Behavioural Biases in the European Football Betting Market

Master Thesis Behavioural Economics

Stefan van der Lee – 333124
Supervisor: dr. T.L.P.R. Peeters

02-07-2015

This paper investigates the presence of the favourite-longshot bias and the sentiment bias in the European football betting market. The presence of these biases is investigated by comparing the returns of different betting strategies. This paper shows that both the favourite-longshot bias and the sentiment bias are present in the match-odds of European football matches.

ERASMUS UNIVERSITEIT ROTTERDAM

Table of Contents

1. Introduction

A growing body of literature examines the efficiency of sports betting markets and investigates the presence of biases. Thaler and Ziemba (1988) point out that there are two definitions of market efficiency. The weak market efficiency condition states that no bets should have positive expected values. The strong market efficiency condition states that all bets should have equal (non-positive) expected values. This expected value would then be equal to the bookmaker's take.

The first papers that examined the efficiency of betting markets focussed mostly on horse racing. Some of them (e.g. Griffith, 1949; Ali, 1977) found an anomaly called 'the favourite-longshot bias', while others did not find this anomaly. The favourite-longshot bias means that favourites win more often than the subjective market probabilities imply and longshots (underdogs) win less often (Cain, Law & Peel, 2000). When this is the case it is more profitable to bet on the favourite than on the longshot. There are some possible explanations for the presence of this bias, which will be discussed in this paper. However, clearly, such a finding violates at least the strong market efficiency condition of a market as given by Thaler and Ziemba (1988) and violates the weak market condition when this bias leads to bets with positive expected values.

When betting on other sports became more popular, researchers also examined the efficiency of other sports betting markets than the horse race betting market. They investigated whether the favourite-longshot bias was also present in these markets. Woodland and Woodland (1994) investigated this bias in the Major League Baseball betting market and they found the reversal of the bias, called the reverse favourite-longshot bias. This means that a simple strategy of only betting on underdogs would lead to smaller losses than those implied by market efficiency.
Cain, Law & Peel (2000) did the same for the UK football betting market and they found evidence for the presence of the favourite-longshot bias, similar to the findings in horse racing.

Another bias that is repeatedly found in the sports betting market is the sentiment bias. In the football betting market, the sentiment bias means that the odds are influenced by the popularity of a football team. Forrest and Simmons (2008) found that, in Spanish and Scottish football, supporters of the more popular team were offered more favourable odds. This means that a betting strategy of betting on teams with the most supporters leads to a higher return than a betting strategy of betting on the less popular teams. Feddersen, Humphreys and Soebbing (2013) found a similar result for NBA matches.

The aims of this paper are to investigate the presence of these two behavioural biases in the odds of the European football betting market and to see whether the weak market efficiency condition given by Thaler and Ziemba (1988) is violated. In other words, the research questions of this paper are: Are the favourite-longshot bias and the sentiment bias present in the European football betting market? And does this lead to profitable betting strategies? The research questions will be answered by comparing returns of different betting strategies over a large dataset of European football matches. The data consists of ten seasons from five large football competitions. As measures of favouritism (probability to win a match) the odds for the particular match and the total amount of wages paid out in a year by the teams in the match are used. To determine the popularity of a club two measures are used as well. First, the average home attendance over one season is used. Second, the number of Twitter followers per club is used. The latter is, to the best of my knowledge, not done before.

The main findings of this paper are that the favourite-longshot bias and the sentiment bias are both present in the odds (independent of the measure that is chosen for favouritism and popularity) and in this dataset they seem to lead to betting strategies with a positive return. However, the opportunities to exploit these profitable betting strategies are scarce.

The odds that are used in this paper are an average of fixed odds given by a large amount (from 6 to over 70) of bookmakers. Odds tell the bettor how much he or she will get paid when a certain event happens. Fixed odds means that the odds do not change between the publication of the odds and the match. This means that the bettor knows what the odds are at the moment he or she places the bet. This is in contrast with pari-mutuel betting, which is, in particular, used in horse racing. In pari-mutuel betting the winning odds are given by the total amount that is bet on the winner divided by the sum of all bets minus track take. This means that in a pari-mutuel betting market the bettors do not know the exact odds beforehand.

The rest of this paper is organized as follows. Section 2 gives an overview of the existing literature on the two biases and presents my expectations of this research, based on the existing literature. Section 3 provides behavioural explanations for the two biases. Section 4 describes the data. Section 5 explains the methods that are used in this paper. Section 6 presents and discusses the results. Section 7 presents the main conclusions of this paper and contains a discussion on the results and methods of this paper.

