tests where implicit probabilities $\phi_{i,m}^j$ for home wins, away wins and draws are pooled together and divided into 10 groups with equal number of observations⁵. When two outcomes from the same match are assigned to the same group, outcomes are randomly withdrawn from that group so that only one outcome from one match could be in the same group. n_m is number of observations in each group and $int.$ is the probability interval for each group. Significance codes: * = 0.05, ** = 0.01 and *** = 0.001.

⁵ Implicit probabilities are initially pooled together and divided into groups with equal number of observations. However, as outcomes are randomly withdrawn so that only one outcome from each match can be assigned to the same group, the total number of implicit probabilities are reduced from 13680 to 12677. This will lead to that the final groups do not have the same number of observations, which can be seen in Table 7 above.

Figure 5: *Graphical Illustration of the Results from the Fourth Category of Grouped Z-tests*



Note: Proportion of winning bets γ_m plotted against average implicit probability $\bar{\phi}_m$ in each group when home wins, draws and away wins are pooled together and divided into 10 groups with equal number of observations, along with a dashed 45-degree line.

7. Discussion

7.1. Interpretation and Implications of Results

The results of this study show that the aggregate fixed-odds betting market for Premier League football matches is statistically weak form inefficient. This means that prices of odds do not reflect all available information on historical prices of odds and therefore are biased predictors of actual match outcomes. Furthermore, odds for away wins are shown to be characterized by a favorite-longshot bias. In other words, highly likely away wins (i.e. when the away team is considered the favorite in a given match) are undervalued by the market, while unlikely away wins (i.e. when the away team is considered the longshot in a given match) are overvalued by the market. The implication of this is that odds for away favorite wins reflect a probability higher than the estimated true probability of such outcomes, while odds for away longshot wins reflect a probability lower than the estimated true probability. From the perspective of the sophisticated bettor, away longshot bets should therefore be considerably less attractive than away favorite bets. In contrast, odds for home wins and draws are shown to be priced rather weak form efficiently, meaning that they reflect all available information on historical odds. Therefore, the sophisticated bettor should on average be indifferent between betting on a favorite and a longshot for these outcomes.

The evidence for a statistically weak form inefficient market is consistent with findings in several other betting markets in general (see e.g. Snyder, 1978; Asch *et al.*, 1982, 1984; Tuckwell, 1983) as well as in some football betting markets in particular (see e.g. Pope and Peel, 1989; Kuypers, 2000; Nyberg, 2014). Furthermore, the evidence for a favorite-longshot bias is consistent with previous findings in a wide range of betting markets, including some betting markets for football matches (see section 3 for an extensive review). As described above, the favorite-longshot bias is found only for away wins in this study, while odds for home wins and draws show no indications of a favorite-longshot bias. This finding is somewhat related to the so-called away-favorite bias, implying that away favorites are undervalued to a larger extent than home favorites (Vlastakis *et al.*, 2009; Nyberg, 2014). While this is true also for the results in this study, there is a considerable difference between that home favorites are *less* undervalued than away favorites and that home favorites are not undervalued at all. To the best of our knowledge, the circumstance that odds only for away wins are mispriced accordingly to the favorite-longshot bias, whereas efficiency cannot be rejected for home win and draw odds, has not been found in any previous study.

There could be several different explanations for these results. As discussed in section 3.4., the favorite-longshot bias has previously been explained by that bettors are either risk-loving or irrational which makes them overbet longshots. In the case of fixed-odds betting markets, such tendencies would force bookmakers to lower the odds for these outcomes, and correspondingly

raise the odds for favorite outcomes, in order to avoid substantial losses. Consequently, longshots become overvalued while favorites become undervalued, resulting in the emergence of the favorite-longshot bias. Furthermore, bookmakers may have other incentives to price odds so that they systematically deviate from the estimated true probabilities of certain outcomes. These incentives include e.g. that bookmakers are better predictors of outcomes than individual bettors which can be exploited by intentionally set biased odds in order to make profits, that bookmakers raise favorite bets to attract bettors as these are more competitive than longshot bets or that bookmakers adjust odds in response to the presence of insider bettors in the market.

