Aalto University
School of Business

IS THERE A FAVORITE-LONGSHOT BIAS IN ONLINE FOOTBALL BETTING MARKETS?

Master's Thesis
Thuy Trang Duong
Aalto University School of Business
Economics
Fall 2020

Author Thuy Trang Duong

Title of thesis Is there a favorite-longshot bias in online football betting markets?

Degree Master of Science in Economics and Business Administration

Degree programme Economics

Thesis advisor(s) Professor Marko Terviö

Year of approval 2020 **Number of pages** 70 **Language** English

Abstract

Online sports betting markets have significantly expanded during the last decades. Furthermore, because of its similarities with financial markets, sports betting markets have become popular empirical settings for researchers to study the working of information markets and the efficient market hypothesis. Although many studies have showed that those markets seem to surprisingly efficient, biases still exist. Of which, the favorite-longshot bias, the infamous regularity that betting odds underestimate the winning chance of favorites and overestimate that of longshots, has been documented in many different types of sports-betting markets.

In this thesis, we focus on football betting markets and study the favorite-longshot bias using online betting data from 11 different well-known football competitions in Europe during seasons 2016/17 and 2017/18. This is done by comparing the subjective probability implied by bookmakers' odds and the objective probability obtained from matches' results. Two methods that have been commonly used in literature are employed with different alternative specifications in the study.

In general, the result shows that the European online football betting markets are quite efficient. Odds provide a good prediction for the actual result of matches. The subjective probability implied by odds is not different from the objective probability obtained from the matches' results at a 5% significant level.

However, in these markets, some biases still exist, including the home-field bias and the favorite-longshot bias. More specifically, there is an existence of strong home-field bias: bookmakers tend to shorten the odds for the victory of away teams more relative to the odds for the victory of home teams. When taking into account this home-field bias, these online betting markets show the presence of a favorite-longshot bias in odds for betting on both victory of home and away teams but not on draw outcome, and this bias decreases as the match's date is approaching. The favorite-longshot bias in these markets persists even with changes in specifications.

Keywords Favorite - longshot bias, Sports betting market, Online betting

Contents

List of Figures

List of Tables

1 Introduction

Online betting markets are a booming industry that has been gaining worldwide popularity over the last decades. The European gaming betting association (EGBA) stated that the global market size for online betting was valued at approximately 45.12 billion euro in 2018 and was expected to expand from 2019 to 2025 at a compound annual growth rate of 11.5% (EGBA (2019)). Furthermore, in EU-28 countries alone, online gambling reached a total revenue of 22.2 billion euro, accounting for 23.2% of the total EU gambling market and 49.2% of the total global online gambling market. Global online gambling revenues have experienced double-digit growth since 2016, and this trend is set to persist through 2023. Year after year, online gambling takes a larger portion of the overall global gambling market. The EU market for online gambling is expected to reach 29.3 billion euro in 2022. The most common online gambling activity is sports betting, accounting for 42.5% of all activities. In addition to the importance of this industry for the economy, the betting market is an interesting field for researchers who study the operating of information markets as well as the implications for risk preferences and the degree of market participants' rationality.

Betting markets have long been a field of interest in the economics literature. Initially, sports betting markets have been seen as a better candidate than the traditional financial market to test the efficient-market hypothesis. As pointed out in Thaler and Ziemba (1988), betting markets are more suitable for testing the market efficiency and decision making under uncertainty. The reasons are that each bet has a well-defined termination point at which its value becomes certain, and the conditions, called "quick and repeated feedback", usually facilitate learning. Because of these characteristics, the betting markets are expected to be more efficient than the traditional financial markets: market prices (odds) provide the actual probabilities that events happen, and profitable betting strategies are nonexistent on that market.

Over the last few decades, the sports betting markets receive even more attention because it is one form of new and emerging markets: prediction markets. The prediction markets, also known as information markets, are markets in which participants trade on contracts which payoffs depend on unknown future events (Wolfers and Zitzewitz (2004)). These events could be anything, varying from future political events such as "will Donald Trump be reelected?' or 'UK to rejoin the EU as a full member before 2026?", financial contracts, science and technology events to results of sport events. Based on the efficient-market hypothesis, the market price, or odds in this context, will be the best predictor of the likelihood that an event happens because it reflects all available information. Therefore, no information can be used to improve this prediction. The most important application of these kinds of markets is their prediction power. According to Wolfers and Zitzewitz (2004), although these markets can yield highly accurate predictions and outperform large-scale polling organizations, they still display some sort of deviations from the perfect efficient market. These biases may lead to the prediction market's poor performance in forecasting some particular types of events.

One of the most well-known biases in prediction markets is "favorite-longshot bias", which has been documented in many sports betting markets. The favorite-longshot bias describes a phenomenon where odds provide biased estimates of winning chance: teams with a high probability of winning the match (so-called the 'favorite') are undervalued, whereas teams with a meager winning chance (the 'longshot') are overvalued. In betting, odds allow us to calculate how much money we could win when we place a bet; therefore, they imply anticipated probabilities of outcomes associated with those odds: the higher the odds are, the lower the probability of that outcome is. The favorite-longshot bias implies that the favorite team wins more often than the probability suggested by the odds, and the longshot team wins less often. This favorite-longshot bias has been found in various papers that study different sport betting markets including pari-mutuel and fixed-odds betting systems.

In the pari-mutuel betting system, the odds are socially determined because all money placed in a particular event is put together into a pool, and the payoff is calculated by sharing the pool among all winners after deducting all transaction costs and fees. In this setting, the favorite - longshot bias in odds, if it exists, is because bettors tend to over bet on longshots and under bet on favorites. On the other hand, in a fixed-odds betting system, bettors are placing wagers against a bookmaker instead of against other bettors. It is called 'fixed-odds' because the bookmaker will announce the odds, and it is fixed at the time the bet is placed. Hence, the outcome of each bet is predetermined, which puts the bookmaker at risk by informed bettors. Therefore, bookmakers could still possibly overprice longshots and underprice favorites and display favorite - longshot bias not only to maximize their profit in accordance with bettors' preferences but also to protect themselves from insider trading.

The most significant difference between the traditional fixed-odds bookmakers and modern online betting bookmakers is that online bookmakers can easily update the odds while taking into account new exploited information. With the expansion of the internet and availability of a tremendous amount of information, as well as the development of learning machine and other tools to develop prediction models for the outcome of matches with the help of computer power, bookmakers are better and better at predicting the actual probability of each different outcomes of matches. Therefore, the online betting market is expected to operate more efficiently and favorite-longshot bias, if it exists, to be smaller than traditional fixed-odds bookmakers.

Research questions: This thesis aims to study the favorite thoroughly - longshot bias in the European online football betting markets. Using an extensive data set of online betting odds from 11 different football competition in Europe during season 2016/17 and 2017/18, I aim to answer the following questions:

1. How well does the European online football betting market perform as a prediction market? In other words, how accurately do the odds provided by bookmakers in these markets can predict the result of matches or the actual probabilities of winning for teams in those matches?

2. Is there a favorite - longshot bias in these betting markets? Is the bias varied when the team is playing at home or away position? What is the magnitude of this favorite - longshot bias if it exists?

This research could contribute to the literature on this topic by extending studies of favorite-longshot bias in a new different setting, emerging online football betting markets, instead of pari-mutuel and traditional fixed odds betting market.

2 Theoretical background and related studies

2.1 Betting odds, probabilities and bookmaker margins

In betting, odds allow us to calculate how much money we could win when we place a bet. There are three main types of odds: fractional (British) odds, decimal (European) odds, and American (moneyline) odds. These are different ways to present the same thing. First, Fractional odds are the ratio of the amount won to the stake. For example, if the fractional odds are 11/4, for every \$4 one bets, this person could win \$11, plus the return of the stake. It means that the total payout for a \$4 bet at odds of 11/4 is \$15. Second, decimal odds represent the amount return for every \$1 bet, including the money won and the stake. For example, an decimal odds of 3.75 means that for every \$1 bet, one could receive the payout of \$3.75 (\$2.75 of profit and \$1 of returning stake). Last, American odds, depending on the negative or positive sign, either indicate the amount needed to bet to win \$100 or that would win for every \$100 staked.

Since odds indicate the payout rate of winning bets, they imply anticipated probabilities of events. If the odds that team A win the match is quoted as 11/4 or 3.75, then the implied probability that team A will win the match is calculated as

$$\frac{1}{1+\frac{11}{4}} = \frac{1}{3.75} = 26.67\%.$$

In general, fractional odds of A/B indicate an implied probability

$$p = \frac{B}{A+B} = \frac{1}{1+\frac{A}{B}}.$$

The implied probability of decimal odds is the inverse of that odd.

The odds in sports betting can be socially determined by bettors in pari-mutuel settings or are quoted in advance by bookmakers in markets. In pari-mutuel betting, all bets of particular types are placed together into a pool, and the payoff is divided equally among winners after tax and commission deductions. In this setting, the odds reflect bettors' beliefs in the probability of occurrence of each outcome. Meanwhile, in fixed-odds betting, bookmakers post odds in advance of the matches, and they are fixed at the time bet is placed. In this setting, the bookmakers can set the odds in such a way that the outcome will always be in their favor: the sum of the probabilities for all possible outcomes will excess of 100%. This overround will guarantee the profit for bookmakers in case of a balanced book - the odds attract relatively enough bets for each outcome so that the profit margin and the payout for winners can be fully covered by the stake of losers regardless of the actual outcome. In this type of setting, the bookmaker could also possibly adjust the mark-up on odds for favorites differently than for longshots to balance the book or exploit any biases in bettor's beliefs to achieve more profit.

In fixed-odds betting, the bookmaker margin, which is also called "overround", is a hidden transaction cost for bettors when placing a bet. It is calculated as the sum of implied probabilities of all possible outcomes of an event subtract 100%. For a football match with three implied probability that home team wins, away team wins and draw ($\hat{p}_H, \hat{p}_A$, and $\hat{p}_D$), bookmaker margin is calculated as

$$M = \hat{p}_H + \hat{p}_A + \hat{p}_D - 1$$

Because of bookmaker margin, implied probability does not provide the true view of bookmakers about the probability of a possible outcome. When taking this margin into account and assuming that the bookmaker's margin is allocated evenly among all possible outcome of one betting event, the implied probability with no bookmaker margin can be calculated as

$$p_i = \frac{\hat{p}_i}{1+M} = \frac{\hat{p}_i}{\sum \hat{p}_i}; \qquad i \in \{H, A, D\}$$

It also should be noted that probabilities implied by bookmakers' odds are seen as subjective probability, the possibility that an outcome will happen based on bookmakers and bettors' judgment. The objective probability is things that can not be known for sure, and we can only obtain an estimation of it based on actual outcomes.

2.2 Odds setting by bookmakers

According to Levitt (2004), there are at least three ways for bookmakers to set the odds for profits. Bookmakers can either balance the books, set the odds at the market-clearing price level, or actively set the odd to exploit the bias in bettor preference.

Balancing the book means that bookmakers set odds to equilibrate money placed on each team, so they make money no matter how the game turns out. When large betting amounts are placed on a team, bookmakers will lower the odds for this team and increase odd for the opponent team to attract more money placed on the opposite side until betting balances. Using this pricing strategy, bookmakers do not need to have any particular skill in predicting the true probabilities of different outcomes for a sports event but need to be good at forecasting betting' behavior.

The second strategy is to set odds in accordance with bookmakers' predictions of the actual outcome. Using this strategy, bookmakers may not achieve balancing the book in every individual match because their prediction may not be in line with bettors' preferences. However, this strategy is still profitable on average if they are consistently better than bettors in forecasting outcomes of sports events.

The third strategy is to set odds using both bookmakers' predictions of actual outcome and their understanding of bettors' preferences. Bookmakers maximize their profits by altering the odds so that they are deviated from the actual probabilities to exploiting bettors' preferences. This strategy, therefore, may lead to favorite-longshot bias. When bettors tend to prefer betting on longshots to betting on favorites, bookmakers will adjust the odds accordingly by lowering odds for longshots and increasing odds for favorites. Moreover, since any inaccurate prediction of probability for longshots' win will cost bookmakers more than that for favorites' win, bookmakers will also tend to shorten odds for longshots more relatively to favorites to protect themselves in case some bettors have more information than they do. To use this strategy, bookmakers need to have the ability to predict both the outcome of matches and betting behaviors. Makropoulou and Markellos (2011) states that this pricing strategy is optimal for bookmakers under uncertainty.

According to Makropoulou and Markellos (2011), no matter what pricing strategy is chosen, bookmakers will all do to maximize their expected profit, which will always lead to inefficiencies in betting markets.

2.3 Market efficiency and behavioral biases

Betting markets are expected to be efficient because of their characteristics that facilitate learning (Thaler and Ziemba (1988)). According to the efficient market hypothesis, the prices should reflect all available information, and there is no room for arbitrage. The degree of efficiency can be defined as weak, semi-strong, or strong, depending on the information that is reflected in the prices. If all past and current public information is reflected in odds, the betting market is weak-form efficient. It is semi-strong-form efficient when odds reflect all publicly available information and strong-form efficient when odds reflect all publicly available and private information. A betting market is considered efficient if bookmakers' odds are unbiased predictors of the actual outcome, and the expected return of all bets is zero. Most previous studies found that betting markets are surprisingly efficient, at least in weak-form within transaction costs(Kuypers (2000)). It means that it is difficult to earn any abnormal return on these markets. However, biases still exist in these markets. Some of the most well-known biases in sports betting are favorite-longshot bias, home field bias, and sentimental bias. Levitt (2004) and Kuypers (2000) found that a profit-maximizing bookmaker will set inefficient odds by exploiting those biases.

In sports such as football, teams play on their home field win more often than their relative strength suggests; however, this advantage is underestimated by bookmakers (Elaad (2020)). This creates the home field bias, which is the phenomenon when odds quoted by bookmakers underestimate the chance that home teams win and overestimate the chance that away teams win. Elaad (2020) found that home-field bias exists in online betting markets in two English competitions League One and League Two, during season 2010/11 and 2017/18. Golec and Tamarkin (1991) also found home-field bias and favorite-longshot bias in American football league NFL and US college football leagues for 15 years from 1973 to 1987. Since all these biases could be presented in the market, home-field bias needs to be taken into account when estimates the favorite-longshot bias.

Another notable bias is sentimental bias, where bettors prefer to bet on specific teams because of some irrational reasons. For example, a team's fan is more likely to bet on that team even when the chance of winning is low. It could lead to a sentimental bias: odds are affected by the relative number of fans of each team in a football match. Kuypers (2000) predicts that this bias leads to shorter odds for heavily supported teams.

2.4 Evidence of favorite - longshot bias

The first time favorites - longshot bias has been noted is by Griffith in 1949. Griffith (1949) used the data containing odds from 1386 horse racing in the US in the pari-mutuel system to study the US bettors' behavior. He grouped horses into 11 categories according to their odds and use win pool shares as the estimates of subjective probability. By comparing the subjective probability and the objective probability (the frequencies of horse winning) in each group, he found that subjective probability is underestimated the winning chances of short-odded horses and overestimated those of long-odded horses. Griffith (1949) then concludes that a bias, which later called "favorite-longshot bias", may exist in horse racing data in the US in the pari-mutuel system. After Griffith (1949), numerous researchers have documented this longshot bias. It has not only been found in race-track betting data around the world but also covers both bookmaker and pari-mutuel market systems. Sauer (1998), Snowberg and Wolfers (2008) and Ottaviani and Sørensen (2008) surveyed the literature of favorite-longshot bias in race-track betting. There are only a few exceptions, which documented in Busche and Hall (1988) for the Hong Kong race-track, Busche (1994) for Hong Kong and Japanese race-tracks, and Swidler and Shaw (1995) for small American race tracks. This reversed bias is still an unsolved puzzle with little theoretical analysis.