2. Literature Review

This section discusses some of the most important and relevant papers including research to the favourite-longshot bias and the sentiment bias. The first part summarises some papers that examined betting market efficiency and searched for or found the favourite-longshot bias or the reverse favourite-longshot bias. The second part discusses some papers that found the sentiment bias. The third part explains why in some betting markets the favourite-longshot bias is found and in other markets the reversal of this bias is found. The fourth and last part of this section presents my expectations, based on the existing literature, for this study.

2.1 Evidence for the favourite-longshot bias in sports betting markets

The first paper in which the favourite-longshot bias was found was the paper of the psychologist Griffith (1949). He was interested in the sports betting behaviour of people. The horse race betting market was a pari-mutuel betting market, which means that the odds on the horses are a function of the proportion of the total money that is bet on each horse and are therefore socially determined. This gave him the opportunity to see how well the people were able to predict the winners. He was the first who compared the probabilities extracted from the odds with the percentage of winners. In other words, he compared the psychological odds with the a-posteriori probabilities. He did this for horse races in the UK and found that the odds were good predictors of the winning probabilities. The only deviation he found was in the extremes. Too little was bet on the short-odded horses (the favourites) and too much was bet on the long-odded horses (the longshots). This was the first evidence of, what was later labelled, the favourite-longshot bias.

William McGlothlin (1956) studied the betting behaviour of people at a horse race track in the United States. McGlothlin did the same as Griffith (1949). He divided the horse races into groups according to the odds and compared the objective probabilities with the subjective established odds. The first purpose of his study was to see in what way the relationship between the odds and the probabilities changed over a day. McGlothlin was interested whether people changed behaviour during a betting day. He found some results for that, but for this study the most interesting finding was that in his dataset the same bias was present as in Griffith's. McGlothlin states that: 'It appears that subjects can be expected to accept low expected values when low probability-high prize combinations are involved, while demanding higher expected values in the case of high probability-low prize combinations.' This means that people bet, relatively seen, too much on the longshots and this makes it more profitable to bet on the favourites. This is again evidence of the favourite-longshot bias.

Weitzman (1965) defined an average bettor at the horse race track. According to the findings of Weitzman, the average man at the race track does not behave according to the expected value hypothesis. He bets too much on the longshot and too little on the favourite. This means that also with the approach that is chosen in this paper, evidence for the favourite-longshot bias was found.

In Ali (1977), Mukhtar M. Ali criticized the findings of Griffith (1949) and McGlothlin (1956). He states that both studies have serious estimation errors in the estimated probabilities. In his paper he analyzed the public betting behaviour in harness horse races. The amount of races is larger than the amount of races Griffith (1949) and McGlothlin (1956) used in their studies. He obtained subjective and estimated objective winning probabilities for all these races. He ordered the horses in each race from most favourite to less favourite according to the odds. He finds that for the horses with the largest probability to win, the objective probability to win is higher than the subjective probability to win. For the longshots, the subjective probability is significantly higher than the objective probability. Although he first criticized their findings, his results are in line with Griffith (1949) and McGlothlin (1956) and point in the direction of the favourite-longshot bias.

Woodland and Woodland (1994) tested the efficiency of the baseball betting market and examined the presence of the favourite-longshot bias in this market. One difference with the studies above is the type of betting. Horse races use pari-mutuel betting where the winning odds are given by the total amount that is bet on the winner divided by the sum of all bets minus track take. This means that the bettors do not know the exact odds beforehand. In the Major League Baseball betting market this is different. Here fixed odds are used. At the moment the bet is placed the bettor knows the odds. The paper tests weak form market efficiency. When the market is efficient each betting strategy should lead to the same loss, equal to the take of the bookmaker. The first finding is that there are no significant differences between the objective and the subjective win probabilities. This would lead to the conclusion that the baseball betting market is efficient. However, another finding is that bettors overbet the favourites, rather than the underdogs. This means that a simple strategy of only betting on underdogs would lead to smaller losses than those implied by market efficiency. This is a reversal of the favourite-longshot bias found in horse racing and is called the reverse favourite-longshot bias. In this dataset the bias is too small to create a profitable betting strategy, but it leads to the conclusion that the baseball betting market is not fully efficient.

The paper of Cain, Law and Peel (2000) is the first paper discussed here that focused on the football betting market. Specifically, they focused on the fixed odds betting market on UK football. Their dataset contains one season of Football league matches. Fixed odds means that once the bookmaker declares the odds (several days before the match) the odds cannot change anymore. In their paper, match outcomes (win by the home team, win by the away team or a draw) and match scores (the exact score of the match, e.g. 1-3) are examined. For the match outcomes they found evidence for the favourite-longshot bias. Bets on longshots led to lower returns than bets on favourites. In their sample this did not lead to profitable betting strategies. In the match scores they find the favourite-longshot bias as well. In some cases this can lead to profitable betting strategies, but these are scarce.