Even though all of the above explanations provide reasonable theories about how the favorite-longshot may emerge in betting markets, none of them could explain why the bias would be present only for away wins. The away-favorite bias is explained with that the home field advantage (i.e. that it in general is advantageous to play at home) is overestimated which leads to that away favorites are undervalued to a larger extent than home favorites (Vlastakis *et al.*, 2009). This theory alone could however not be used to explain why the favorite-longshot bias exists for away wins while odds for draws and home wins are efficient. The reason is that, if the home field advantage was overestimated, odds for home wins would as well be mispriced which should be manifested in inefficiencies. Instead, another possible explanation goes as follows.

If it is assumed that there is a home field advantage⁶ and that bookmakers account for this in their price-setting of odds, away win longshots would on average be more of a clear longshot outcome than home win longshots (i.e. the probability that away longshots win should on average be less than the probability that home longshots win). Prospect theory now suggests that small probabilities tend to be overweighted, implying that the subjective estimation of a small probability is larger than the actual probability (Kahneman and Tversky, 1979). As the probability for an away longshot win on average is smaller than the probability for a home longshot win, it could be so that this subjective misestimation is larger for away longshots than for home longshots. If the probabilities for home longshots moreover are large enough, these probabilities may not even be misestimated to a significant extent. Consequently, away longshot wins are overvalued by the market while home longshot wins are not, and conversely away favorite wins are undervalued (when the longshot outcome is overvalued the favorite outcome must be undervalued as the probabilities of all outcomes in a given match should sum to one) while home favorites are not. In

⁶ The data set used in this study supports the existence of a home field advantage as the frequency of home wins is equal to 46.23%, which can be compared to the frequencies of draws and away wins which are 25.53% and 28.24%, respectively.

the end, the implication is that prices of away win odds are inefficient and characterized by a favorite-longshot bias, while prices of home win odds in contrast are efficient.

7.2. Limitations of Study

There are three main limitations of this study which may affect the validity of the results. Firstly, it is assumed that bookmaker margins are equally distributed across match outcomes, i.e. that the margin is equal for home wins, draws and away wins. This assumption is necessary to make in order to calculate the implicit probabilities based on the odds as there is no data neither on bookmaker margins nor distribution of bets available. If this assumption does not hold, the implicit probabilities calculated are incorrect which would make the results less valid. However, as pointed out by Forrest *et al.* (2005), overall margins in betting markets have decreased as a result of intensified competition, which makes this assumption less problematic.

Secondly, it is assumed that odds are determined on the basis of true probabilities, i.e. bookmakers aim to set odds that as accurately as possible reflect the true probability of a certain outcome occurring. As pointed out by Levitt (2004), bookmakers may intentionally set odds which deviate from true probabilities as they are better predictors of match outcomes than individual bettors, which would imply that this assumption does not hold. On the other hand, once again as a consequence of intensified competition, the incentives for bookmakers to set odds which deviate too much from true probabilities have decreased in order to avoid heavy losses.

Thirdly, although this study uses a large data set including prices of odds over a relatively long period of time, it is always favorable to include as much information as possible when testing for market efficiency as such tests investigate whether all available information is reflected in prices. In particular, given that there is a home field advantage implying that home favorites on average should be more clear favorites than away favorites, and that away longshots on average should be more clear longshots than home longshots, there may be a lacking number of observations for clear away favorites and home longshots, which could have impact on the results. Thus, in order to perfectly validate the results found in this study, it would be advantageous to perform the same tests on a substantially larger data set, perhaps as well including odds for football matches in several different countries.