The empirical research also extended to other sport betting markets. In Cain et al. (2000), the authors used the data that comprise the results of 2855 football league matches played in the UK during 1991/92 season with associated odds against each simple outcome (home win, away win, or draw) and against each score quoted by bookmaker William Hill. They found that the individual fixed-odds betting market on UK football exhibits the same favorite-longshot bias found in horse-racing. Cain et al. (2003) also tried to verify the existence of the favorite-longshot bias in a variety of fixed-odds sports betting markets. They found the bias existed in betting on boxing, cricket, football, greyhound races, horse races, tennis, and even on baseball, which contrasts with the findings in Woodland and Woodland (1994) and Woodland and Woodland (2003) that a reversed bias exists in a study of fixed-odds betting on US baseball market. Gandar et al. (2002) reexamined the US baseball betting market and concluded that the bias is not solely a bias involving favorites and longshots and the relationship between subjective probabilities and returns seem to be suggestive of a favorite - longshot bias rather than its reverse. Home-field advantage, which arises from the fact that a team plays better on its home ground, has shown to be consistently underestimated by bookmakers. In conjunction with the reverse favorite-longshot bias, this home-field bias gives rise to what in the literature so-called home-underdog bias.

In spite of voluminous lists of studies related to favorites-longshot bias in horse racing, there are only a few studies of favorite-longshot bias in fixed-odds football betting markets, and most of them are only related to the traditional fixed-odds bookmakers where the odds are declared and unchanged until the matches. Cain et al. (2000) and Rossi et al. (2011) both claimed that the fixed-odds betting markets exhibit the favorite - longshot bias in their data on the UK's and Italian football league. Rossi et al. (2011) suggested that the bias is both caused by bettor's behavior and bookmaker's strategy to deal with insiders. Makropoulou and Markellos (2011) compared the odd between an online betting bookmaker and a traditional fixed-odds bookmaker and found out that both exhibits favorite-longshot bias, and the bias is more pronounced for the latter.

2.5 Measurement of favorite-longshot bias

There have been several ways to measure the favorite-longshot bias in the literature. Previous studies have used different methods to verify the favorite-longshot bias by either comparing subjective probability implied by odds and estimated objective probability for a favorite/longshot team to win from the actual outcome of a match or comparing the returns of bets on favorite and longshot.

For the first method, the grouping technique to calculate subjective and objective probabilities also varies from study to study. Therefore, the magnitude of the bias among those studies is not compatible.

Cain et al. (2000) and Makropoulou and Markellos (2011) reported the difference in implicit and objective probabilities in the form of 'returns to a unit bet' - calculating the average return of betting $1 on each odds in various bins of implicit probabilities from those odds. The returns to a unit bet are equivalent to the percentage difference between the objective probability and the implied probability of odds in each category. According to Cain et al. (2000), in their data on odds quoted by the bookmaker William

Hill of 8655 football matches played in the UK during 1991/92 season, it is shown that bets on longshots with an implied probability of winning lower than 0.2 had substantially low returns of -0.155 compare to that of -0.017 for bets on favorites with a probability of winning over 0.6. In other words, betting on long shots yielded a loss of -15.5% while betting on favorites yielded a loss of only -1.7%. However, there is no evidence that betting on favorites is a potentially profitable strategy because of the transaction cost and the profit margin (overround) of bookmakers. Makropoulou and Markellos (2011), using the same technique with the result of 8363 matches played during 2002-2004 and associated odds quoted by two bookmakers, a leading European-based online bookmaker and a major European-based bookmaker who issues fixed-odds coupons, found that both bookmakers exhibit the favorite-longshot bias in their odds. The return for a bet on longshot (implicit probability under 0.2) with online bookmaker and bookmaker who issues coupons are respectively -35.58% and -40.57%, whereas those for favorite (implicit probability over 0.8) are -0.5% and -3.95%. They found the favorite - longshot bias is much more pronounced for the bookmaker with fixed-odds coupons than the online bookmaker.

Rossi et al. (2011) investigated the favorite-longshot bias in the Italian football betting markets using the data consists of the result and odds quoted by three different bookmakers on 289 matches played in the 2007/08 Serie A season. By computing the bookmaker's take-out rate - the percentage point difference between the implied probability from odds and the actual frequency of winning for each sub-group with the cut-off at distribution standard deviation of frequency, he found out that on average, the take-out rate was 6% for longshots and 1% for favorites. It means bookmakers shorten the odds of longshots more relatively more than favorites, confirming the existence of favorite-longshot bias in the Italian football wagering markets.

2.6 Theoretical explanations for favorite-longshot bias

There are still unclear explanations for this phenomenon. There are many theoretical explanations suggested for this bias, such as locally risk-loving attitude, misperceptions of probability, or the existence of insider trading.

First, the neoclassical theory suggests that the prices bettors are willing to pay can be used to recover their utility function. The locally risk-loving agent with unbiased belief and maximizing expected utility function can accept lower average returns on longshots (riskier bets). Thus, the favorite-longshot bias can be rationalized by positing a risk-loving utility function (Snowberg and Wolfers (2010)).

Another explanation, suggested by behavioral theories, is that cognitive errors and misperceptions of probabilities play a role in market mispricing. People are systematically bad at discerning small and tiny probabilities and exhibit a strong preference for certainty over extremely likely outcomes. As in Kahneman (1979), people might have the probability weighting function that overweights low probabilities and underweights high probabilities, leading to overestimate the winning possibilities of longshots and underestimate those of favorites.

In Snowberg and Wolfers (2010), the authors used an extensive data set to implement an empirical test that can discriminate between these competing theories in the context of the pari-mutuel betting system on horse racing. The test shows evidence in favor of the view that misperceptions of probability drive the favorite-longshot bias.

The third explanation of insider trading is first suggested by Shin et al. (1991), Shin (1992), and Shin et al. (1993). Shin models assume that bookmakers take into account the presence of 'insiders' - who have superior information, which is better than what bookmakers have - in their odds-setting procedures. The bookmakers have to determine the odds inefficiently to raise enough revenue from 'outsiders' to pay insiders their winnings. One consequence of this behavior is that the betting odds tend to understate the winning probability of the favorites and to overstate those of the longshots.

Makropoulou and Markellos (2011) also developed a model in fixed-odds setting. Instead of having insiders who possess private information as in Shin et al. (1991), Shin (1992), and Shin et al. (1993), bookmakers have to deal with experts or informed bettors who exploit public information that arrives after the declaration of odds. Fixed-odds bookmakers, therefore, require a consistently higher margin than bookmakers that may practically update odds (for example, online betting bookmakers) as compensation for the additional risk that informed bettors will account for information arriving after the odds have been set and exploit any mispricing on behave of the bookmaker. The theoretical pricing model in Makropoulou and Markellos (2011) identifies two sources for the favorite-longshot bias as a result of an optimal pricing policy by the bookmaker. The first is the existence of behavioral biases among bettors. If bettors tend to overestimate the probability of winning for longshots relative to that for favorites, the bookmaker distorts the prices accordingly by quoting relatively shorter odds on longshots to optimize profits. The second is the variance of new information that may be exploited by informed bettors.

According to Ottaviani and Sørensen (2008), the different theories which explain the favorite-longshot bias is difficult to distinguish from each other empirically. In order to compare the performance of those alternative explanations for favorite - longshot bias, most previous literature has tried to test whether those theories are able to simultaneously explain both the favorite - longshot bias and other regularities in odd adjustment and the timing of bets.

3 Data

The data used in this study consists of results for simple outcomes (a home win, an away win, or a draw) of 7,669 matches played in 11 well-known football leagues during season 2016/17 and 2017/18, along with associated odds for each outcome of those matches quoted by 28 online bookmakers.

Matches' results are derived from the website `www.worldfootball.net`. This website provides data on past football match results, includes all competitions that are considered in this study. Each observation consists of the date and time of the match, names of home and away teams, and the final score of the match. The final score is used to define which team won that specific game. Those results of football games will be matched with odds on Oddschecker based on the match's date and time and the name of the teams.

The associated odds for those matches is given by Marko Terviö. The data is scraped on `www.oddschecker.com`. Oddschecker provides live tables of all odds bookmakers and betting exchanges offer currently on an event. Odds on each match are recorded in a 2-hour interval since the day they start to appear on Oddschecker. Usually, the odds are quoted for the first time at around two weeks before the match is played. Each data point includes the date and time the data is recorded, the date and time of the matches, names of teams and odds offered for a result of the victory of either the home or away team, or draw by different bookmakers. Oddschecker provides odds quoted by a total of 28 online bookmakers and betting exchanges and all odds recorded in the form of fractional odds. It is also noted that some bookmakers announce their odds as soon as two or three weeks prior to the match, but others may publish their odds only a few days before kick-off. Therefore, the number of bookmakers giving odds for a match varies during the pre-match period, and there is a maximum of 28 bookmakers quoting odds simultaneously. All odds are converted to decimal-form for easier calculation and interpretation, and all odd mentioned from now on is decimal odd. The formula to

Table 1: List of Competitions and the Number of Matches and Odds Quoted in the Data

Total number of matches and numbers of odds quoted by all bookmakers in the data set. Column "Day -1" and "Day -7" represent the numbers of odds quoted one day and seven days before kick-off respectively.

League Name		Match played	Odds quoted		
			All	Day -1	Day -7
UEFA	Champions League	217	165,807	90,813	39,645
English	Premier League	760	692,472	7,125	5,169
	Championship	1,114	579,231	20,640	7,599
	League One	1,114	504,339	90,978	26,214
	League Two	1,114	502,800	90,585	23,781
	National League	1,114	369,240	82,791	7,914
	FA Cup	342	128,598	62,058	54,657
	EFL Cup	93	69,414	8,214	2,295
Spanish	La Liga Primera	760	628,965	61,245	49,356
	La Liga Segunda	932	374,649	73,047	9,702
	Copa Del Rey	109	40,521	17,439	11,529
Total(N)		7,669	4,056,036	604,935	237,861

convert a fractional odds A/B to decimal odd is:

$$O_{decimal} = 1 + \frac{A}{B}$$

All data and R codes for data cleaning are available in the supplementary file of this thesis.

Table 1 shows the number of matches and odds quoted by bookmakers in football competitions that I use in my data. If a bookmaker only provides one or two odds for a particular match, those odds are dropped from the data set due to missing value. There are in total 4,056,036 odds, quoted by 28 different bookmakers and betting exchanges for 7,669 matches in the data. The number of odds quoted increases as the kickoff date is closer. The total number of odds offered one day before kickoff is approximately 2.54

Table 2: Number of bookmakers offering odds for matches on each day prior to kickoff

	No. matches	No. bookmakers Mean(SD)		No. matches	No. bookmakers Mean(SD)
Day -1	7636	26.4 (2.45)	Day -8	2789	17.2 (9.60)
Day -2	7638	26.1 (3.03)	Day -9	2278	18.5 (7.91)
Day -3	7573	25.3 (4.08)	Day -10	2039	18.8 (7.36)
Day -4	7177	22.8 (7.67)	Day -11	1927	18.2 (7.67)
Day -5	6672	21.9 (7.47)	Day -12	1737	18.5 (7.31)
Day -6	6365	19.6 (7.99)	Day -13	1561	18.0 (7.14)
Day -7	5520	14.4 (9.20)	Day -14	1388	15.5 (8.16)

times more than that offered seven days before. Table 2 presents the average number of bookmakers providing odds each day in fourteen days before kickoff and their standard deviations. In most games, the odds are quoted about one week earlier until the games play. Two weeks before kickoff, only about 20% of matches in the data set received odds by bookmakers. That number increases day by day until the matches start. The average number of bookmakers give offers on each day also increases day after day and reaches a maximum of 26.4 on the last day before the match dates. Almost all bookmakers provide odds on the few last days before the matches. Fewer bookmakers offer odds one or two weeks earlier than the dates of matches, especially for matches in the preliminary or first round. Figure 1 compares the number of bookmakers quoting odds for a match in two periods: one day and seven days before matches play. Most bookmakers announced odds within one day before matches; meanwhile, the number of bookmakers giving odds for a match one week earlier varies from less than 5 to 28.

In online betting markets, bookmakers tend to frequently update the odds in the pre-match period and adjust the odds when new information comes to light. For example, bookmakers will reduce the odds for longshots and increase the odds for favorite if a sudden large amount of money is placed on the longshot team. It is important to note

Figure 1: Number of bookmakers giving odds for a match before that match play

that bettors who placed a bet prior to a change in odds will still play for the odds they put the money on. The constant odds updating of bookmakers, therefore, could potentially lead the odds to represent betting behavior better: the probability implied by these odds will continuously capture new information available in the market and move towards the true objective probability, which increases market efficiency.

One limitation of my data is that it could possibly contain some 'stale odds' when something went wrong on Oddschecker, such as an odd that was no longer available with a bookmaker but still be on the website or a few odds extracted are wrong due to a technical error in the retrieval process. To deal with outliers' problems, I will use the median of odds quoted simultaneously by all bookmakers for my analysis.

Table 3 provides some descriptive statistics of the data. It shows the average of the mean, median, max, and standard deviation of odds quotes by different bookmakers simultaneously on three different outcomes: home win, away win, or draw. On average, the standard deviation of odds among different bookmakers is 0.208; the average odds in the whole data is 3.43. On average, odds for betting on a home win are lower than betting on a draw or an away win. As shown in this table, the average and the median of bookmakers' odds are close to each other, suggesting that the median of odds is a

Table 3: Descriptive Statistics

The average, median and maximum of odds and standard deviation of odds quotes by different bookmakers simultaneously for one match on average

		Average Odds	Median Odds	Max Odds	SD
Day -1	Home Win	2.54	2.54	2.76	0.092
	Draw	3.70	3.70	3.98	0.135
	Away Win	4.09	4.09	4.59	0.231
	All	3.44	3.44	3.77	0.153
Day -7	Home Win	2.38	2.40	2.56	0.148
	Draw	3.50	3.51	3.73	0.195
	Away Win	3.83	3.84	4.35	0.383
	All	3.24	3.25	3.55	0.242
All	Home Win	2.50	2.51	2.69	0.119
	Draw	3.67	3.67	3.94	0.172
	Away Win	4.12	4.12	4.68	0.332
	All	3.43	3.44	3.77	0.208

good representative of odds by different bookmakers. When comparing odds seven days and one day before matches' dates, odds for betting on the away team's victory has been shortened while odds for betting on that of the home team have been lengthened. The standard deviations of odds quoted by bookmakers concurrently are also smaller one day before matches' dates than that seven days before, which suggests that odds are more concentrated near the mean as matches' dates approach.

4 Empirical Methodology

Investigating the favorite-longshot bias has been done in several ways in earlier literature by looking at the subjective and objective probability. The former is derived from what the odds implied while the latter from the actual results.

The subjective probability is the probability that an individual ascribes to a specific outcome. So this subjective probability is based on the odds given prior to the match. It is also noted that the sum of the implied probabilities for all three possible outcomes (a home win, an away win, or a draw) will excess of 100%. Hence, the subjective probabilities are normalized so that the sum of chances for a home win, an away win and a draw of one match is equal to 1.