In Woodland and Woodland (2001) the market efficiency of and the presence of the favourite-longshot bias in the betting market of the National Hockey League are examined. The data consists of matches of 6 seasons of the NHL. Their conclusion was that the NHL betting market was (somewhat) inefficient and that the reverse favourite-longshot bias was present. This means that the strategy of betting on underdogs leads to higher average returns than betting on favourites. In addition, the strategy of betting on heavy underdogs that play away is a profitable strategy in the last four seasons of their dataset.

Woodland and Woodland were interested whether the inefficiency and the profitable betting strategy would still be present in the National Hockey League some years after the publication of Woodland and Woodland (2001). In Woodland and Woodland (2010) they find that this reverse favourite-longshot bias has disappeared and that the NFL betting market was efficient at that time. In the baseball betting market the reverse favourite-longshot bias is still present.

2.2 Evidence for the sentiment bias in sports betting markets

A growing body of evidence supports the idea that the sentiment bias is, next to other financial markets, also present in the sports betting market. In the football betting market, the sentiment bias means that the odds are influenced by the popularity of a football team. When this is the case, the odds are not a perfect reflection of the relative probabilities of match outcomes, but are affected by the number of fans of each team.
In a pari-mutuel betting markets the direction of the sentiment bias is clear. More bets will be placed on the popular team. This leads to lower odds for the more popular team and higher odds for the less popular team. When this is the case it makes it more profitable to bet on the non-favourite team (the longshot). Off course, well-informed bettors will recognize this, but when the bookmaker's take is 'sufficiently' high the informed bettors cannot make a profit out of this bias and will not recover the balance in the odds (Forrest and Simmons, 2008).

In a fixed odds betting market the direction of the sentiment bias is less clear. Although Avery and Chevalier (1999) found that in the National Football League betting market, losses for 'glamorous' teams were abnormally high, most studies have found that the losses are smaller for the more popular teams.

Forrest and Simmons (2008) examined the sentiment bias in Spanish and Scottish Football. They used the differences in home attendance of last season between the two teams as popularity-measure. They found that supporters of the more popular team were offered more favourable odds. This leads to higher returns when you would 'follow the crowd'. In their dataset this bias is not large enough to overcome the bookmaker's take and make systematically profits by exploiting this.

Braun and Kvasnicka (2013) studied national sentiment. By analyzing odds of online bookmakers all over Europe they found that the odds were biased for the national team from the country where the bookmaker is settled. This finding is based on online bookmaker odds from 12 European countries for matches of the qualification of UEFA Euro 2008.

In the paper of Feddersen, Humphreys and Soebbing (2013) the sentiment bias is found as well. They examined National Basketball Association (NBA) matches. Their dataset consists of more than 33.000 matches over a period of 30 years. They found that, in line with Forrest and Simmons (2008), bookmakers offer more favourable point spreads for more popular teams. Their measures of popularity are relative match attendance and team All Star votes. Next to the observation that the bias is present, they explain why bookmakers offer more favourable odds to more popular teams. Feddersen, Humphreys and Soebbing (2013) state that the reason why bookmakers offer more favourable odds to more popular teams is to increase the profits. It may lead to unbalanced betting, but it will (probably) lead to a higher betting volume as well. This higher betting volume leads to higher profits because of the take that is earned by bookmakers. Humphreys (2010) showed that unbalanced betting leads to higher profits than balanced betting on either side of games. Besides (when not shifting the odds too much) it does not lead to a profitable betting strategy for informed bettors.

2.3 Explanations for the mixed findings in the literature

The first two sections of this literature review showed that there is a large amount of evidence on the presence of the favourite-longshot bias and the sentiment bias. However, not in all cases the direction of the bias is the same. Woodland and Woodland (1994) found the reverse favourite-longshot bias in the baseball betting market, where Cain, Law and Peel (2000) found the favourite-longshot bias itself for the football betting market. In the case of the reverse favourite-longshot bias it is more profitable to bet on longshots, while in the case of the favourite-longshot bias betting on favourites leads to higher returns than betting on longshots.

For the sentiment bias mixed results are found as well. Although almost all other studies found that the odds for popular teams are more favourable than the odds for less popular teams, Avery and Chevalier (1999) found the opposite.

In some studies explanations for these mixed results are given. Williams and Paton (1998) argued that possible explanations of the favourite-longshot bias are the presence of relatively high transaction costs (bookmaker's take) and the presence of uninformed or noise bettors in the market. In contrast, low transactions costs and the presence of informed bettors that derive extra utility from betting on favourites can possibly lead to the reverse favourite-longshot bias. This is how they explain the finding of the reverse favourite-longshot bias in the US Baseball League by Woodland and Woodland (1994), as the Baseball betting market is characterized by low transactions costs. They say as well that, in a fixed odds market, the presence of insiders can lead to a positive favourite-longshot bias. So, according to Williams and Paton, it depends on the level of transaction costs and the presence of informed bettors or insiders whether the odds reflect a favourite-longshot bias, a reverse favourite-longshot bias or no bias at all.