8. Concluding Remarks

Ever since Griffith (1949) found evidence for the favorite-longshot bias, the efficiency of prices in betting markets has been given lots of consideration in the academic literature. The results have overall been somewhat ambiguous – in betting markets for certain sports, such as horse racing, most conclusions point towards the existence of a favorite-longshot bias, while in betting markets for other sports, odds appear to be priced rather efficiently. This study aims to evaluate the statistical weak form market efficiency of odds in the fixed-odds betting market for English Premier League football matches. ‘Statistical’ efficiency refers to if prices of odds reflect information available for market participants (in contrast to economic efficiency, which concerns if it is possible to find profitable investment strategies in betting markets), while ‘weak form’ efficiency narrows the definition of information to only information on historical prices of odds (in contrast to semi-strong or strong form efficiency which expand the definition to include as well all information publicly available in the market and insider information, respectively). This is done by performing a multinomial logit regression as well as grouped z-tests with four different grouping procedures, using a data set consisting of odds from 11 different bookmakers for all Premier League matches between the 2006/2007 season up until the 2017/2018 season, and thereby answering the research questions,

1. *Is the fixed-odds betting market for English Premier League football matches statistically weak form efficient?*
2. *Does this betting market suffer from a favorite-longshot bias, i.e. are low probability outcomes overvalued and high probability outcomes undervalued?*

The results show that the aggregate fixed-odds betting market for Premier League football matches is statistically weak form inefficient. In other words, implicit probabilities given by odds seem to aggregately deviate from estimated objective probabilities. Furthermore, evidence for the favorite-longshot bias is found, however, only for away win odds, while efficiency cannot be rejected for home win and draw odds. To the best of our knowledge, this particular finding, i.e. that the favorite-longshot bias exists only for away wins, is the first of its kind and is coined the ‘away-favorite-longshot bias’. Even though this study does not aim to evaluate the underlying factors behind why certain betting markets are characterized by the favorite-longshot bias, a suggested explanation for the away-favorite-longshot-bias in particular is presented. Assuming a home field advantage and applying ideas from prospect theory (Kahneman and Tversky, 1979), it could be so that the probabilities for away longshot wins are overweighted to a larger extent than

the probabilities for home longshot wins, as the former probabilities on average are smaller than the latter. If the probabilities for home win longshots furthermore are large enough, they may not be overweighted to a significant extent, resulting in that home win odds are priced rather efficiently while away odds are characterized by the favorite-longshot bias.

This study aims to contribute to the current body of knowledge in two separate ways. Firstly, betting markets have grown significantly during the last couple of years, partly because of the emergence of online betting, which most likely has had an impact both on the behavior of bettors and the incentives of bookmakers. Consequently, it is relevant to re-examine market efficiency and the favorite-longshot bias in the light of these rapid changes. Secondly, previous research has come to rather ambiguous conclusion about whether betting markets for football matches indeed are efficient, if they are characterized by a favorite-longshot bias or any other type of bias. Thus, by using a large data set and applying several sophisticated statistical models, the study aims to give further insight about the characteristics of this particular betting market.

Regarding the first intended contribution, the findings of this study suggests that the efficiency of prices may have been affected by the transformed betting industry environment, as a new version of the favorite-longshot bias is found to be present. Thus, in addition to the explanation for the away-favorite-longshot bias proposed above, this anomaly may as well be a result of bettors' and bookmakers' changed behavior in response to a reshaped reality of betting markets. Because even though the above explanation may hold, it fails to explain why betting markets have not been characterized by the away-favorite-longshot bias previously. In order to do this, explanations such as that bettors have become more informed (as information is more accessible), spend more money and commit more errors as online betting becomes more popular, or that bookmakers' incentives to set odds which deviate from true probabilities most likely has decreased due to intensified competition, may be included. However, at this point, it would be careless to draw any conclusion about how these factors interact and in the end affect the efficiency of prices in betting markets.

Regarding the second intended contribution, the results give further insight about the characteristics of betting markets for football matches as a new version of the favorite-longshot bias is exposed. On the other hand, ironically, the results as well enhance the ambiguity about in what specific ways football betting markets are inefficient. Thus, in the end, it appears to be difficult to accurately determine consistent characteristics of such markets. This could certainly be a result of that football betting markets e.g. in different countries and across different time periods actually differ regarding how efficiently prices of odds are set, but could also be a question of methodology

(as a wide range of different models previously have been applied, see section 3.3. for details) or small data samples.