Since all 28 bookmakers' odds are highly correlated and usually move closely to each other, I first construct synthetic odds, which represent the odds offered by all different bookmakers. To tackle the problem that 'stale' odds may possibly exist in the data set, I choose the synthetic odds are the median of the odds offered by all bookmakers at the same moment.

The objective probability, on the other hand, is the probability derived after the match. If we have only one match with the result is one of the three outcomes: a home win, an away win, or a draw, then it is impossible to know the probability of each outcome. However, if we have many matches, we can calculate the objective probability for each outcome of the whole group of matches by dividing the number of times the outcome occurs by the total number of matches played. The method to calculate the objective probability here is similar to the method used in Griffith (1949) and Ali (1977), which also is followed in many previous studies investigating favorite longshot bias. The categories of the subjective winning probabilities can now be derived with a constant interval. The used intervals between categories in this study are steps of 5% since this could better reflect the objective probability distribution of different probabilities.

For example, teams in the category of subjective probabilities 80-85% might win more often than the subjective probabilities implied, while teams in the category of subjective probabilities 85-90% might win less often. When these winning probabilities are viewed together in the interval 80-90%, the calculated objective winning probabilities might give a biased view of the actual winning probabilities for some teams.

In each of those categories, I estimate the objective probability and the confidence interval of this estimation based on each bet's result. The "Result" is 1 if the bet is associated with the outcome that actually happened, and 0 otherwise. The random variable "Result" is expected to follow a binomial distribution $B(n, p)$ in which n is the number of bets and p is objective probability. The estimated objective probability is calculated as the average of variable "Result":

$$\text{Objective Probability} = \hat{\pi} = \frac{\sum \text{Result}_i}{n}$$

The 95% confidence interval is estimated using the Clopper-Pearson interval, which is based on the cumulative probabilities of the binomial distribution. The method is an early and common method for calculating binomial confidence intervals, which Clopper and Pearson introduced in their paper Clopper and Pearson (1934). Compared to other methods that estimate the confidence intervals based on an approximation, the advantage of this method is that it can obtain a more accurate estimation with a small sample size n or the success probability p is close to 0 or 1. The interval obtained from this method will never cover less than the nominal coverage while other methods may cover less than the nominal coverage (Brown et al. (2001)). It means that a 95% confidence interval obtained from this method may be wider than it is needed to achieve 95%.

To test whether there are any differences in the objective and subjective probabilities in each category, the null hypothesis $\pi = p$ is tested. This hypothesis considers whether, on average, in each category, the number of times the outcome that is bet on actually happened is in line with the average (or mid-point) subjective probability p of that category. The random variable Result is expected to follow the binomial distribution

$B(n, p)$. Therefore, a binomial test is suitable for this purpose.

Because binomial tests are performed simultaneously in every category and considered together to test for the existence of favorite-longshot bias, the multiple testing problem arises. If only one test is performed at the 5% significance level, then there is only a 5% chance to reject the null hypothesis incorrectly. In other words, the false-positive error rate of one test is 5%. However, if 20 tests are performed simultaneously at the 5% significance level each, the expected number of false-positive is 1. In order to control for this so-called multiple comparison problem, each test's error rate must be more stringent than $\alpha = 5\%$ so that the probability of making at least one false-positive error among all tests remains at 5%. The most conservative method to do that is the Bonferroni correction: to obtain a familywise error rate of α when testing the hypothesis on m categories, the false positive error rate for each test must be no higher than $\frac{\alpha}{m}$.

To test for the existence of favorite-longshot bias as a whole, I use the following standard linear regression:

$$Result - p = \beta_0 + \beta_1 p + \varepsilon$$

The variable "Result" is equal to 1 if the event that is wagered on happened and 0 if it does not. The variable p is the normalized subjective probability implied by the median of odds. The regression is applied separately for different types of odds: home win, away win, or draw odds. Since three possible bets on each match are not independent, odds for a draw from a set of bets on the same match are dropped from regression.

In the absence of any bias, the coefficients would be $\beta_0 = 0$ and $\beta_1 = 0$. Meanwhile, the favorite-longshot bias would be indicated by $\beta_0 < 0$ and $\beta_1 > 0$. Two t-tests are conducted to test the hypotheses that $\beta_0 = 0$ and $\beta_1 = 0$.

The t-statistic has the following formula for testing the hypothesis that $\beta = 0$:

$$t = \frac{\beta - 0}{s/\sqrt{n}} = \frac{\beta}{SE}$$

It's also note that the difference in objective and subjective probability could possibly be affected not only by the favorite-longshot bias but also by home-field bias, a second linear regression should be used, controlling for the effect of home-field bias if existence:

$$Result - p = \beta_0 + \beta_1 p + \beta_2 Home + \beta_3 p \times Home + \varepsilon$$

In this regression, β_2 estimates the home-field bias and β_3 estimates the effect of home-field bias on the favorite - longshot bias.

5 Results

In this chapter, I will use the discussed methodology to investigate whether the favorite longshot bias exists in the sample of data consisting of quoted odds within one day before matches start.

My first step of analysis is to construct synthetic odds, which are the medians of odds offered simultaneously by all bookmakers, and calculate the implied subjective probabilities associated with each of those odds. This implied subjective probability is then normalized so that the sum of subjective probabilities on all three possible outcomes is equal to 1. This normalization relies on the assumption that the bookmaker's margin is evenly distributed among all these three bets.

Figure 2 shows the histogram of normalized subjective probability implied by the synthetic odds. The frequency distributions of the constructed odds are quite normal for the home/away team's victory, but the frequency distribution for the draw odds is strongly leptokurtic and asymmetric.

5.1 Favorite - longshot bias in each probability category

The subjective probabilities are divided into different categories by step of 5%. In each category, the objective probability and the 95% confidence interval are calculated. A binomial test is conducted to test whether there are any differences in the objective and subjective probabilities. For each category, the result of the bet is either win or loss; therefore, it is expected to follow the binomial distribution B(n,p), where n is the number of observations in that category and p is the average of subjective probability in that same category. A two-sided binomial test is performed to test whether the objective probability is equal to subjective probability in each category.

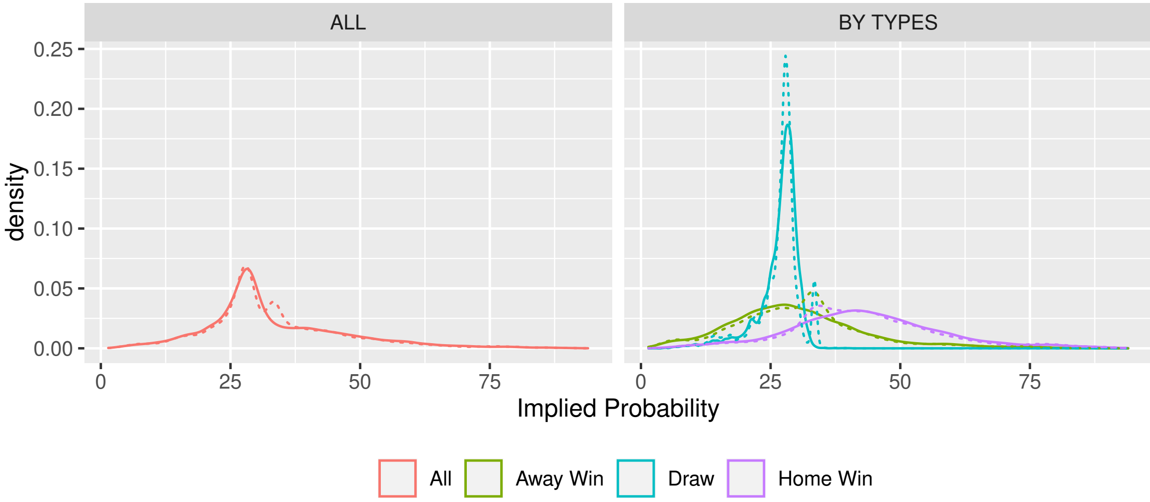

Figure 2: Distribution of (Normalized) Subjective Probability

This figure illustrates the density of subjective probability implied by the synthetic odds. The synthetic odds are constructed as the median of odds from all bookmakers at the same time. The solid line represents the density of implied probability within one day before matches play. The dashed line represents the density seven days before matches play.

To further investigate the favorite-longshot bias, also taking into account home-field bias, the data is divided into three groups for different types of odds: home win odds, away win odds, and draw odds. Similar binomial tests are also performed on each category for each type of odds.

Figure 3 presents the average subjective probability and estimated objective probability in each category within one day before matches start. The results used to construct this figure is shown in Appendix A1. Overall, subjective probability is a good predictor of objective probability. In Figure 3, points depict subjective and objective probabilities relationship is very close to the 45-degree line. As shown in Figure 3, overall, the subjective probabilities for longshots, which is points with low subjective probability, is higher than its paired objective probability; meanwhile, the objective probability for favorites, which is points with high subjective probability, is lower than its paired objective probability.

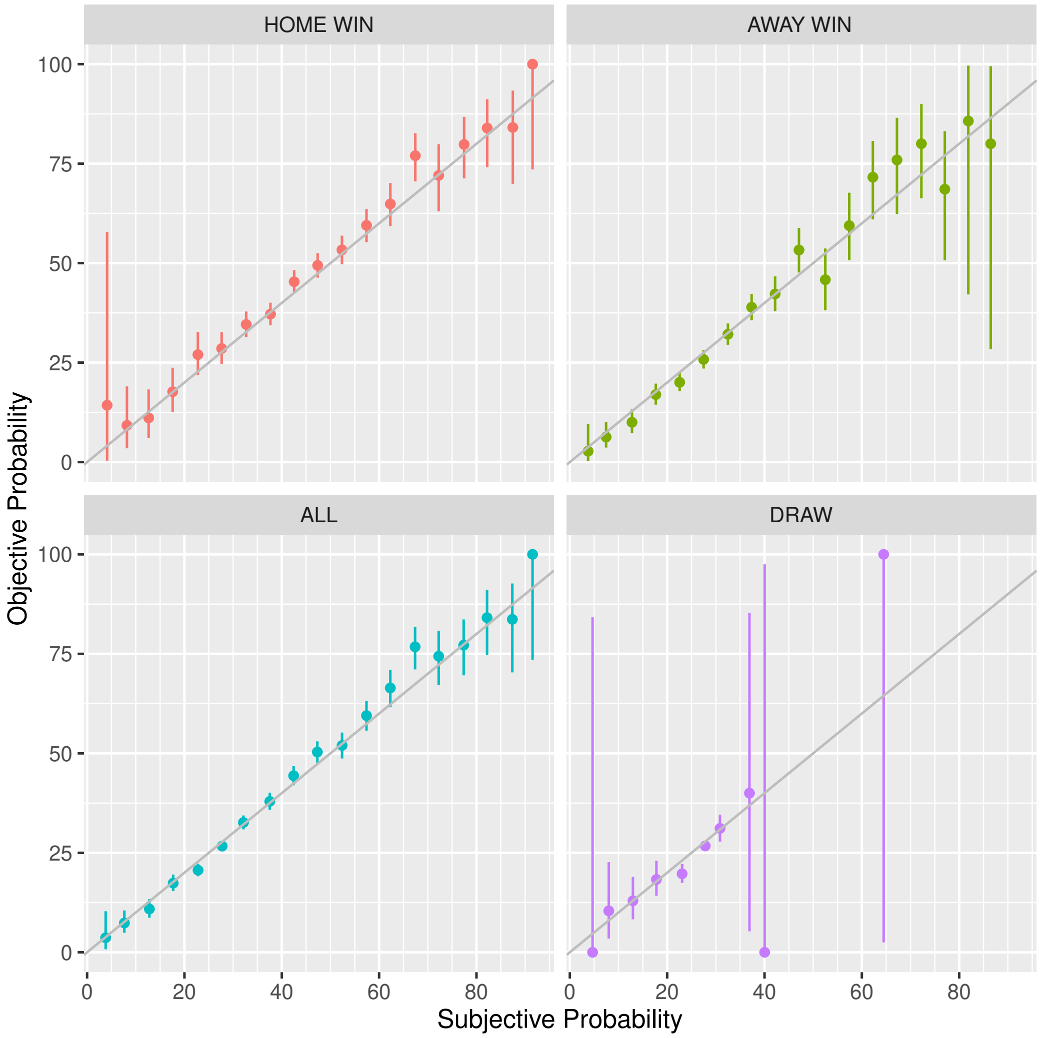

Figure 3: Favorite - longshot bias one day before kickoff: a comparison of subjective and objective probabilities

This figure presents the objective and subjective probability points with the 45-degree line. Points above (below) this line depict categories where subjective probability is lower (higher) than estimated objective probability. Horizontal line segments describe the 95% confidence intervals of estimated objective probabilities. The categories here are constructed by dividing subjective probabilities into 19 groups with a step of 5%: 0-5%, 5-10%, 10-15%, ..., 90-95% (There is no subjective probability higher than 95%). The estimated objective probability and confidence interval are shown in Appendix A1.

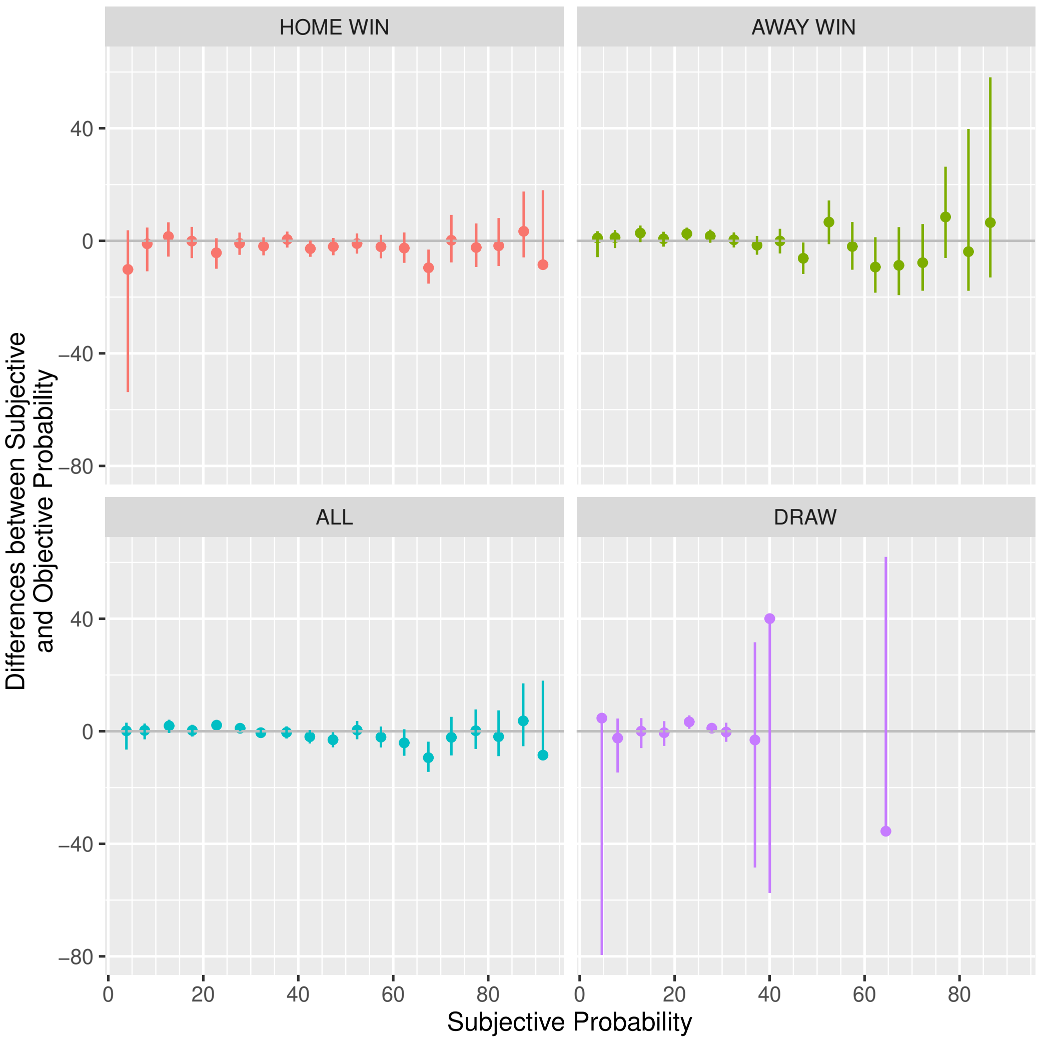

Figure 4: Favorite - longshot bias one day before kickoff: the differences between subjective and objective probabilities

This graph presents the differences between subjective and objective probability in each category. Horizontal line segments describe the 95% confidence intervals of these estimations. The categories here are constructed by dividing subjective probabilities into 19 groups with a step of 5%: 0-5%, 5-10%, 10-15%, ..., 90-95% (There is no subjective probability which is higher than 95%). The estimated objective probability and confidence interval are shown in Appendix A1.