Ottaviani and Sorensen (2010) created a theory that predicts that the favourite-longshot bias is larger when the number of (informed) bettors increases, bettors have more private information, the recreational value of the event decreases or the bookmaker's take increases. This is in line with the findings of Williams and Paton (1998). Ottaviani and Sorensen (2010) argue that they can explain most of the findings in the literature with their theory. They believe as well that one of the explanations of the presence of the reverse favourite-longshot bias is low transaction costs. According to them this can explain why the bias is still present in baseball and is not present in hockey, as the transaction costs in baseball are half of the transactions costs in hockey.

2.4 Hypotheses based on the existing literature

Based on the existing literature, my expectations for this research are that both the favourite-longshot bias and the sentiment bias will be present in the European football betting market. In line with the findings of Cain, Law and Peel (2000), who examined the UK football betting market, I expect that it will be more profitable (higher returns will be achieved) to bet on favourites than on longshots. In terms of the odds this would mean that the odds for favourite teams (teams with the highest probability to win) are more favourable than the odds for longshots. Regarding the sentiment bias I expect that the odds will be more favourable for the more popular teams. This would be in line with almost all (except for Avery and Chevalier, 1999) literature that is discussed in this paper. When this expectation is true the betting strategy of betting on the most popular teams should result in higher returns than the betting strategy of betting on the less popular teams. This would be in line with the findings of Forrest and Simmons (2008) and Feddersen, Humphreys and Soebbing (2013) and could be explained by profit maximizing behaviour of the bookmakers. Next to this, I do not expect that there will be large opportunities to obtain systematic profits by using simple betting strategies that make use of these two biases. I expect that the biases will be present, but that they are not large enough that informed betters can make sure profits.

3. Behavioural explanations for the biases

The literature review showed that there is a large amount of evidence for the favourite-longshot bias and the sentiment bias. An interesting question is: Why are these biases present? What behaviour leads to the existence of these biases? This section provides some possible explanations.

The favourite-longshot bias means that people overbet longshots and underbet favourites. Because of this behaviour the odds are adjusted and the bias is present in the odds as well. It becomes more profitable to bet on the favourites than on the longshots. Why do people overbet the longshot? How can this behaviour be explained?

One explanation for the favourite-longshot bias is that people have the tendency to overestimate small probabilities. This is the explanation Griffith (1946) gave for the favourite-longshot bias he found in the horse race betting market. When people overestimate small probabilities they overestimate the winning probabilities of longshots. They think the probability that the longshot will win is larger than the real probability. This leads to too much bets on the longshots, which will result in lower odds for the longshots and higher odds for the favourites.

Close to the explanation that people overestimate small probabilities is the explanation that people give higher weights to small probabilities. Subjective probabilities (estimated probabilities) are used in the rank-dependent utility model (Quiggin, 1982) and in cumulative prospect theory (Tversky and Kahneman, 1992). These two models make use of subjective probabilities instead of objective probabilities. Tversky and Kahneman (1992) showed that most people give higher weights to small probabilities and lower weights to large probabilities. According to them the most common probability weighting function looks like the function in figure 1. The red line shows the weight people give to a certain probability and the dotted line shows the objective probability. It can be seen that for small probabilities the red line is above the dotted line. This means that for small probabilities the probability weight is larger than the objective probability. For large probabilities it is the other way around. There, the probability weight is smaller than the objective probability. When people have a probability weighting function like this, it means that they do not overestimate small probabilities, but they give them more weight. This overweighting of small probabilities leads to the same betting behaviour as the overestimation of small probabilities. People give higher weights to the probability that the longshot will win and give lower weight to the probability that the favourite will win. Because of this, people overbet the longshots and underbet the favourites. This will result in lower odds for the longshots and higher odds for the favourites.