Lastly, one may ask – why does any of this even matter? Firstly, betting markets make up an enormous industry with large annual turnovers, which furthermore are expected to increase significantly during the coming years. Thus, tendencies for systematic mispricing of odds will be of highest interest both for bettors who seek to earn profits (even though it has been shown to be difficult) and for bookmakers who aim to set odds that maximize profits (assuming that this is their main incentive). Secondly, and more importantly, prices of odds in betting markets may expose insights about human behavior. As pointed out by Thaler and Ziemba (1988), betting markets may be better suited for testing rationality and market efficiency than other markets as each asset has a well-defined termination point where its value becomes certain. Thus, findings that prices of odds in betting markets are set inefficiently may as well say something about agents' irrationality, about how they act and make decisions in uncertain situations involving risk and money, about the incentives of bookmakers and the aggregate knowledge of the general public. Price-setting of odds in various forms of betting markets, studied in different settings and from different perspectives, should therefore be further treated in the academic literature in order to better understand human behavior and decision-making under risk.

9. Suggestions for Further Research

Two main suggestions for further research are presented. Firstly, considering the persistent ambiguity regarding market efficiency in football betting markets, it is suggested that further tests make use of significantly larger data sets including odds from several different countries in order to evaluate whether a general pattern can be found. Preferably, new methods with the ability to handle the above stated assumptions (i.e. margins distribution, true probabilities) are developed, however, this may be more of a data-related issue. Secondly, underlying factors behind the different forms of inefficiencies, partly in the light of the rapidly changing betting industry, is suggested to be studied in various kinds of betting markets. More specifically, it would be interesting to further study how the emergence of online betting has affected both bettors' behavior and bookmakers' incentives, and in turn which implications this has for the price-setting of odds. Additionally, possible explanations for the away-favorite-longshot bias found in this study are suggested to be considered and evaluated.

10. References

- Akhtar, S., Scarf, P. 2012. 'Forecasting Test Cricket Match Outcomes in Play', *International Journal of Forecasting*, Vol. 28, pp. 632-643.
- Ali, M. M. 1977. 'Probability and Utility Estimates for Racetrack Bettors', *Journal of Political Economy*, Vol. 85, No. 4 (August, 1977), pp. 803-815.
- Andrikogiannopoulou, A., and Papakonstantinou, F. 2011. 'Market Efficiency and Behavioral Biases in the Sports Betting Market', *HEC Geneva, working paper*.
- Asch, P., Malkiel, B. G., Quandt, R. E. 1982. 'Racetrack Betting and Informed Behavior', *Journal of Financial Economics*, Vol. 10, pp. 187-194.
- Asch, P., Malkiel, B. G., Quandt, R. E. 1984. 'Market Efficiency in Racetrack Betting', *The Journal of Business*, Vol. 57, No. 2 (April, 1984), pp. 165-175.
- Beem, B., and Mikler, J. 2011. 'National Regulations for a Borderless Industry: US versus UK Approaches to Online Gambling', *Policy and Society*, Vol. 30 (2011), pp. 161-174.
- Berkowitz, J. P., Depken, C. A., and Gandar, J. M. 2017. 'A Favorite-Longshot Bias in Fixed-Odds Betting Markets: Evidence From College Basketball and College Football', *The Quarterly Review of Economics and Finance*, Vol. 63, pp. 233-239.
- Busche, K. 1994. 'Efficient Market Results in an Asian Setting', in *Efficiency of Racetrack Betting Markets*, eds. Hausch, D., Lo, V., and Ziemba, W. T., pp. 615-616, New York: Academic Press.
- Busche, K., Hall, C. D. 1988. 'An Exception to the Risk Preference Anomaly', *The Journal of Business*, Vol. 61, No. 3 (July 1988), pp. 337-346.
- Cain, M., Law, D., and Peel, D. 2000. 'The Favourite-Longshot Bias and Market Efficiency in UK Football Betting', *Scottish Journal of Political Economy*, Vol. 47, No. 1 (February, 2000), pp. 25-36.
- Cain, M., Law, D., and Peel, D. 2003. 'The Favourite-Longshot Bias, Bookmaker Margins and Insider Trading in a Variety of Betting Markets', *Bulletin of Economic Research*, Vol. 55, pp. 263-273.
- Camerer, C. F. 1998. 'Prospect Theory in the Wild: Evidence from the Field', Social Science Working Paper, No. 1037, Division of the Humanities and Social Sciences, California Institute of Technology.
- Cortis, D. 2015. 'Expected Values and Variances in Bookmaker Payouts: A Theoretical Approach Towards Setting Limits On Odds', *The Journal of Prediction Markets*, Vol. 9, No. 1, pp. 1-14.