An alternative way to compare subjective and objective probabilities is to compare the differences between subjective and objective probabilities with zero. Figure 4 illustrates the differences between subjective and objective probabilities in each category. As shown in Figure 4, overall, the differences are slightly positive for points with low subjective probability (longshots) and slightly negative for points with high subjective probability (favorites), which suggests that the longshots' winning probability is overvalued and the favorites' winning chance is undervalued.

However, the 95% confidence interval line segments show that these differences are not statistically significant. It is also notable in Figure 3 and Figure 4 that the subjective probability of betting on home wins is all lower than the objective probability in the same category; this is described as negative differences between subjective and objective probability for bets on victories of home teams in Figure 4. It could imply a strong home-field bias.

Table 4: Binomial test: Is objective probability equal to subjective probability one day before kickoff?

This table presents the unadjusted p-values of exact binomial tests whether the objective probability equal to the subjective probability in that category.

Category	All	Home	Draw	Away	Category	All	Home	Draw	Away
(0,5]	1	0.2328	1	1	(50,55]	1	0.4949		0.146
(5,10]	0.9218	0.6529	0.4322	0.6303	(55,60]	0.2479	0.2462		0.8619
(10,15]	0.2408	1	0.8205	0.1289	(60,65]	0.0349	0.1965	1	0.0674
(15,20]	0.5882	1	0.8834	0.6111	(65,70]	0.0092	0.0205		0.3005
(20,25]	0.0087	0.0964	0.0066	0.0395	(70,75]	0.6645	0.8352		0.2694
(25,30]	0.0318	0.5292	0.0585	0.1189	(75,80]	0.9222	0.6536		0.2298
(30,35]	0.3618	0.327	0.4949	1	(80,85]	0.5753	0.6593		1
(35,40]	0.7821	0.8791	0.6266	0.5712	(85,90]	0.3935	0.5011		0.4427
(40,45]	0.0704	0.041	1	0.8208	(90,95]	0.6161	0.6161		
(45,50]	0.0712	0.4417		0.0225					

Table 4 shows the result of binomial tests for each category in all sample data and three sub-sample (bets on a home win/ a draw/ an away win). Bonferroni correction is used to control for the multiple testing problem. To obtain a familywise error rate of 5% when testing the hypothesis on 19 different categories, the error rate per comparison must be $\overline{\alpha} = 5\%/19 \approx 0.263\% = 0.00263$. There is no p-value in Table 4 that is small than $\overline{\alpha} = 0.00263$; therefore, the objective probability and subjective probability is not significantly different from each other.

5.2 Favorite - longshot bias across probability categories

Figure 3 plots the objective probability against subjective probability in each category for all odds and for each type of odds. Except for outliers in the first or last category for draw odds due to the small size of the sample, linear regression is fitted well to observed data.

Favorite-longshot bias across all categories is tested by performing the following regression:

$$Result - p = \beta_0 + \beta_1 p + \varepsilon$$

This regression estimates the effect of subjective probability on the difference in objective and subjective probabilities. In the absence of any bias, β_0 and β_1 is equal to 0. The regression is conducted in 4 sample groups: mixed types, home win odds, away win odds, and draw odds. The first group is constructed by dropping draw odds.

The coefficient of regression and p-value for the t-test are reported in Table 5. For the market as a whole, the favorite-longshot bias is presented, indicated by a negative constant $\beta_0 = -2.419$ and a positive coefficient $\beta_1 = 0.081$. The estimates are significant at a 5% significance level. As shown in the result of β_0 and β_1, there is an away-longshot bias in this market: the objective probability is much higher than the subjective probability for betting on the home teams, especially for the home-favorite teams, while it is lower than its paired subjective probability for betting on away teams

Table 5: Regression summary: The favorite-longshot bias in the whole market one day before kickoff

Coefficient		Estimate	Std. Error	p-value
A. Mixed-type Odds				
Constant	β_0	-2.419	0.953	0.0111
Subj. Probability	β_1	0.081	0.024	0.0006
N =	15,272			
B. Home Win Odds				
Constant	β_0	0.727	1.745	0.677
Subj. Probability	β_1	0.020	0.038	0.599
N =	7,636			
C. Away Win Odds				
Constant	β_0	-3.571	1.242	0.00404
Subj.Probability	β_1	0.101	0.038	0.00761
N =	7,636			
D. Draw Odds				
Constant	β_0	-3.520	3.329	0.290
Subj.Probability	β_1	0.088	0.124	0.477
N =	7,636			

or draw results. The favorite-longshot bias is most pronounced in sample of away win odds, with $\beta_0 = -3.571$ percentage point and $\beta_1 = 0.101$ percentage point. These estimates are significantly different from 0 at a 5% significance level with correction for multiple testing. In other words, at a 5% significance level, the probability of a victory of an away-favorite team is underestimated by bookmakers' odds while that for an away - longshot team is overestimated.

Because there is no evidence of favorite-longshot bias existence in odds for betting on a draw, and there is a home-away bias in the data set. I use the following regression to separate the effect of favorite-longshot bias and home-field bias in a sample of bets on a home win or an away win:

$$Result - p = \beta_0 + \beta_1 p + \beta_2 Home + \beta_3 p \times Home + \varepsilon$$

In which, $Home$ is a dummy variable that takes value 1 if the bet is on a home win and 0 if the bet is on an away win.

Table 6: Regression summary: The favorite-longshot bias and home bias estimates one day before kickoff

Coefficient		Estimate	Std. Error	p-value
Constant	β_0	-3.571	1.303	0.00616
p	β_1	0.101	0.040	0.01099
Home	β_2	3.976	2.094	0.05763
$p \times Home$	β_3	-0.071	0.054	0.18252

Table 6 indicates the result of the regression. The favorite-longshot bias persists in betting on both home and away teams at a 5% significance level, which is indicated by a negative constant $\beta_0 = -3.571$ and a positive main coefficient $\beta_1 = 0.101$. Coefficient β_3 is slightly negative but insignificantly different from 0. It means the biases are not much differences between betting on home and away teams. However, the probability for betting on home win is underestimated by approximately 3.6 percentage points compared to that for betting on away win. Although the home-field bias is statistically insignificant due to high standard error, the unobserved favorite-longshot bias in home win odds could possibly be dampened by the home-field bias.

6 Alternative Specifications

The result in chapter 5 shows that there is some evidence that a favorite-longshot bias exists for the whole sample data, although it is not significant. In this chapter, I reexamine the results with alternative specifications. This investigation is performed to verify the existence of favorite-longshot bias, whether the above result is sensitive to the chosen specifications.

6.1 Odds grouping: equal ranges or equal observations

In the previous section, odds are divided into 20 different groups based on implied subjective probability. The categories are constructed so that the range of implied probability is divided into 20 equal intervals of 5% and the odds associated with similar implied probability are in the same group. Odds grouping by equal interval gives 20 categories of subjective probability with equal ranges but unequal numbers of observations. Using equal interval provides information about favorite-longshot bias in the group of extreme favorite or longshot odds at the cost of higher standard errors because of small numbers of observations.

Another way to group odds is by making 20 groups with an approximately equal number of observations in each category. Figure 5 plots the objective and subjective probability in each category using the odds grouping method that divides odds into 20 groups with approximately equal numbers of observations. The method is applied for all odds quoted one day before kickoff and for each of three types of bets (bet on a home win, a draw, or an away win). Each category contains approximately 1150 observations in the whole sample and 383 observations in each sub-sample of different types. Because of the large number of observations in each category, the 95% confidence interval is smaller; therefore, the favorite-longshot bias is easier to be observed.

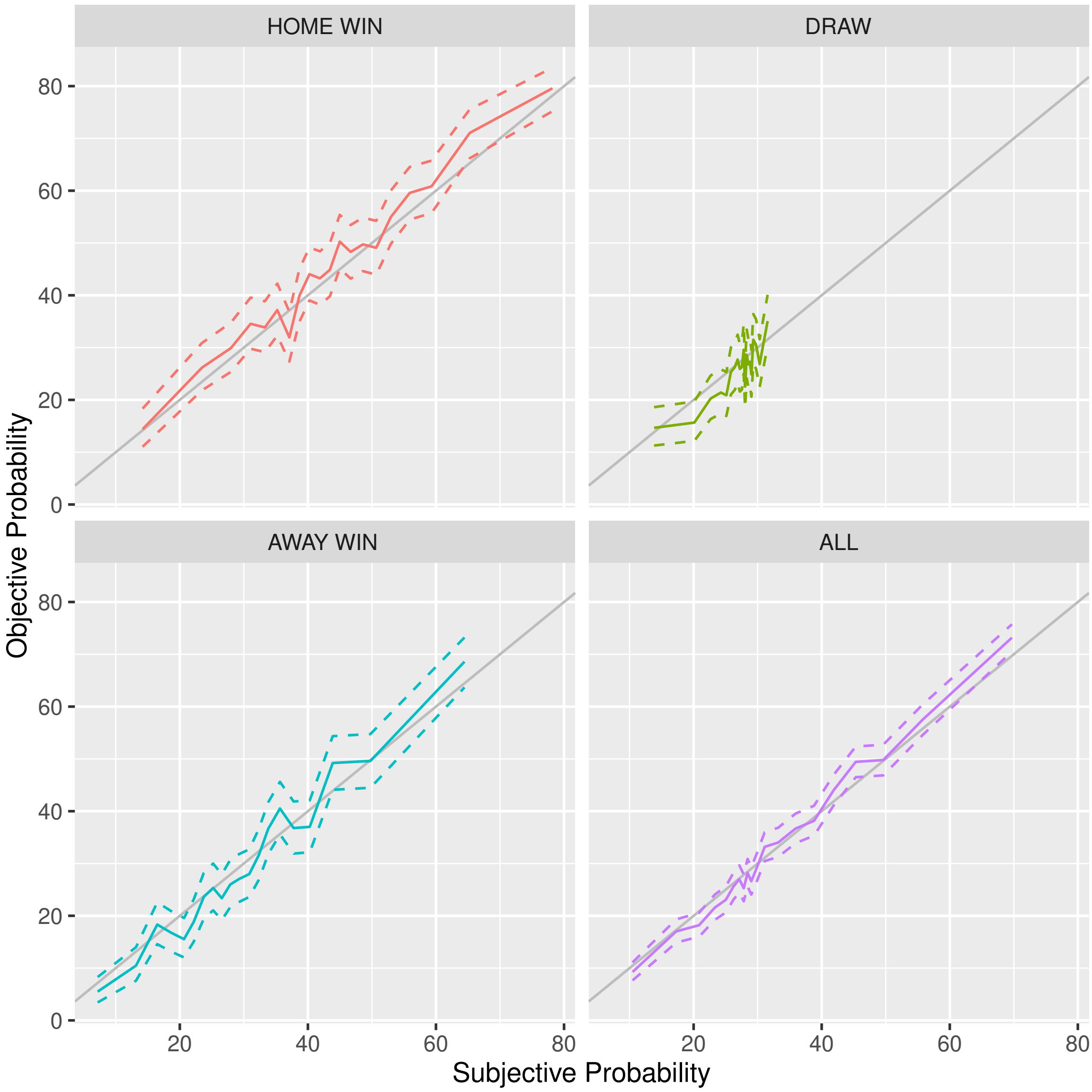

Figure 5: A comparison of Objective and subjective probabilities (Alternative specification: equal number of observations in each category)

Dashed lines are upper and lower bounds of the 95% confidence interval. Each category contains approximately 1,150 observations for whole sample and around 383 observations in each sub-group of different types (betting on a home win, a draw, or an away win). The ranges of subjective probabilities in each category is shown in Table 7.

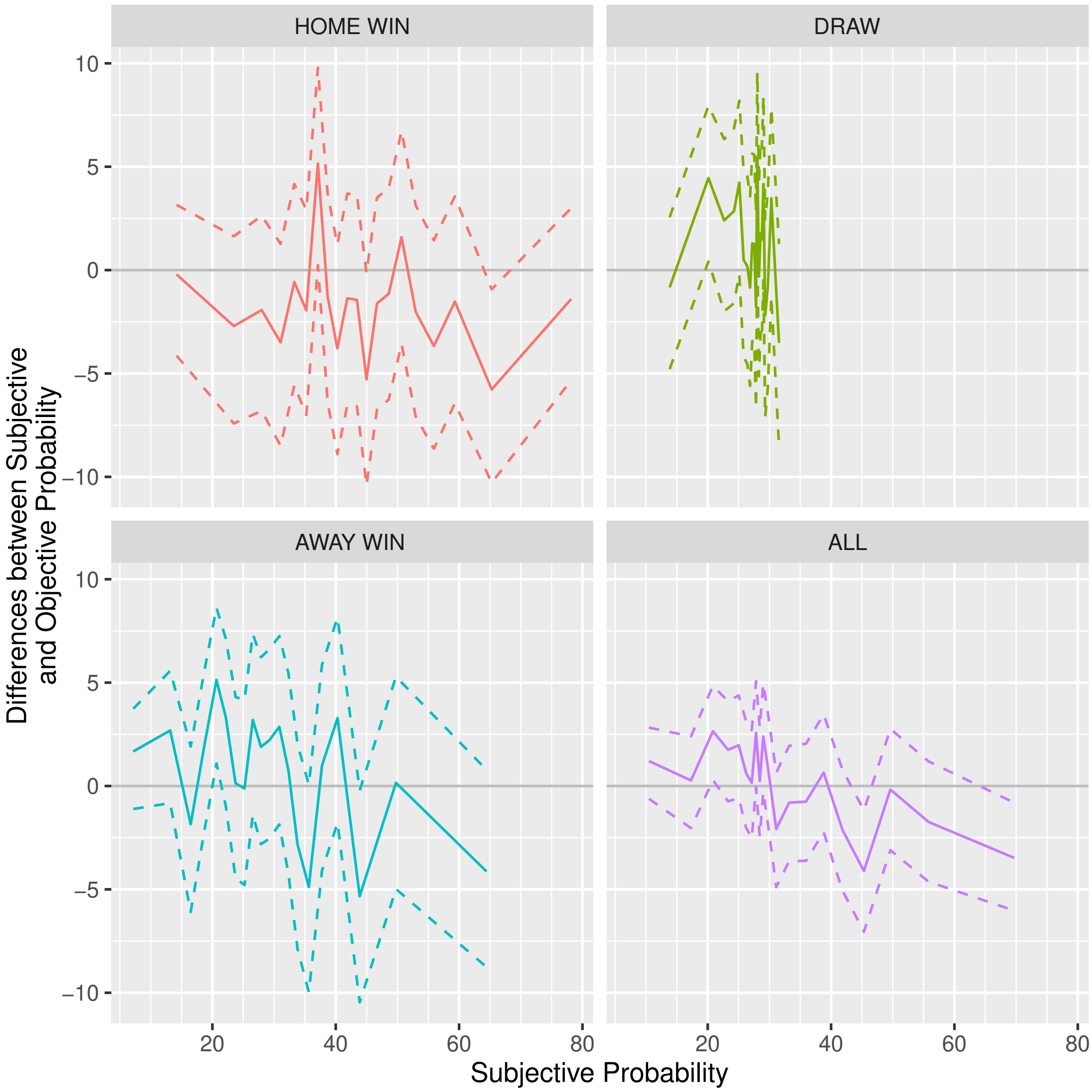

Figure 6: Difference between objective and subjective probabilities (Alternative specification: equal number of observations in each category)

Dashed line are upper and lower bounds of the 95% confidence interval. Each category contains approximately 1,150 observations for whole sample and around 383 observations in each sub-group of different types (betting on a home win, a draw, or an away win). The ranges of subjective probabilities in each category is shown in Table 7.