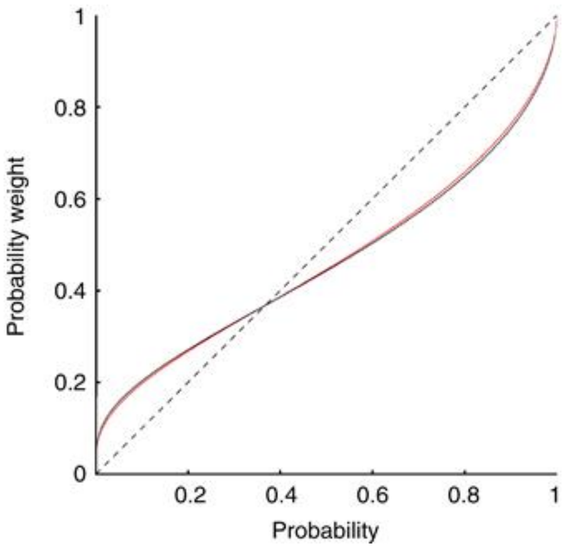

Figure 1: The most common probability weighting function

Weitzman (1965) argued that the favourite-longshot bias he found could be explained by the fact that bettors are risk lovers. According to him they are able to estimate the (small) probabilities, but they just love the risk of betting on longshots. Ali (1977) gave the same explanation in his paper. When bettors love the risk of betting on longshots, the odds for longshots will go down and the odds for favourites will go up.

Most researchers assume that the behaviour of the bettors creates these biases. Dixon and Pope (2004) argued that it is possible that the odd-setter (the bookmaker) displays the cognitive bias as well. When the odd-setter suffers from the favourite-longshot bias he will assign too low odds to longshots and too high odds to favourites. In this case the inefficiency of the odds is not a reaction on the behaviour of the bettors, but the bias is with the odd-setter.

The sentiment bias means that people bet too much on the team they support. This leads to the fact that more bets are placed on popular teams. Why do people bet too much the team they support?

An important explanation is wishful thinking. As Babad and Katz (1991) showed, self-reported intensity of fanhood or preference for a given team influences predictions of game outcomes. They found that even a moderate preference (e.g. supporting another team, but having only a slightly preference for one team over another) leads to wishful thinking in predictions of game outcomes. People that support or favour a certain team estimate the probability that this team will win higher than other people. This explains why people bet more on the team they support. They estimate the probability that their team will win larger than other people. Babad and Katz (1991) label this phenomenon as wishful thinking, but you may argue that you can label this as overconfidence as well. In general, the overconfidence bias means that people overestimate their own capabilities. In this case, people are overconfident about the capabilities of the team they support. They think they are better (have a larger probability to win the match) than they actually are.

Whether wishful thinking or overconfidence bias, this behaviour leads to too many bets on the popular teams and too little bets on the less popular teams and this will influence the odds.

4. Data

The dataset used in this study consists of matches out of five large football competitions in Europe over the seasons 2001-2002 up to and included 2010-2011. The competitions used are the Premier League (England), the Championship (England), The Primera División (Spain), The Serie A (Italy) and the Ligue 1 (France). In total, the dataset consists of 20289 matches.

For all these matches the results and the match odds are present. For a large part of this dataset (almost 85 percent) the total amount of wages per team per year is known and for an even larger part (almost 98 percent) the amount of Twitter followers (which is used as proxy for team popularity) is present and used. There are some clubs in the dataset that do not have an official twitter account. These are Venezia, Piacenza, Ancona, Grenoble and Arles. When analysing the sentiment bias with the number of Twitter followers as measure of popularity the matches that include one of those teams are deleted from the dataset. As a second proxy for team popularity the average home attendance for every team per season is included. This is known for all clubs in the dataset. The odds that are used are averages of a large number (from 6 to over 70) of bookmaker odds.

The odds that are used are fixed decimal odds. Odds tell the bettor how much money he or she will get when a certain event happens. Events that will happen with a large probability (according to the bookmaker) have small odds and events that will happen with a small probability have large odds. Odds can be seen as inverse implied probabilities. Fixed odds means that the odds do not change between the publication of the odds (several days before the match) and the match. This means that the bettor knows what the odds are at the moment he or she places the bet.

Decimal odds is a way of displaying the odds. While in the United Kingdom fractional odds are most popular to use by bookmakers, the use of decimal odds is favoured in the rest of Europe. The decimal odds tell the bettor how much he or she will get for every euro that he or she bets. When the odd for a home win is 1.42 this means that for every euro that is bet you will receive 1.42 euro when the home team wins. When 10 euro is bet and the home team wins, the bettor ends with 14,20 euro's and makes a profit of 4.20 euro's.

The match results and the odds come from www.football-data.co.uk , the numbers of twitter followers come from www.folos.im and the average home attendances come from www.european-football-statistics.co.uk.

Table 1, on the next page, shows summary statistics for the measures that are used in this paper. The measures with their statistics are split up per competition. For the odds the table shows the average home odds, the average odds for draws and the average away odds. For the other measures, the table shows the average, the standard deviation and the minimum and maximum per competition. The wages for the Primera División, the Serie A and the Ligue 1 are in million euro's, the wages for the Premier League and the Championship are in million pounds and the attendances and Twitter followers are in thousands of people.