- Crafts, N. F. R. 1985. 'Some Evidence of Insider Knowledge in Horse Race Betting in Britain', *Economica*, Vol. 52, pp. 295-304.
- Deschamps, B., Gergaud, O. 2007. 'Efficiency in Betting Markets: Evidence from English Football', *The Journal of Prediction Markets*, Vol. 1, pp. 61-73.
- Deutscher, C., Ötting, M., Schneemann, S., and Scholten, H. 2019. 'The Demand for English Premier League Soccer Betting', *Journal of Sports Economics*, Vol. 20, No. 4, pp. 556-579.
- Direr, A. 2013. 'Are Betting Markets Efficient? Evidence from European Football Championships', *Applied Economics*, Vol. 45, pp. 343-356.
- Dixon, M. J., and Coles, S. G. 1997. 'Modelling Association Football Scores and Inefficiencies in the Football Betting Market', *Journal of the Royal Statistical Society*, Vol. 46, No. 2, pp. 265-280.
- Dixon, M. J., and Pope, P. F. 2004. 'The Value of Statistical Forecasts in the UK Association Football Betting Market', *International Journal of Forecasting*, Vol. 20, pp. 697-711.
- Dowie, J. 1976. 'On the Efficiency and Equity of Betting Markets', *Economica*, Vol. 43, pp. 139-150.
- Fama, E. F. 1970. 'Efficient Capital Markets: A Review of Theory and Empirical Work', *The Journal of Finance*, Vol. 25, No. 2, pp. 383-417.
- Figlewski, S. 1979. 'Subjective Information and Market Efficiency in a Betting Market', *Journal of Political Economy*, Vol. 87, No. 1 (February, 1979), pp. 75-88.
- Forrest, D., Goddard, J., and Simmons, R. 2005. 'Odds-Setters as Forecasters: The Case of English Football', *International Journal of Forecasting*, Vol. 21, pp. 551-564.
- Forrest, D., and McHale, I. 2007. 'Anyone for Tennis (Betting)?', *The European Journal of Finance*, Vol. 13, No. 8, pp. 751-768.
- Forrest, D., and Simmons, R. 2000. 'Forecasting Sport: the Behavior and Performance of Football Tipsters', *International Journal of Forecasting*, Vol. 16, pp. 317-331.
- Franck, E., Verbeek, E., and Nüesch, S. 2010. 'Prediction Accuracy of Different Market Structures – Bookmakers versus a Betting Exchange', *International Journal of Forecasting*, Vol. 26, pp. 448-459.
- Gabriel, P. E., and Marsden, J. R. 1990. 'An Examination of Market Efficiency in British Racetrack Betting', *Journal of Political Economy*, Vol. 98, No. 4 (August, 1990), pp. 874-885.
- Gandar, J. M., Zuber, R. A., and Johnson, R. S. 2001. 'Searching for the Favourite-Longshot Bias down under: an examination of the New Zealand pari-mutuel betting market', *Applied Economics*, Vol. 33, No. 13, pp. 1621-1629.