Table 7: Binomial test: Is objective probability equal to subjective probability one day before kickoff? Alternative Specification: Equal observations

p-value of an exact binomial test whether the objective probability equal to the subjective probability in that category. Bets are group by equal observation in each category

ALL		Away Win		Home Win		Draw	
Group	p-value	Group	p-value	Group	p-value	Group	p-value
[1.58,14.9]	0.2476	[1.58,10.9]	0.4284	[3.73,19.9]	1	[4.6,18]	0.6565
(14.9,19.4]	0.7549	(10.9,15.2]	0.0957	(19.9,26.2]	0.1474	(18,21.5]	0.0152
(19.4,22.1]	0.0199	(15.2,17.5]	0.4081	(26.2,29.7]	0.4595	(21.5,23.7]	0.5413
(22.1,24.3]	0.2639	(17.5,19.9]	0.3584	(29.7,32.3]	0.2931	(23.7,24.8]	0.0834
(24.3,25.7]	0.1517	(19.9,21.4]	0.0078	(32.3,34.3]	0.4467	(24.8,25.5]	0.0867
(25.7,26.8]	0.4402	(21.4,23.1]	0.2478	(34.3,36.2]	0.8725	(25.5,26.2]	0.5231
(26.8,27.5]	0.842	(23.1,24.6]	0.5431	(36.2,37.9]	0.1709	(26.2,26.7]	0.4842
(27.5,28.1]	0.0521	(24.6,25.9]	0.3784	(37.9,39.5]	0.9579	(26.7,27.1]	0.8628
(28.1,28.7]	0.6957	(25.9,27.2]	0.3565	(39.5,41.1]	0.0538	(27.1,27.4]	0.5664
(28.7,29.3]	0.0633	(27.2,28.6]	0.6051	(41.1,42.6]	0.6399	(27.4,27.7]	0.8635
(29.3,30.3]	0.217	(28.6,30]	0.1029	(42.6,44.2]	0.3046	(27.7,27.9]	0.6082
(30.3,32]	0.151	(30,31.6]	0.5428	(44.2,45.7]	0.2153	(27.9,28.1]	0.0995
(32,34.4]	0.7079	(31.6,33]	0.7419	(45.7,47.5]	0.504	(28.1,28.5]	0.3622
(34.4,37.4]	0.4214	(33,34.6]	0.3041	(47.5,49.6]	0.7983	(28.5,28.7]	0.2385
(37.4,40.2]	0.7161	(34.6,36.6]	0.1366	(49.6,51.7]	1	(28.7,28.9]	0.0768
(40.2,43.5]	0.1124	(36.6,38.9]	0.5598	(51.7,54.4]	0.2819	(28.9,29.1]	0.1417
(43.5,47.2]	0.0126	(38.9,41.8]	0.7541	(54.4,57.5]	0.3027	(29.1,29.5]	0.3991
(47.2,52.2]	0.8365	(41.8,46.2]	0.1801	(57.5,61.5]	0.4661	(29.5,30]	0.6197
(52.2,59.6]	0.297	(46.2,54.8]	0.7591	(61.5,69.5]	0.0241	(30,30.7]	0.3382
(59.6,93.8]	0.0084	(54.8,87.3]	0.0876	(69.5,93.8]	0.5767	(30.7,63.2]	0.1511

In Figure 5, the subjective probability approximately estimates the actual objective probability, described by the solid line move closely along the 45-degree line. It is also confirmed by the binomial tests for the hypothesis that subjective and objective probability is equal in each category (See Table 7 for the p-value of these tests). In Figure 6, the downward trend of the line illustrates the favorite - longshot bias: odds provide a favorable probability of a win for the longshot team and underestimates the probability that the favorite team wins. The exception is for betting on home wins, where favorites and longshots are both underestimated by odds quoted by bookmakers and favorites are more underestimated than longshots. This exception is evidence of the existence of home-field bias in this data.

6.2 Odds grouping: alternative number of categories

The second alternative specification is altering the number of categories. In the primary analysis, there are 20 groups of odds. This part examines whether the results significantly change if odds are divided into forty groups or only ten groups. When subjective probabilities are divided into smaller groups, the number of observations in a group decreases, leading to a higher estimation error. In other words, for analysis with more number of categories, the 95% confidence interval will be broader. In contrast, when the bin steps are larger, the number of observations in each category increases, lowering the estimation error and narrowing the confidence interval of objective probability estimation. However, larger bin steps also lead to losing information on extreme favorites and longshots.

Figure 7, 8 and Figure 9, 10 illustrate the objective and subjective probability when odds are divided into 40 or 10 categories from low to high subjective probability. Figure 7 depicts the objective and subjective probability when odds are divided into 40 categories with the steps of 2.5%. Figure 8 presents the differences between subjective and objective probabilities in each category shown in Figure 7. The confidence intervals are large, but the favorite-longshot bias pattern still presents: compare to the objective probabilities, the subjective probabilities are higher for longshots and lower for favorites. Figure 9 and Figure 10 displays the case of 10 categories. When altering the numbers of groups, the favorite longshot bias still presents: odds underestimate the favorite team's victory chance and overestimate the longshot team's victory chance. However, the bias is still statistically insignificant as presented in Table 8 and Table9. It is also notable that in each of these above scenarios, subjective probabilities for home teams' victories underestimate the actual winning possibility of these teams more than that for away teams', suggesting the existence of home-field bias.

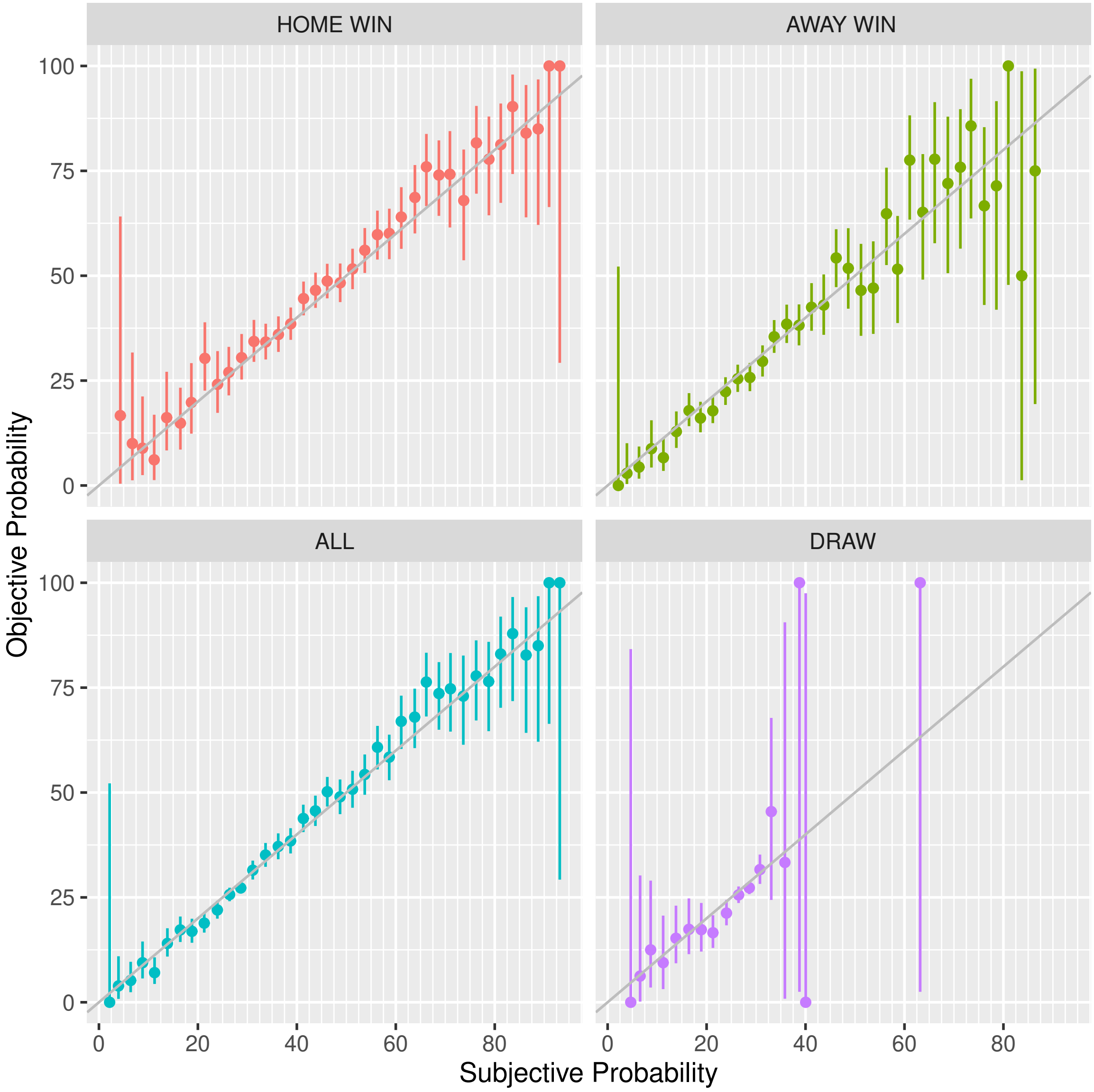

Figure 7: A Comparison of Objective and Subjective Probability (Alternative specification: group into 40 categories)

Reproduce results in the previous section in the new scenario when the subjective probability is divided into 40 categories with steps of 2.5% from 0-2.5%, 2.5-5%, ..., 92.5-95%

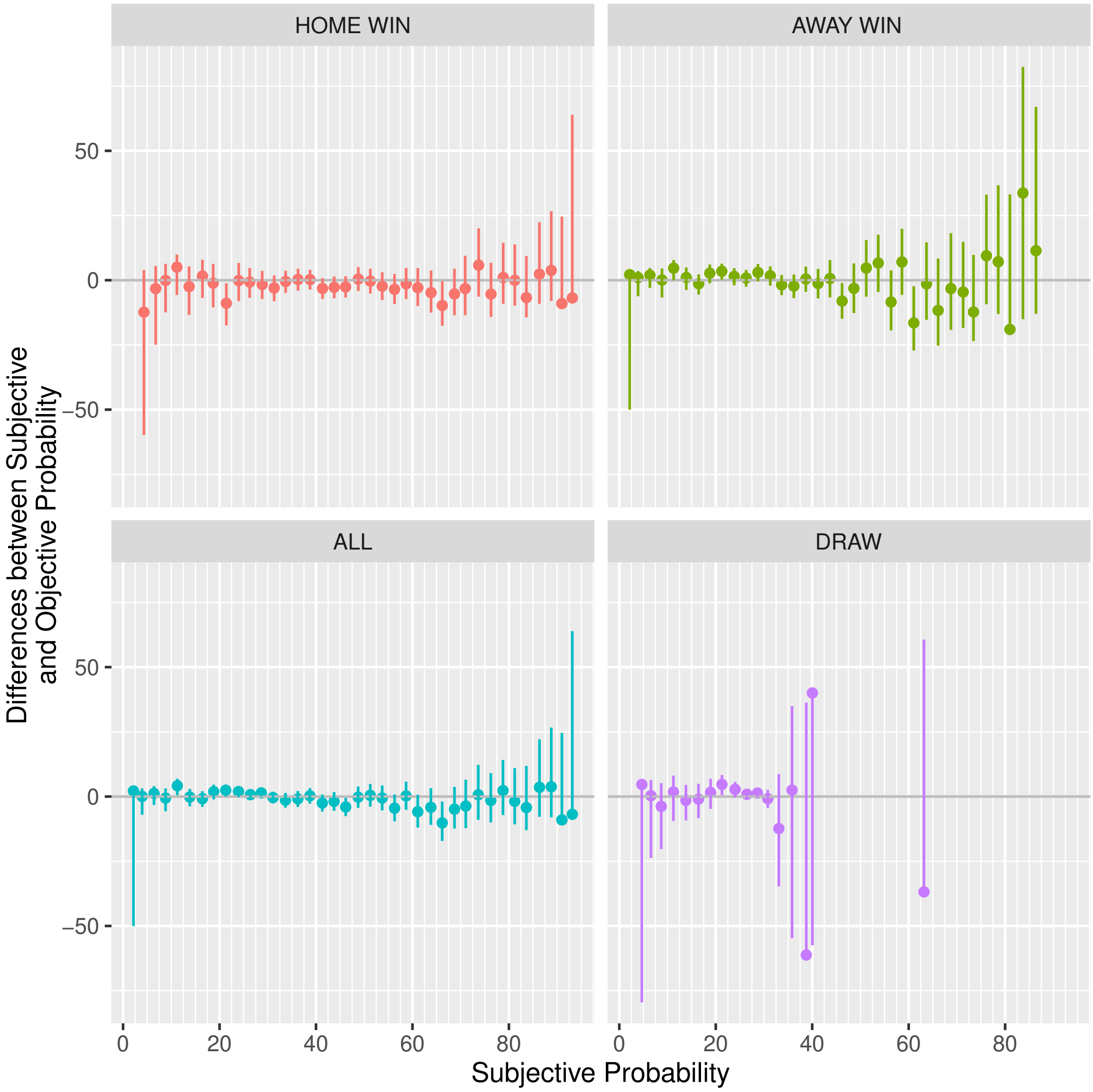

Figure 8: Differences between Subjective and Objective Probability (Alternative specification: group into 40 categories)

Reproduce results in the previous section in the new scenario when the subjective probability is divided into 40 categories with steps of 2.5% from 0-2.5%, 2.5-5%, ..., 92.5-95%

Table 8: Binomial test: Is objective probability equal to subjective probability? Alternative Specification: No. category is 40

the p-values of exact binomial tests that whether the objective probability equal to the subjective probability in the same category in each category where subjective probability is divided by steps of 2.5%.