	Odds			Wages (x1.000.000)				Twitter Followers (x1000)				Attendances (x1000)				Years	
	Avg. Home	**Avg. Draw**	**Avg. Away**	**Average**	**St. dev**	**Min.**	**Max.**	**Average**	**St. dev**	**Min.**	**Max.**	**Average**	**St. dev**	**Min.**	**Max.**	**Min.**	**Max.**
Premier League	2.50	3.58	4.39	48.5	30.8	11.4	172.6	1374.8	1992.7	51.4	5782.2	34.8	12.9	15.8	75.8	2001	2011
Championship	2.15	3.30	3.55	10.8	6.0	2.7	47.4	130.5	180.3	12.0	2458.4	17.0	6.6	2.8	43.4	2001	2011
Primera División	2.30	3.50	4.23	39.5	45.2	4.9	240.6	1779.5	4588.4	15.1	15926.2	28.4	17.3	8.0	79.3	2001	2011
Serie A	2.39	3.31	4.53	49.3	50.5	0.4	234.0	385.3	660.0	0.9	2546.8	24.0	15.2	5.1	63.6	2001	2011
Ligue 1	2.17	3.16	4.04	31.3	20.2	7.2	111.7	359.3	541.5	0.4	2207.2	20.9	11.6	3.0	53.0	2001	2011

Table 1: Summary statistics of the data

5. Method

This section explains how the research questions are answered and why some methodological choices are made. Important to note is that the term ‘favourite’ is used for the team with the largest probability to win and the term ‘popular’ is used for a team with a large amount of fans. This choice of terms is consistent over the whole paper.

The presence of the favourite-longshot bias and the sentiment bias will be investigated by comparing returns of different betting strategies. For every match it is checked which team is the favourite and which team is the more popular team. To measure favouritism (probability to win) and popularity different proxies are used. The choice of these different proxies is explained in the next two paragraphs. By constructing betting strategies that make use of the measures for favouritism and popularity, it can be tested whether the favourite-longshot bias and the sentiment bias are present in the odds.

For the favourite-longshot bias it is necessary to state which team is the favourite. One possible way to do this is by looking at the odds. The team with the lowest odd is the favourite to win this particular match, according to the bookmaker. Although some studies showed that the odds are efficient in forecasting match results, it might be a bit arbitrary to use only the odds as a measure of favouritism when we test whether the same odds include biases. Therefore another measure to indicate the favourite is used as well: The total amount of wages paid out by a club in a year. Underlying here is the assumption that clubs who can pay higher wages do have the better players and are the better teams. According to Peeters and Szymanski (2014) this is a reasonable assumption.

For the sentiment bias a measure of popularity is needed. The first measure that is used is the number of Twitter followers per team. One assumption is made here, as the number of twitter followers in May 2015 is used. This means that it is assumed that popular teams now where also popular from the seasons 2001-2002 until 2010-2011. For most of the teams this will be the case, but I am aware of the fact that for some teams the popularity has changed over the years. This is the reason why I make use of a second measure for popularity as well. As a second measure for team popularity the average home attendance (the average number of people that were in the stadium during a home match) of a club in a season is used. This proxy for popularity is used by Forrest and Simmons (2008) as well.

There are some clubs (and years) in the dataset for which the wages are unknown and some clubs that do not have an official Twitter account. When the favourite-longshot bias is investigated with the wages as measure for favouritism the matches that involve these teams are excluded from the analysis. The same holds for the teams that do not have an official Twitter account in analysing the sentiment bias.

After indicating which teams are the favourites (by one of the two measures), three betting strategies are possible to investigate the presence of the favourite-longshot bias: Betting on the favourites, betting on the longshots and betting on draws. These three betting strategies are examined and their returns are compared. Something similar is done for the sentiment bias. After indicating which teams are the most popular teams (by one of the two measures), three betting strategies are possible: Betting on the most popular teams, betting on the least popular teams and betting on draws. The returns of these betting strategies are compared. A betting strategy is constructed as follows: For every match 1 imaginary euro is bet on a match outcome that follows the chosen betting strategy. Then the match results are examined and for every match the return of this betting strategy is calculated. At the end the (average) return for this particular betting strategy is calculated and is reported as the return of the betting strategy over the (chosen part of the) dataset.

To investigate the presence of the two biases, the returns of the betting strategies are compared for the whole set of matches, but as well for parts of the dataset to get a clearer view on the betting strategies and the biases. In this way it is investigated whether the returns of the betting strategies are different when the degree of favouritism or popularity is 'very large' or 'smaller'. In the case of the odds these distinctions are made by dividing the matches according to the odds given for the favourite team. Steps of 0.10 are made between two categories. For the wages, the Twitter followers and the average home attendance something similar is done. Here, the divisions are made by using the standard deviations of the different measures. In case of the wages, you get matches where the difference in wages between two opponents is smaller than one standard deviation, between one and two standard deviations, between two and three standard deviations, etc. etc. For the other proxies the same principle is used.