- Gandar, J. M., Zuber, R. A., Johnson, R. S., and Dare, W. 2002. 'Re-examining the betting market on Major League Baseball Games: is there a reverse favourite-longshot bias?', *Applied Economics*, Vol. 34, pp. 1309-1317.
- Gandar, J. M., Zuber, R. A., O'Brien, T., and Russo, B. 1988. 'Testing Rationality in the Point Spread Betting Market', *The Journal of Finance*, Vol. 43, No. 4 (September, 1988), pp. 995-1008.
- Gainsbury, S., Wood, R., Russel, A., Hing, N., and Baszczynski, A. 2012. 'A Digital Revolution: Comparison of Demographic Profiles, Attitudes and Gambling Behavior of Internet and non-Internet Gamblers', *Computers in Human Behavior*, Vol. 28, pp. 1388-1398.
- Gainsbury, S., Wood, R., Russel, A., Hing, N., Baszczynski, A., and Lubman, D. 2015. 'How the Internet is Changing Gambling: Findings from an Australian Prevalence Survey', *Journal of Gambling Studies*, Vol. 31, No. 1, pp. 1-15.
- Goddard, J. 2005. 'Regression Models for Forecasting Goals and Match Results in Association Football', *International Journal of Forecasting*, Vol. 21, pp. 331-340.
- Goddard, J., and Asimakopoulous, I. 2004. 'Forecasting Football Results and the Efficiency of Fixed-Odds Betting', *Journal of Forecasting*, Vol. 23, pp. 51-66.
- Golec, J., and Tamarkin, M. 1991. 'The Degree of Inefficiency in the Football Betting Market', *Journal of Financial Economics*, Vol. 30, pp. 311-323.
- Gray, P. K., and Gray, S. F. 1997. 'Testing Market Efficiency: Evidence from the NFL Sports Betting Market', *The Journal of Finance*, Vol. 52, No. 4 (September, 1997), pp. 1725-1737.
- Griffith, R. M. 1949. 'Odds Adjustments by American Horse-Race Bettors', *The American Journal of Psychology*, Vol. 62, No. 2 (April, 1949), pp. 290-294.
- Hausch, D. B., Ziemba, W. T., and Rubinstein, M. 1981. 'Efficiency of the Market for Racetrack Betting', *Management Science*, Vol. 27, No. 12 (December, 1981), pp. 1435-1452.
- Henery, R. J. 1985. 'On the Average Probability of Losing Bets on Horses with Given Starting Price Odds', *Journal of Royal Statistical Society*, Vol. 148, No. 4, pp. 342-349.
- Hodges, S. D., Tompkins, R. G., and Ziemba, W. T. 2003. 'The Favorite / Long-shot Bias in S&P 500 and FTSE 100 Index Futures Options: The Return to Bets and the Cost of Insurance', *EFA 2003 Annual Conference Paper*, No. 135.
- Jullien, B., and Salanié, B. 2000. 'Estimating Preferences under Risk: The Case of Racetrack Bettors', *Journal of Political Economy*, Vol. 108, No. 3, pp. 503-530.
- Kahneman, D., and Tversky, A. 1979. 'Prospect Theory: An Analysis of Decision Under Risk', *Econometrica*, Vol. 47, No. 2 (March, 1979), pp. 263-292.