Category	All	Home	Draw	Away	Category	All	Home	Draw	Away
(0,2.5]	1			1	(50,52.5]	0.8602	0.8851		0.3906
(2.5,5]	1	0.2328	1	1	(52.5,55]	0.8468	0.3889		0.2326
(5,7.5]	0.6412	0.3957	1	0.4806	(55,57.5]	0.0881	0.2574		0.1877
(7.5,10]	0.7026	1	0.3561	1	(57.5,60]	0.9556	0.6645		0.2561
(10,12.5]	0.0237	0.3639	0.8295	0.0457	(60,62.5]	0.0749	0.4854		0.0186
(12.5,15]	0.8896	0.5949	0.5955	0.714	(62.5,65]	0.275	0.2807	1	1
(15,17.5]	0.5897	0.7884	0.7307	0.4512	(65,67.5]	0.016	0.0378		0.2283
(17.5,20]	0.2237	0.7932	0.6341	0.1678	(67.5,70]	0.2884	0.2817		0.8316
(20,22.5]	0.049	0.019	0.0269	0.0346	(70,72.5]	0.4892	0.6752		0.685
(22.5,25]	0.0802	1	0.083	0.4036	(72.5,75]	0.8953	0.3497		0.3211
(25,27.5]	0.387	0.8262	0.3946	0.6413	(75,77.5]	0.8961	0.3663		0.3096
(27.5,30]	0.0413	0.561	0.0881	0.0891	(77.5,80]	0.6563	0.8677		0.5163
(30,32.5]	0.7315	0.2143	0.6246	0.3807	(80,82.5]	0.8612	1		0.591
(32.5,35]	0.3113	0.8126	0.2572	0.3636	(82.5,85]	0.6422	0.4653		0.2995
(35,37.5]	0.5706	0.9269	1	0.3293	(85,87.5]	0.5852	0.7685		0.4427
(37.5,40]	0.8733	0.9351	0.3877	0.8773	(87.5,90]	0.4853	0.4853		
(40,42.5]	0.1304	0.117	1	0.6844	(90,92.5]	1	1		
(42.5,45]	0.3042	0.2017		0.8847	(92.5,95]	1	1		
(45,47.5]	0.0231	0.2144		0.0192					
(47.5,50]	0.934	0.854		0.51					

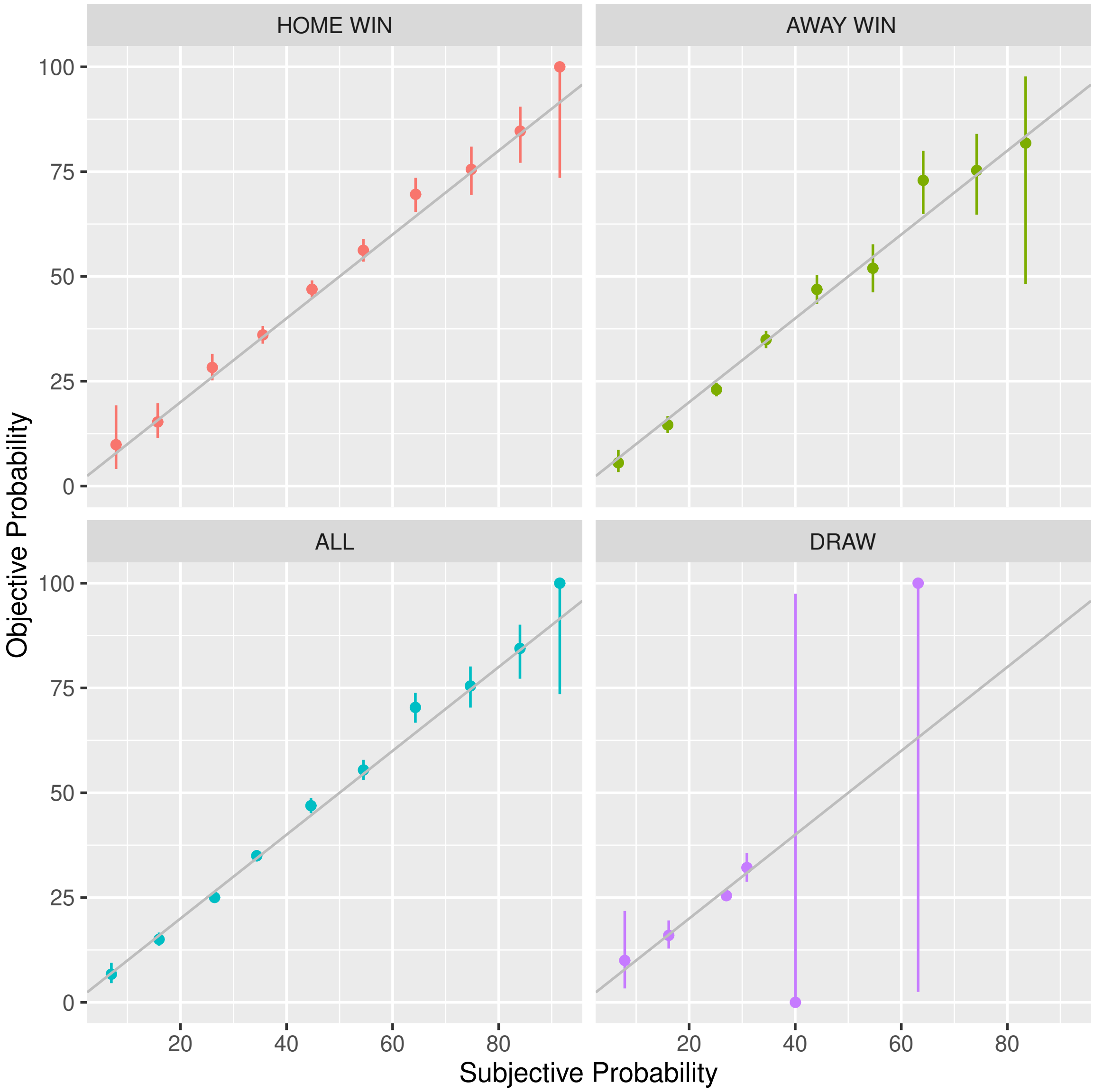

Figure 9: A Comparison of Objective and Subjective Probability (Alternative specification: group into 10 categories)

Reproduce results in the previous section in the new scenario when the subjective probability is divided into 10 categories with steps of 10% from 0-10%, 10-20%, ..., 90-100%

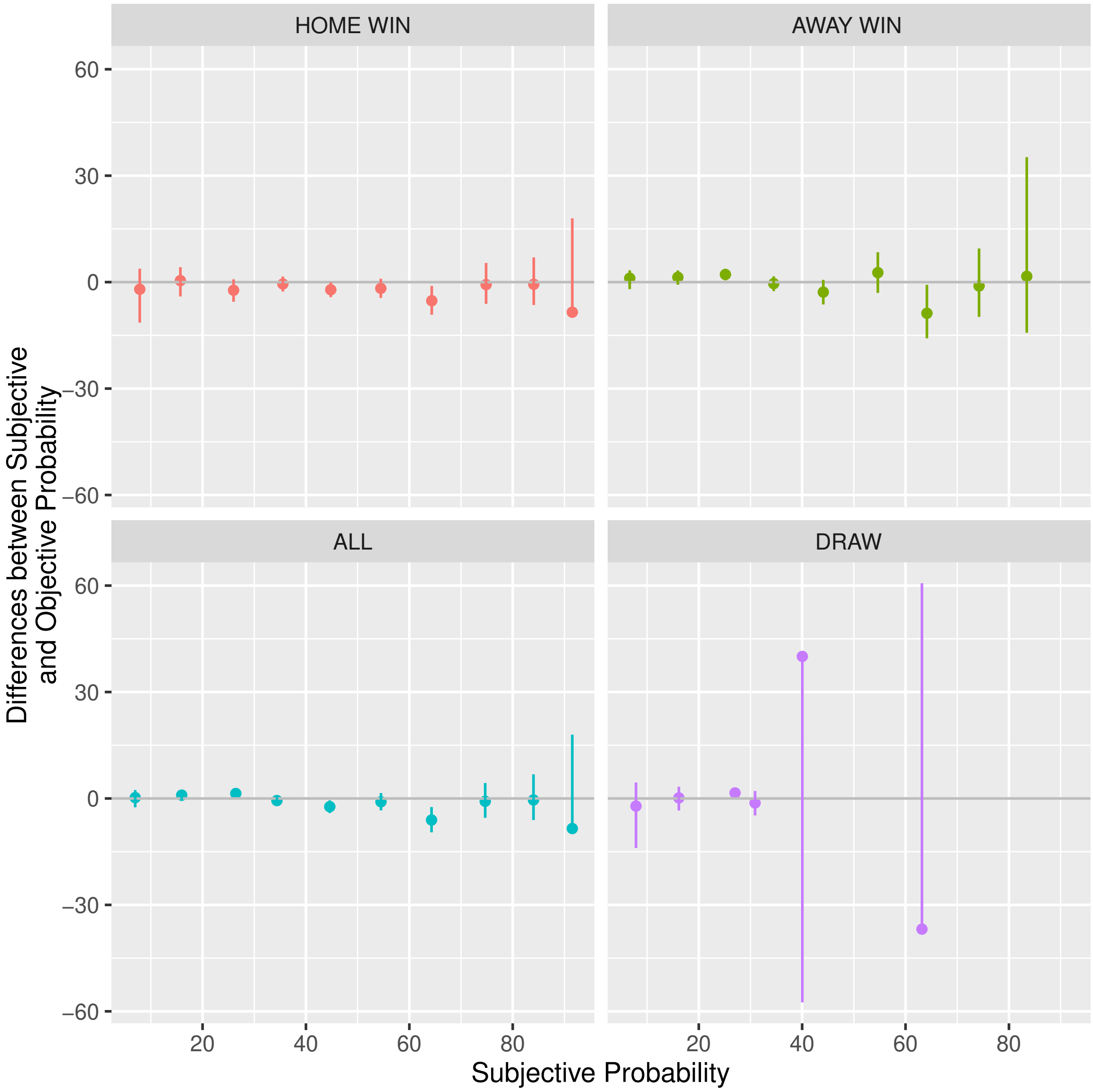

Figure 10: Differences between Subjective and Objective Probability (Alternative specification: group into 10 categories)

Reproduce results in the previous section in the new scenario when the subjective probability is divided into 10 categories with steps of 10% from 0-10%, 10-20%, ..., 90-100%

Table 9: Binomial test: Is objective probability equal to subjective probability? Alternative Specification: No. category is 10

the p-values of exact binomial tests that whether the objective probability equal to the subjective probability in the same category in each category where subjective probability is divided by steps of 10%

Category	All	Home	Draw	Away
(0,10]	0.9259	0.5048	0.5933	0.5034
(10,20]	0.2629	0.9382	1	0.1995
(20,30]	0.0015	0.1376	0.005	0.0107
(30,40]	0.3947	0.6248	0.4712	0.6933
(40,50]	0.0108	0.0429	1	0.1064
(50,60]	0.4565	0.2053		0.3585
(60,70]	0.0011	0.0128	1	0.0296
(70,80]	0.7953	0.879		0.9016
(80,90]	1	1		0.7018
(90,100]	0.6161	0.6161		

6.3 Alternative lag

This section investigates the favorite-longshot bias with odds quoted seven days instead of one day before kickoff. Figure 11 depicts the subjective and objective probability in each category seven days before kickoff. Points above the 45-degree lines describe categories where objective probability is higher than subjective probability, while points below the lines describe categories where objective probability is lower than subjective probability. Figure 12 illustrates the difference between subjective and objective probability in each category. The results are consistent with the previous section. Except for the outliers due to small observations, the objective probability in each category is well predicted by the average subjective probability of that same group. The differences between subjective and objective probabilities are insignificantly different from zero. However, the difference decreases as subjective probability increases, which supports the existence of a favorite-longshot bias.

Table 10 and Table 11 presents the results of regressions on sample of odds seven day before kickoff. Compare to Table 5 and Table 6, the results in Table 10 and 11 are very similar. At a 5% confidence interval, the favorite-longshot bias exists in this market seven days before kickoff. β_1 in all regression one day before kickoff is smaller than that seven days before kickoff, suggesting that the favorite-longshot bias is less pronounced as the date of match approaches.

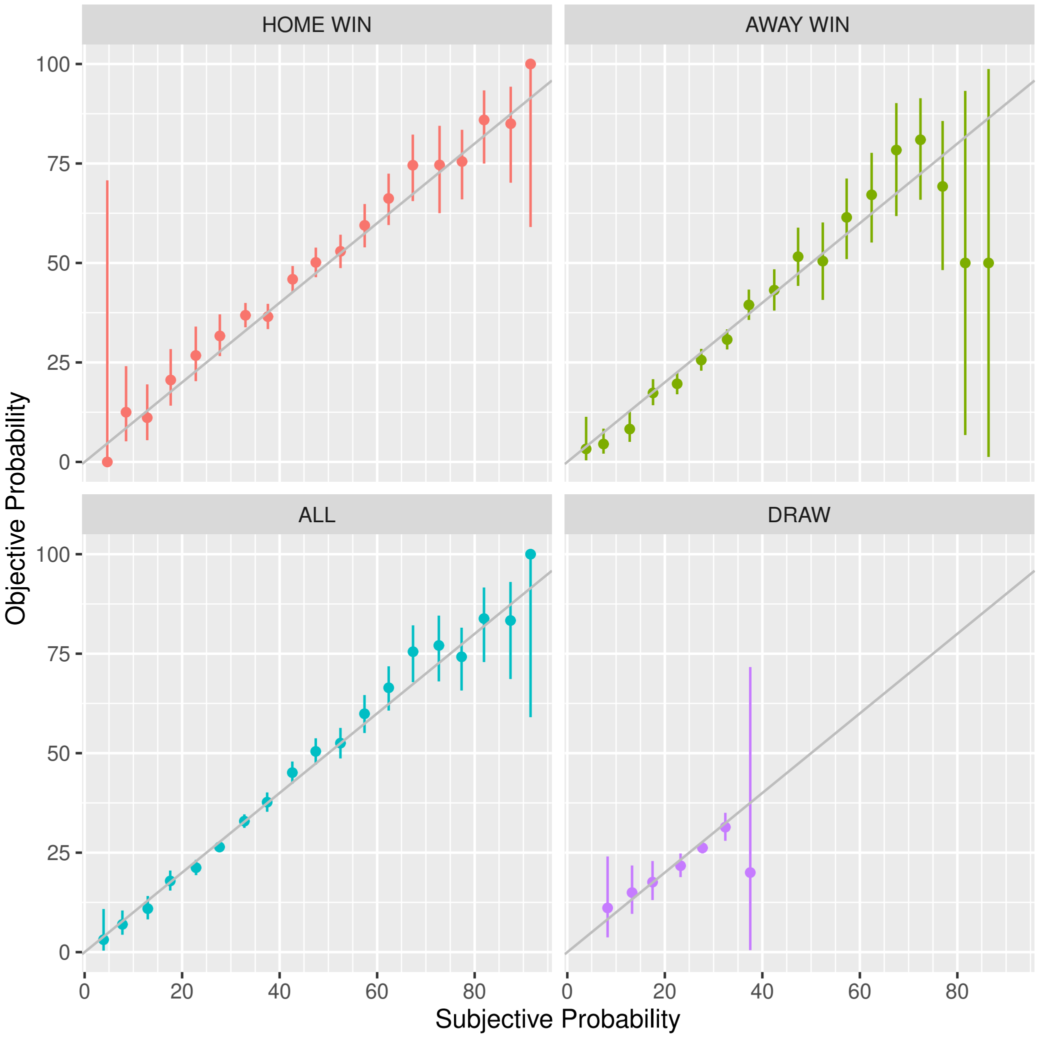

Figure 11: A Comparison of Objective and Subjective Probability 7 days before kickoff