Because in a lot of cases the more popular teams will be the best (and therefore favourite) teams, a 'two by two design' is included. In this two by two design, four different betting strategies that use both the odds and the number of twitter followers are compared. The four betting strategies in this two by two design are: Betting on the teams that are the favourite according to the odds *and* have more Twitter followers than the opponent (1), betting on the teams that are the favourite according to the odds *and* have less Twitter followers than the opponent (2), betting on teams that are not the favourite according to the odds *and* have more Twitter followers than the opponent (3), betting on teams that are not the favourite according to the odds *and* have less Twitter followers than the opponent (4).

This approach makes sure that when looking at the sentiment bias it is really the sentiment bias we are looking at and not the favourite-longshot bias. For the favourite-longshot bias the same holds. In other words, this approach makes it possible to control for the other bias.

For all the analyses of betting strategies that are described in this section, the mean returns of the strategies will be compared. The comparison is only informative when it can be checked whether differences that are found are significantly different as well. To do this the variance (or standard deviation) of the returns is necessary. As the distribution of the returns is unknown, the bootstrapping method is used for creating the standard deviations. I have chosen to use 10.000 repetitions for the bootstrapping method, because this is large enough and the calculation time is still acceptable. With the standard deviations from the bootstrapping method, 95 percent confidence intervals are constructed. By examining the 95 percent confidence intervals of the returns of the different betting strategies it can be checked whether the returns of the betting strategies are significantly different at a 5 percent significance level.

6. Results

This section presents and discusses the results and will answer the research questions of this study. The results are split up in three parts. The first part presents the results on the favourite-longshot bias only, the second part presents the results on the sentiment bias only and the third part presents the results on the combination of the two biases: the two by two design.

6.1 The favourite-longshot bias

The first results that I discuss are on the favourite-longshot bias where the different betting strategies are performed over the whole dataset. When the odds are used as the measure of favouritism, this is possible for 20289 matches. The returns of three betting strategies are compared. The results can be seen in table 2. The betting strategy of betting on favourite teams leads to a return of -5,5 percent, the strategy of betting on only draws leads to an average return of -10,7 percent and the betting strategy of betting on all the longshots leads to a return of -16,7 percent. Examining the 95 percent confidence intervals it can be seen that they do not overlap. Therefore it can be concluded that the returns of the betting strategies that are found are significantly different at a 5 percent significance level.

Odds-range favourites	# of games	ROI favourites	ROI draw	ROI longshots
1.04-2.72	20289	-5,5% [-6,84%, -4,16%]	-10,7% [-12,75%, -8,67%]	-16,7% [-19,03%, -14,35%]

Table 2: Favourite-longshot bias with the odds as the measure of favouritism.

When the wages paid out by both clubs is used as the measure of favouritism 17.237 matches are left, because for the other 3052 matches the wage-information of at least one of the two teams in a match is missing. This means that 84,95 percent of the initial dataset is left for this analysis. The returns of the strategies of betting on the teams that pay the highest wages, betting on the teams that pay the lowest wages and betting on draws can be compared. The results are similar to those when the odds are used as a measure of favouritism and are showed in table 3. Again, betting on the favourites leads to a higher return (-7,1 percent) than betting on the longshots (-15,8 percent) and again the confidence intervals do not overlap. Therefore it can be concluded that the different returns of the betting strategies that are found are significantly different at a 5 percent significance level.

# of games	ROI Highest Wages	ROI draw	ROI Lowest Wages
17273	-7,05% [-8,76%, -5,41%]	-10,27% [-12,47%, -8,15%]	-15,82% [-18,12%, -13,35%]

Table 3: Favourite-longshot bias with the wages as the measure of favouritism.

The same type of analysis has been done for different categories of matches. The matches are split up based on 'degree of favouritism'. The matches are divided according to the odds given for the favourite team. Steps of 0.10 are made between two categories. The results are shown in table 4. The categories '1.00-1.10' and '1.10-1.20' are taken together, because there are not so many matches in these categories. The same holds for the categories '2.60-2.70' and '2.70-2.72'. The smallest odd in the dataset is 1.04 and the largest for a favourite is 2.72. Therefore these values are in table 4 and not 1.00 and 2.80.