- Kairouz, S., Paradis, C., and Nadeau, L. 2012. 'Are online gamblers more at risk than offline gamblers?', *Cyberpsychology, Behavior and Social Networking*, Vol. 15, No. 3, pp. 175-180.
- Koning, R. H. 2012. 'Regression Tests and the Efficiency of Fixed Odds Betting Markets', *International Journal of Sports Finance*, Vol. 7, pp. 262-274.
- Kuypers, T. 2000. 'Information and Efficiency: An Empirical Study of a Fixed Odds Betting Market', *Applied Economics*, Vol. 32, pp. 1353-1363.
- Levitt, S. D. 2004. 'Why are Gambling Markets Organised so Differently From Financial Markets?', *The Economic Journal*, Vol. 114, No. 495 (April, 2004), pp. 223-246.
- McGlothlin, W. H. 1956. 'Stability of Choices Among Uncertain Alternatives', *The American Journal of Psychology*, Vol. 69, No. 4 (December, 1956), pp. 604-615.
- Montes, K. S., and Weatherly, J. N. 2017. 'Differences in the Gambling Behavior of Online and Non-Online Student Gamblers in a controlled Laboratory Environment', *Journal of Gambling Studies*, Vol. 33, No. 1, pp. 85-97.
- Nyberg, H. 2014. 'A Multinomial-Logit Based Statistical Test of Association Football Betting Market Efficiency', *Helsinki Center of Economic Research: Discussion Paper*, No. 380 (June, 2014).
- Ottaviani, M., and Sørensen, P. N. 2005. 'Parimutuel versus Fixed-Odds Markets', Working Paper, Department of Economics, University of Copenhagen: Copenhagen, Denmark.
- Paul, R. J., and Weinbach, A. P. 2005. 'Bettor Misperceptions in the NBA', *Journal of Sports Economics*, Vol. 6, No. 4 (November, 2005), pp. 390-400.
- Paul, R. J., Weinbach, A. P., Borghesi, R., and Wilson, M. 2008. 'Using Betting Market Odds to Measure the Uncertainty of Outcome in Major League Baseball', *International Journal of Sport Finance*, Vol. 4, pp. 255-263.
- Pope, P. F., and Peel, D. A. 1989. 'Information, Prices and Efficiency in a Fixed-Odds Betting Market', *Economica*, Vol. 56, No. 223 (August, 1989), pp. 323-341.
- Quandt, R. E. 1986. 'Betting and Equilibrium', *The Quarterly Journal of Economics*, Vol. 101, No. 1 (February, 1986), pp. 201-208.
- Rosett, R. N. 1965. 'Gambling and Rationality', *Journal of Political Economy*, Vol. 73, No. 6 (December, 1965), pp. 595-607.
- Rue, H., and Salvesen, O. 2000. 'Prediction and Retrospective Analysis of Soccer Matches in a League', *Journal of the Royal Statistical Society*, Vol. 49, No. 3, pp. 399-418.
- Ruhm, D. L. 2003. 'Distribution-based pricing formulas are not arbitrage-free', *Proceedings of the Casualty Actuarial Society*, Vol. 90, pp. 97-129.
- Sauer, R. D. 1998. 'The Economics of Wagering Markets', *Journal of Economic Literature*, Vol. 36, pp. 2021-2064.

- Sauer, R. D., Brajer, V., Ferris, S. P., and Marr, M. W. 1988. 'Hold Your Bets: Another Look at the Efficiency of the Gambling Market for National Football League Games', *Journal of Political Economy*, Vol. 96, No. 1, pp. 206-213.
- Schnytzer, A., and Weinberg, G. 2008. 'Testing for Home Team and Favorite Biases in the Australian Rules Football Fixed-Odds and Point Spread Betting Markets', *Journal of Sports Economics*, Vol. 9, No. 2 (April, 2008), pp. 173-190.
- Shin, H. S. 1991. 'Optimal Betting Odds Against Insider Traders', *The Economic Journal*, Vol. 101, No. 408 (September, 1991), pp. 1179-1185.
- Shin, H. S. 1992. 'Prices of State Contingent Claims with Insider Traders, and the Favourite-Longshot Bias', *The Economic Journal*, Vol. 102, No. 411 (March, 1992), pp. 426-435.
- Shin, H. S. 1993. 'Measuring the Incidence of Insider Trading in a Market for State-Contingent Claims', *The Economic Journal*, Vol. 103, No. 420 (September, 1993), pp. 1141-1153.
- Shmanske, S. 2005. 'Odds-Setting Efficiency in Gambling Markets: Evidence from the PGA Tour', *Journal of Economics and Finance*, Vol. 29, No. 3, pp. 391-402.
- Snowberg, E., and Wolfers, J. 2010. 'Explaining the Favorite-Longshot Bias: Is it Risk-Love or Misperceptions?', *Journal of Political Economy*, Vol. 118, No. 4, pp. 723-746.
- Snyder, W. W. 1978. 'Horse Racing: Testing the Efficient Markets Model', *The Journal of Finance*, Vol. 33, No. 4 (September, 1978), pp. 1109-1118.
- Statista, 2019a. 'Sports Betting and Gambling Market/Industry – Statistics and Facts', *Statista*, Accessed 2019-03-22, <https://www.statista.com/topics/1740/sports-betting/>.
- Statista, 2019b. 'Gambling industry in the United Kingdom (UK) – Statistics and Facts', *Statista*, Accessed 2019-04-09, <https://www.statista.com/topics/3400/gambling-industry-in-the-united-kingdom-uk/>.
- Statista, 2019c. 'Premier League – Statistics and Facts', *Statista*, Accessed 2019-04-09, <https://www.statista.com/topics/1773/premier-league/>.
- Terrell, D., and Farmer, A. 1996. 'Optimal Betting and Efficiency in Parimutuel Betting Markets With Information Costs', *The Economic Journal*, Vol. 106, pp. 846-868.
- Thaler, R. H., and Ziemba, W. T. 1988. 'Parimutuel Betting Markets: Racetracks and Lotteries', *Journal of Economic Perspectives*, Vol. 2, No. 2 (Spring 1988), pp. 161-174.
- Tuckwell, R. H. 1983. 'The Thoroughbred Gambling Market: Efficiency, Equity and Related Issues', *Australian Economic Papers*, Vol. 22, No. 40, pp. 106-118.
- U.K. Gambling Commission, 2018. 'Industry Statistics', *U.K. Gambling Commission*, Accessed 2019-04-09, <https://www.gamblingcommission.gov.uk/news-action-and-statistics/Statistics-and-research/Statistics/Industry-statistics.aspx>