Reproduce results in the previous section on odds quoted 7 days before kickoff date of matches. The categories here is constructed by dividing subjective probabilities into 19 groups with steps of 5%: 0-5%, 5-10%, 10-15%, ..., 90-95% (There is no subjective probability which is higher than 95%). The estimated objective probability and confidence interval is shown in Appendices A2

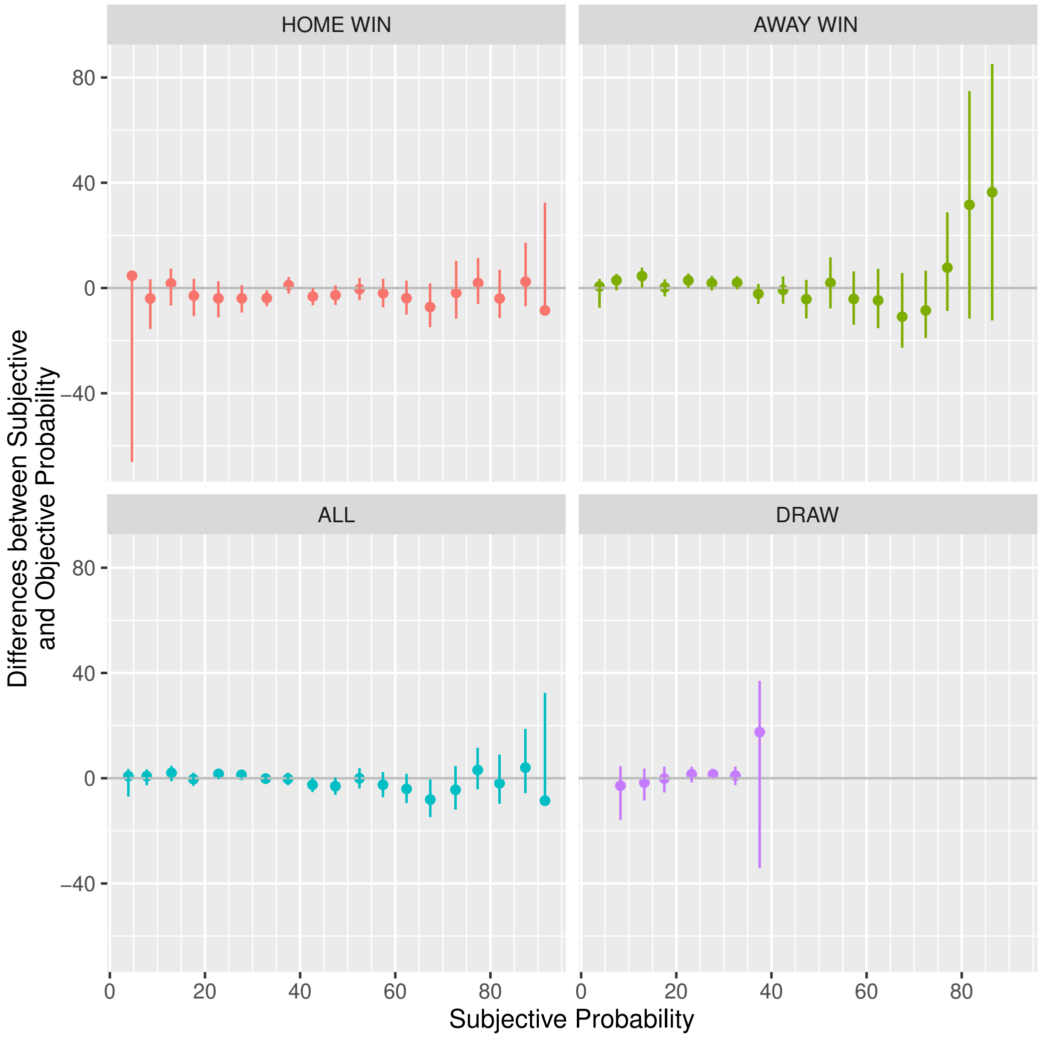

Figure 12: Differences between Subjective and Objective Probability 7 days before kickoff

Reproduce results in the previous section on odds quoted 7 days before kickoff date of matches. The categories here is constructed by dividing subjective probabilities into 19 groups with steps of 5%: 0-5%, 5-10%, 10-15%, ..., 90-95% (There is no subjective probability which is higher than 95%). The estimated objective probability and confidence interval is shown in Appendices A2

Table 10: Regression summary: The favorite-longshot bias in the whole market seven day before kickoff

Coefficient		Estimate	Std. Error	p-value
A. Mixed-type Odds				
Constant	β_0	-2.359	1.149	0.04004
Subj. Probability	β_1	0.084	0.029	0.00368
N =	11,040			
B. Home Win Odds				
Constant	β_0	2.211	2.008	0.271
Subj. Probability	β_1	-0.0003	0.0444	0.995
N =	5,520			
C. Away Win Odds				
Constant	β_0	-4.413	1.513	0.00356
Subj.Probability	β_1	0.120	0.046	0.00847
N =	5,520			
D. Draw Odds				
Constant	β_0	2.946	3.617	0.415
Subj.Probability	β_1	-0.166	0.134	0.217
N =	5,520			

Table 11: Regression summary: The favorite-longshot bias and home bias estimates seven day before kickoff

Coefficient		Estimate	Std. Error	p-value
Constant	β_0	-4.413	1.588	0.00546
p	β_1	0.120	0.048	0.01209
Home	β_2	6.624	2.493	0.00789
$p \times Home$	β_3	-0.121	0.064	0.05966

6.4 Average odds as the representative odds from bookmakers

In this section, the averages of odds are chosen as the synthetic odds representing all odds quoted simultaneously from all bookmakers. Figure 13 and Figure 14 describe the results. Compare to Figure 3 and Figure 4, the results are consistent. The subjective probability still successfully estimates the true objective probability in each category. The difference between subjective and objective probability is small and insignificantly different than zero; however, the difference decreases as subjective probability increases, suggesting the presence of favorite-longshot bias.

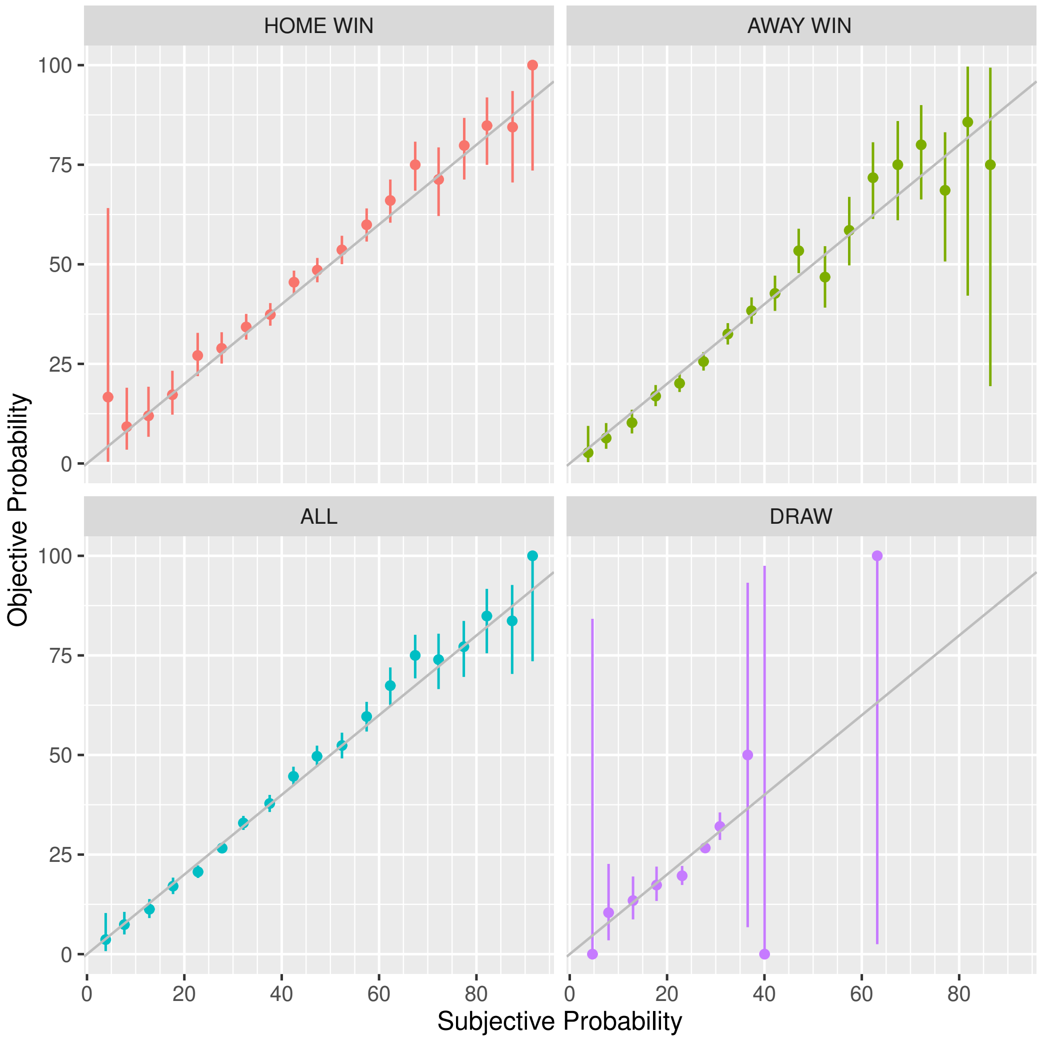

Figure 13: A comparison of Objective and Subjective Probability (Alternative specification: Average of odds as the synthetic odds)

Reproduce results in the previous section with synthetic odds that are the averages of odds quoted by different bookmakers. The categories here is constructed by dividing subjective probabilities into 19 groups with a steps of 5%: 0-5%, 5-10%, 10-15%, ..., 90-95% (There is no subjective probability which is higher than 95%).

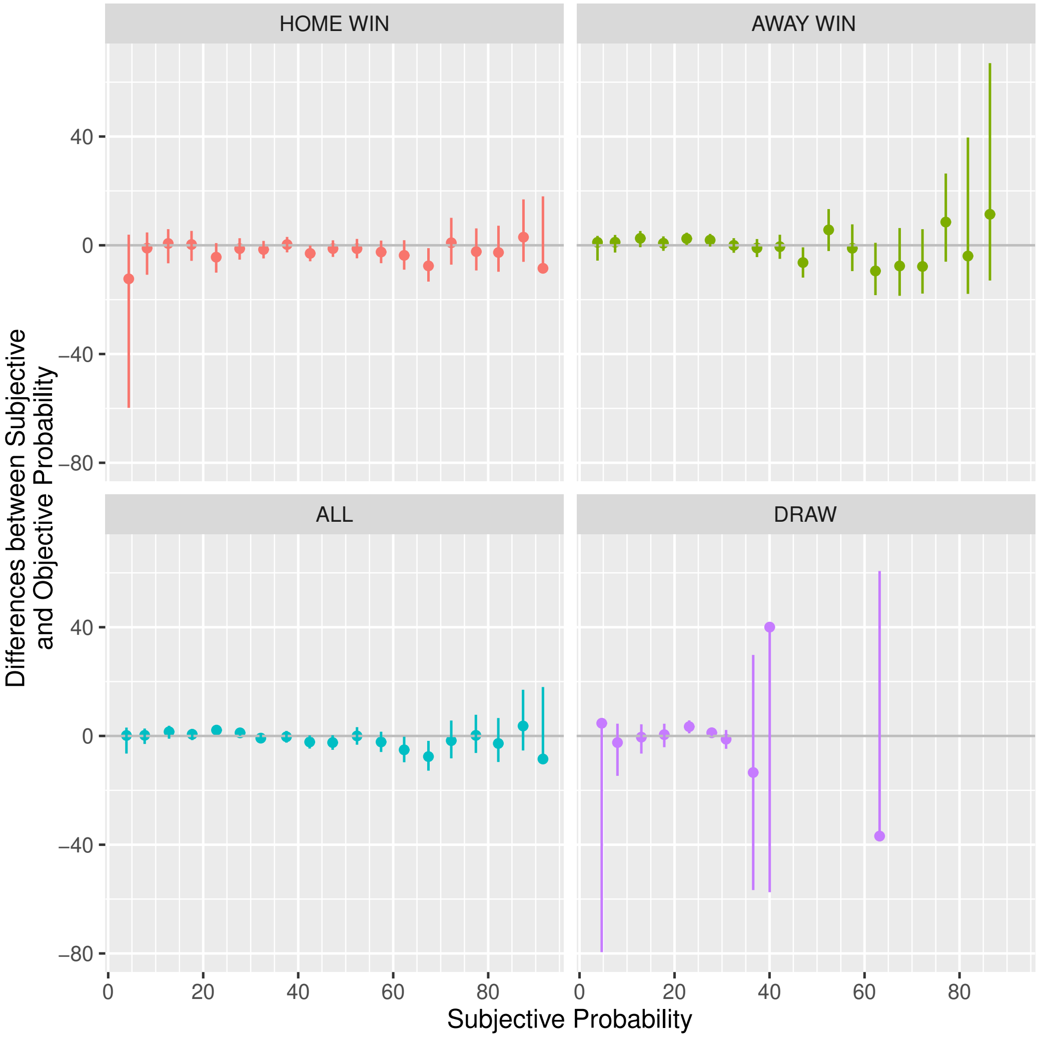

Figure 14: Difference between Subjective and Objective Probability (Alternative specification: Average of odds as the synthetic odds)

Reproduce results in the previous section with synthetic odds that are the averages of odds quoted by different bookmakers. The categories here is constructed by dividing subjective probabilities into 19 groups with a steps of 5%: 0-5%, 5-10%, 10-15%, ..., 90-95% (There is no subjective probability which is higher than 95%).

7 Conclusion

In this thesis, I examine the favorite-longshot bias in the European online football betting market. This thesis extends the previous literature on this field of research by study the favorite-longshot bias in the emerging online betting market.

A total of 7,250 football matches and associated online betting odds from 11 well-known football competition in Europe during season 2016/17 and 2017/18 have been analyzed. The analysis shows that bookmakers' odds are good predictors for the actual winning chance of teams in football matches. The subjective probabilities implied by bookmakers' odds are not statistically different from the objective probabilities of different matches' outcomes. The results are consistent even with changes in specifications. However, there is some evidence that favorite-longshot bias exists in the European online football betting markets. The regression of the difference between subjective and objective probability on subjective probability on the whole data set shows the presence of favorite-longshot bias in these markets, and the bias exists at a 5% significant level.

In European online football betting markets, there is an existence of strong home bias: bookmakers tend to shorten the odds for the victory of away teams more relative to odds for the victory of home teams. The implied probability for home win is lower on average by approximately 3.9 percentage points compared to that for away win one day before kickoff date. When taking this into account, these online markets show the presence of a favorite - longshot bias in both betting on home wins and away wins odds.

References

Ali, M. M. (1977). Probability and utility estimates for racetrack bettors. *Journal of Political Economy*, 85(4):803–815.

Brown, L. D., Cai, T. T., and DasGupta, A. (2001). Interval estimation for a binomial proportion. *Statistical Science*, 16(2):101–117.

Busche, K. (1994). Efficient market results in an asian setting. In *Efficiency of racetrack betting markets*, pages 615–616. New York: Academic Press.

Busche, K. and Hall, C. D. (1988). An exception to the risk preference anomaly. *Journal of business*, pages 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*, 47(1):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*, 55(3):263–273.

Clopper, C. J. and Pearson, E. S. (1934). The use of confidence or fiducial limits illustrated in the case of the binomial. *Biometrika*, 26(4):404–413.