- Vaughan Williams, L., and Paton, D. 1997. 'Why is there a favourite-longshot bias in British Racetrack Betting Markets?', *The Economic Journal*, Vol. 107, pp. 150-158.
- Vecer, J., Ichiba, T., and Laudanovic, M. 2006. 'Parallels Between Betting Contracts and Credit Derivatives: Lessons Learned from FIFA World Cup 2006 Betting Markets', Working Paper, Department of Statistics, Columbia University.
- Vergin, R. C., and Scriabin, M. 1978. 'Winning Strategies for Wagering on National Football League Games', *Management Science*, Vol. 24, No. 8, pp. 789-875.
- Vlastakis, N., Dotsis, G., and Markellos, R. N. 2009. 'How Efficient is the European Football Betting Market? Evidence from Arbitrage and Trading Strategies', *Journal of Forecasting*, Vol. 28, pp. 426-444.
- Walls, W. D., and Busche, K. 2003. 'Broken odds and the favourite-longshot bias in parimutuel betting: a direct test', *Applied Economic Letters*, No. 10, Vol. 5, pp. 311-314.
- Weitzman, M. L. 1965. 'Utility Analysis and Group Behavior: An Empirical Study', *Journal of Political Economy*, Vol. 73, No. 1, pp. 18-26.
- Wood, R. T., and Williams, R. J. 2011. 'A Comparative Profile of the Internet Gambler: Demographic Characteristics, Game-Play Patterns, and Problem Gambling Status', *New Media & Society*, Vol. 13, No. 7, pp. 1123-1141.
- Woodland, L. M., and Woodland, B. M. 1994. 'Market Efficiency and the Favorite-Longshot Bias: The Baseball Betting Market', *The Journal of Finance*, Vol. 49, No. 1 (March, 1994), pp. 269-279.
- Woodland, L. M., and Woodland, B. M. 2001. 'Market Efficiency and Profitable Wagering in the National Hockey League: Can Bettors Score on Longshots?', *Southern Economic Journal*, Vol. 67, No. 4 (April, 2001), pp. 983-995.
- Woodland, L. M., and Woodland, B. M. 2003. 'The Reverse Favourite-Longshot Bias and Market Efficiency in Major League Baseball: An Update', *Bulletin of Economic Research*, Vol. 55, No. 2, pp. 113-123.
- Zuber, R. A., Gandar, J. M., and Bowers, B. D. 1985. 'Beating the Spread: Testing the Efficiency of the Gambling Market for National Football League Games', *Journal of Political Economy*, Vol. 93, No. 4 (August, 1985), pp. 800-806.