EGBA (2019). European online gambling key figures 2018.

Elaad, G. (2020). Home-field advantage and biased prediction markets in english soccer. *Applied Economics Letters*, 27(14):1170–1174.

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*, 34(10):1309–1317.

Golec, J. and Tamarkin, M. (1991). The degree of inefficiency in the football betting market: Statistical tests. *Journal of Financial Economics*, 30(2):311 – 323.

Griffith, R. M. (1949). Odds adjustments by american horse-race bettors. *The American Journal of Psychology*, 62(2):290–294.

Kahneman, D. (1979). Tversky a.(1979). *Prospect theory: an analysis of decision under risk*, pages 263–292.

Kuypers, T. (2000). Information and efficiency: an empirical study of a fixed odds betting market. *Applied Economics*, 32(11):1353–1363.

Levitt, S. D. (2004). Why are gambling markets organised so differently from financial markets?*. *The Economic Journal*, 114(495):223–246.

Makropoulou, V. and Markellos, R. N. (2011). Optimal price setting in fixed-odds betting markets under information uncertainty. *Scottish Journal of Political Economy*, 58(4):519–536.

Ottaviani, M. and Sørensen, P. N. (2008). The favorite-longshot bias: An overview of the main explanations. In *Handbook of Sports and Lottery markets*, pages 83–101. Elsevier.

Rossi, M. et al. (2011). Match rigging and the favorite long-shot bias in the italian football betting market. *International Journal of Sport Finance*, 6(4):317–334.

Sauer, R. D. (1998). The economics of wagering markets. *Journal of Economic Literature*, 36(4):2021–2064.

Shin, H. S. (1992). Prices of state contingent claims with insider traders, and the favourite-longshot bias. *The Economic Journal*, 102(411):426–435.

Shin, H. S. et al. (1991). Optimal betting odds against insider traders. *Economic Journal*, 101(408):1179–1185.

Shin, H. S. et al. (1993). Measuring the incidence of insider trading in a market for state-contingent claims. *Economic Journal*, 103(420):1141–1153.

Snowberg, E. and Wolfers, J. (2008). Chapter 7 - examining explanations of a market anomaly: Preferences or perceptions? In Hausch, D. B. and Ziemba, W. T., editors, *Handbook of Sports and Lottery Markets*, Handbooks in Finance, pages 103 – 136. Elsevier, San Diego.

Snowberg, E. and Wolfers, J. (2010). Explaining the favorite–long shot bias: Is it risk-love or misperceptions? *Journal of Political Economy*, 118(4):723–746.

Swidler, S. and Shaw, R. (1995). Racetrack wagering and the "uninformed" bettor: A study of market efficiency. *The Quarterly Review of Economics and Finance*, 35(3):305–314.

Thaler, R. H. and Ziemba, W. T. (1988). Anomalies: Parimutuel betting markets: Racetracks and lotteries. *The Journal of Economic Perspectives*, 2(2):161–174.

Wolfers, J. and Zitzewitz, E. (2004). Prediction markets. *The Journal of Economic Perspectives*, 18(2):107–126.

Woodland, L. M. and Woodland, B. M. (1994). Market efficiency and the favorite-longshot bias: the baseball betting market. *The Journal of Finance*, 49(1):269–279.

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*, 55(2):113–123.

Appendices

These appendices provide the calculation of estimated objective probability and their 95% confidence interval using in different figures of my thesis. For each match in the data set, there is a set of 3 synthetic odds for betting on 3 different outcome: home win, draw and away win. Each odd associate with a bet on a specific outcome of a specific game. This bet has the result of either win or loss, which is determined by the actual result of that game. The estimated objective probability is calculated by following steps. First, subjective probabilities implied by the synthetic odds are computed and normalized to remove bookmakers' margins (the overround). Then, based on the subjective probabilities, all bets are divided into different categories. Next, in each category, the total number of bets (N) and the number of winning bets (Win) is counted. Because result of these bet is either win or loss, this is a Bernoulli trial. The objective probability is estimated as the portions of bets that result is a win, which is equal to Win divided by N. The 95% confidence interval for the objective probability is a binomial proportion confidence interval, which is computed using Clopper-Pearson method. This method is also called an "exact" method because it is based on the cumulative probability of the actual binomial distribution rather than an approximation. This 95% interval may be well above 95% depend on N and Win but it never has less than 95% coverage.

A1

This section provides 4 tables of the estimated objective probability and the 95% confidence interval in difference probability categories one day before kickoff. These results is used in Figure 3 and Figure 4

Table A1: Objective Probability and Confidence Interval of Different Probability Categories 1 Day before Kickoff

ALL BETS (DAY -1)					
Probability Category	N	Win	Subjective Probability (%)	Objective Probability(%)	Confidence Interval
(0,5]	82	3	3.84	3.66	[0.76;10.32]
(5,10]	364	27	7.67	7.42	[4.94;10.61]
(10,15]	718	81	12.81	11.28	[9.06;13.83]
(15,20]	1318	225	17.68	17.07	[15.08;19.21]
(20,25]	2633	544	22.8	20.66	[19.13;22.26]
(25,30]	7190	1914	27.76	26.62	[25.6;27.66]
(30,35]	2791	919	32.12	32.93	[31.18;34.71]
(35,40]	2007	759	37.52	37.82	[35.69;39.98]
(40,45]	1667	744	42.41	44.63	[42.23;47.06]
(45,50]	1383	687	47.25	49.67	[47.01;52.34]
(50,55]	947	496	52.38	52.38	[49.14;55.6]
(55,60]	689	411	57.44	59.65	[55.88;63.34]
(60,65]	402	271	62.31	67.41	[62.59;71.98]
(65,70]	256	192	67.42	75	[69.23;80.18]
(70,75]	165	122	72.22	73.94	[66.54;80.45]
(75,80]	149	115	77.4	77.18	[69.6;83.65]
(80,85]	86	73	82.13	84.88	[75.54;91.7]
(85,90]	49	41	87.34	83.67	[70.34;92.68]
(90,95]	12	12	91.53	100	[73.54;100]

Table A2: Objective Probability and Confidence Interval of Different Probability Categories for Betting on Home Team 1 Day before Kickoff

BETTING ON HOME TEAM WIN (DAY -1)					
Probability Category	N	Win	Subjective Probability (%)	Objective Probability(%)	Confidence Interval
(0,5]	6	1	4.32	16.67	[0.42;64.12]
(5,10]	65	6	8.18	9.23	[3.46;19.02]
(10,15]	117	14	12.63	11.97	[6.7;19.26]
(15,20]	197	34	17.54	17.26	[12.26;23.27]
(20,25]	273	74	22.74	27.11	[21.92;32.79]
(25,30]	533	154	27.67	28.89	[25.08;32.94]
(30,35]	861	295	32.7	34.26	[31.09;37.54]
(35,40]	1150	430	37.65	37.39	[34.59;40.26]
(40,45]	1167	531	42.52	45.5	[42.62;48.41]
(45,50]	1059	514	47.31	48.54	[45.49;51.59]
(50,55]	776	416	52.36	53.61	[50.03;57.16]
(55,60]	554	332	57.45	59.93	[55.71;64.04]
(60,65]	309	204	62.31	66.02	[60.44;71.29]
(65,70]	204	153	67.42	75	[68.47;80.78]
(70,75]	115	82	72.23	71.3	[62.12;79.35]
(75,80]	114	91	77.49	79.82	[71.28;86.76]
(80,85]	79	67	82.16	84.81	[74.97;91.9]
(85,90]	45	38	87.42	84.44	[70.54;93.51]
(90,95]	12	12	91.53	100	[73.54;100]

Table A3: Objective Probability and Confidence Interval of Different Probability Categories for Betting on Draw 1 Day before Kickoff

BETTING ON DRAW (DAY -1)					
Probability Category	N	Win	Subjective Probability (%)	Objective Probability(%)	Confidence Interval
(0,5]	2	0	4.67	0	[0;84.19]
(5,10]	48	5	7.98	10.42	[3.47;22.66]
(10,15]	171	23	13.01	13.45	[8.72;19.5]
(15,20]	317	55	17.84	17.35	[13.35;21.98]
(20,25]	1103	217	23.09	19.67	[17.37;22.14]
(25,30]	5262	1403	27.84	26.66	[25.47;27.88]
(30,35]	727	233	30.85	32.05	[28.67;35.58]
(35,40]	4	2	36.56	50	[6.76;93.24]
(40,45]	1	0	40.03	0	[0;97.5]
(60,65]	1	1	63.17	100	[2.5;100]

Table A4: Objective Probability and Confidence Interval of Different Probability Categories for Betting on Away Team 1 Day before Kickoff

BETTING ON AWAY TEAM WIN (DAY -1)					
Probability Category	N	Win	Subjective Probability (%)	Objective Probability(%)	Confidence Interval
(0,5]	74	2	3.78	2.7	[0.33;9.42]
(5,10]	251	16	7.48	6.37	[3.69;10.15]
(10,15]	430	44	12.79	10.23	[7.53;13.49]
(15,20]	804	136	17.65	16.92	[14.39;19.69]
(20,25]	1257	253	22.56	20.13	[17.94;22.45]
(25,30]	1395	357	27.49	25.59	[23.32;27.97]
(30,35]	1203	391	32.48	32.5	[29.86;35.23]
(35,40]	853	327	37.35	38.34	[35.06;41.69]
(40,45]	499	213	42.17	42.69	[38.3;47.16]
(45,50]	324	173	47.04	53.4	[47.8;58.93]
(50,55]	171	80	52.44	46.78	[39.13;54.55]
(55,60]	135	79	57.42	58.52	[49.73;66.93]
(60,65]	92	66	62.3	71.74	[61.39;80.64]
(65,70]	52	39	67.39	75	[61.05;85.97]
(70,75]	50	40	72.2	80	[66.28;89.97]
(75,80]	35	24	77.11	68.57	[50.71;83.15]
(80,85]	7	6	81.76	85.71	[42.13;99.64]
(85,90]	4	3	86.4	75	[19.41;99.37]

A2

This section provides 4 tables of the estimated objective probability and the 95% confidence interval in difference probability categories one day before kickoff. These results is used in Figure 11 and Figure 12

Table A5: Objective Probability and Confidence Interval of Different Probability Categories 7 Day before Kickoff

ALL BETS (DAY -7)					
Probability Category	N	Win	Subjective Probability (%)	Objective Probability(%)	Confidence Interval
(0,5]	64	2	3.9	3.12	[0.38;10.84]
(5,10]	303	22	7.76	7.26	[4.61;10.79]
(10,15]	463	48	12.96	10.37	[7.74;13.51]
(15,20]	893	158	17.56	17.69	[15.24;20.36]
(20,25]	1717	365	22.86	21.26	[19.34;23.27]
(25,30]	4932	1304	27.69	26.44	[25.21;27.69]
(30,35]	2745	897	32.73	32.68	[30.92;34.47]
(35,40]	1556	593	37.36	38.11	[35.69;40.58]
(40,45]	1195	538	42.58	45.02	[42.17;47.89]
(45,50]	859	430	47.39	50.06	[46.66;53.45]
(50,55]	637	336	52.49	52.75	[48.79;56.68]
(55,60]	405	242	57.38	59.75	[54.8;64.57]
(60,65]	283	189	62.32	66.78	[60.97;72.25]
(65,70]	154	118	67.27	76.62	[69.14;83.06]
(70,75]	106	83	72.64	78.3	[69.24;85.72]
(75,80]	128	93	77.25	72.66	[64.08;80.16]
(80,85]	69	60	81.84	86.96	[76.68;93.86]
(85,90]	44	35	87.21	79.55	[64.7;90.2]
(90,95]	7	7	91.33	100	[59.04;100]

Table A6: Objective Probability and Confidence Interval of Different Probability Categories for Betting on Home Team 7 Day before Kickoff

BETTING ON HOME TEAM WIN (DAY -7)					
Probability Category	N	Win	Subjective Probability (%)	Objective Probability(%)	Confidence Interval
(0,5]	2	0	4.51	0	[0;84.19]
(5,10]	57	7	8.39	12.28	[5.08;23.68]
(10,15]	91	9	12.83	9.89	[4.62;17.95]
(15,20]	131	28	17.66	21.37	[14.7;29.39]
(20,25]	165	45	22.74	27.27	[20.64;34.74]
(25,30]	303	96	27.73	31.68	[26.48;37.25]
(30,35]	904	333	32.96	36.84	[33.68;40.07]
(35,40]	897	325	37.53	36.23	[33.08;39.47]
(40,45]	839	386	42.66	46.01	[42.59;49.45]
(45,50]	680	337	47.39	49.56	[45.74;53.39]
(50,55]	532	284	52.49	53.38	[49.04;57.69]
(55,60]	312	185	57.42	59.29	[53.62;64.8]
(60,65]	213	143	62.3	67.14	[60.39;73.4]
(65,70]	114	86	67.24	75.44	[66.49;83.02]
(70,75]	66	50	72.81	75.76	[63.64;85.46]
(75,80]	101	75	77.36	74.26	[64.6;82.44]
(80,85]	64	57	81.84	89.06	[78.75;95.49]
(85,90]	42	34	87.25	80.95	[65.88;91.4]
(90,95]	7	7	91.33	100	[59.04;100]

Table A7: Objective Probability and Confidence Interval of Different Probability Categories for Betting on Draw 7 Day before Kickoff

BETTING ON DRAW (DAY -7)					
Probability Category	N	Win	Subjective Probability (%)	Objective Probability(%)	Confidence Interval
(5,10]	45	6	8.28	13.33	[5.05;26.79]
(10,15]	145	20	13.21	13.79	[8.63;20.5]
(15,20]	248	43	17.47	17.34	[12.84;22.64]
(20,25]	766	166	23.2	21.67	[18.8;24.76]
(25,30]	3695	962	27.74	26.04	[24.63;27.48]
(30,35]	617	192	32.35	31.12	[27.48;34.94]
(35,40]	4	0	35.93	0	[0;60.24]

Table A8: Objective Probability and Confidence Interval of Different Probability Categories for Betting on Away Team 7 Day before Kickoff

BETTING ON AWAY TEAM WIN (DAY -7)					
Probability Category	N	Win	Subjective Probability (%)	Objective Probability(%)	Confidence Interval
(0,5]	62	2	3.88	3.23	[0.39;11.17]
(5,10]	201	9	7.47	4.48	[2.07;8.33]
(10,15]	227	19	12.85	8.37	[5.11;12.76]
(15,20]	514	87	17.57	16.93	[13.79;20.45]
(20,25]	786	154	22.55	19.59	[16.87;22.54]
(25,30]	934	246	27.46	26.34	[23.54;29.29]
(30,35]	1224	372	32.74	30.39	[27.82;33.05]
(35,40]	655	268	37.13	40.92	[37.12;44.79]
(40,45]	356	152	42.4	42.7	[37.5;48.02]
(45,50]	179	93	47.39	51.96	[44.38;59.47]
(50,55]	105	52	52.46	49.52	[39.62;59.45]
(55,60]	93	57	57.24	61.29	[50.62;71.22]
(60,65]	70	46	62.36	65.71	[53.4;76.65]
(65,70]	40	32	67.37	80	[64.35;90.95]
(70,75]	40	33	72.36	82.5	[67.22;92.66]
(75,80]	27	18	76.82	66.67	[46.04;83.48]
(80,85]	5	3	81.79	60	[14.66;94.73]
(85,90]	2	1	86.39	50	[1.26;98.74]