Looking at the results in the table, it can be seen that for 12 of the 16 categories the betting strategy of betting on the favourite teams leads to higher returns than the other two betting strategies. More specifically, this is the case for the first 9 categories (for matches where the odds for the favourite are below 2.00). From category 10, this betting strategy leads 'only' in 3 out of 7 categories to the highest returns. It seems that the favourite-longshot bias is strongest (or at least present) when there is a clear favourite (when the odd for one of the teams is low). For 14 of the 16 categories it is more profitable to bet on the favourite teams than on the longshots. Only for the matches where the favourite's odds are between 2.00 and 2.10 and between 2.50 and 2.60 betting on the longshots leads to higher returns than betting on the favourites. Looking at the 95 percent confidence intervals it can be seen that quite a lot of the confidence intervals overlap. The confidence intervals are larger than the confidence interval for the returns over the whole dataset, because there are less matches in each category now. Although some confidence intervals overlap, there seems to be a clear pattern in the returns.

Probably, the most interesting result is the return of the betting strategy of betting on the favourites for the matches with the clearest favourite (odds below 1.20). Over these 247 matches this strategy results in a return of +4,7 percent with a 95 percent confidence interval of [-0.02, +8.45] percent. This means that in this dataset the favourite-longshot bias is largest in the matches where there is one team with a very high probability to win the match and one team with a very low probability to win the match. The probability weighting function showed in figure 1 can explain this result. People give higher weight to the very small probabilities and lower weight to the very high probabilities and therefore the bias is largest in this kind of matches. For all the other categories there are no positive returns for any strategy.

Odds-range favourites	# of games	ROI favourites	ROI draw	ROI longshots
1.04-1.20	247	+ 4,7% [-0,02%, 8,45%]	-54,5% [-71,71%, -30,56%]	-67,6% [-89,13%, -21,18%]
1.20-1.30	621	-2,4% [-6,57%, 1,48%]	-16,9% [-30,63%, -1,11%]	-45,2% [-61,51%, -24,44%]
1.30-1.40	749	-3,1% [-7,72%, 1,15%]	-20,1% [-31,39%, -7,51%]	-23,1% [-39,26%, -4,13%]
1.40-1.50	984	-5,9% [-10,38%, -1,73%]	-10,8% [-20,88%, -0,53%]	-19,5% [-32,58%, -4,48%]
1.50-1.60	1284	-4,6% [-8,73%, -0,40%]	-13,5% [-21,73%, -5,00%]	-17,3% [-28,33%, -5,71%]
1.60-1.70	1587	-6,7% [-10,70%, -2,69%]	-9,5% [-16,81%, -1,98%]	-16,3% [-25,42%, -6,99%]
1.70-1.80	1793	-3,8% [-7,80%, 0,31%]	-12,3% [-18,96%, -5,40%]	-18,8% [-26,58%, -10,40%]
1.80-1.90	1791	-6,9% [-11,25%, -2,61%]	-11,1% [-17,77%, -4,45%]	-12,6% [-20,28%, -4,66%]
1.90-2.00	1626	-2,8% [-7,51%, 1,96%]	-14,4% [-21,28%, -7,51%]	-14,3% [-21,58%, -6,27%]
2.00-2.10	1712	-9,7% [-14,31%, -4,80%]	-8,8% [-15,45%, -2,03%]	-7,7% [-14,80%, -0,29%]
2.10-2.20	1723	-3,9% [-9,12%, 1,07%]	-8,2% [-14,74%, -1,30%]	-18,0% [-24,44%, -11,09%]
2.20-2.30	1818	-8,2% [-13,29%, -3,18%]	-5,1% [-11,48%, 1,62%]	-14,7% [-20,88%, -8,56%]
2.30-2.40	1776	-5,7% [-10,80%, -0,12%]	-7,7% [-14,18%, -1,14%]	-14,1% [-20,09%, -7,87%]
2.40-2.50	1557	-5,3% [-11,10%, 0,74%]	-2,6% [-9,86%, 4,66%]	-18,9% [-24,94%, -12,65%]
2.50-2.60	792	-8,7% [-17,04%, -0,06%]	-14,6% [-23,90%, -4,58%]	-3,0% [-11,82%, 6,26%]
2.60-2.72	224	-0,03% [-16,20%, 17,07%]	-10,13% [-27,37%, 8,71%]	-13,46% [-29,31%, 3,27%]

Table 4: Favourite-longshot bias with the odds as the measure of favouritism.

In figure 2 the results from table 4 are plotted in a graph. It can be seen clearly that the returns are highest for the strategy of betting on favourites for almost all odds-ranges. Next to that, the graph shows that the returns of the strategy of betting on the favourite is largest for favourite teams with very low odds (left side of the graph) and the returns of the other two strategy are lowest in this same part of the graph. The return of the strategy of betting on longshots is lowest for matches where there is a really clear favourite according to the odds.

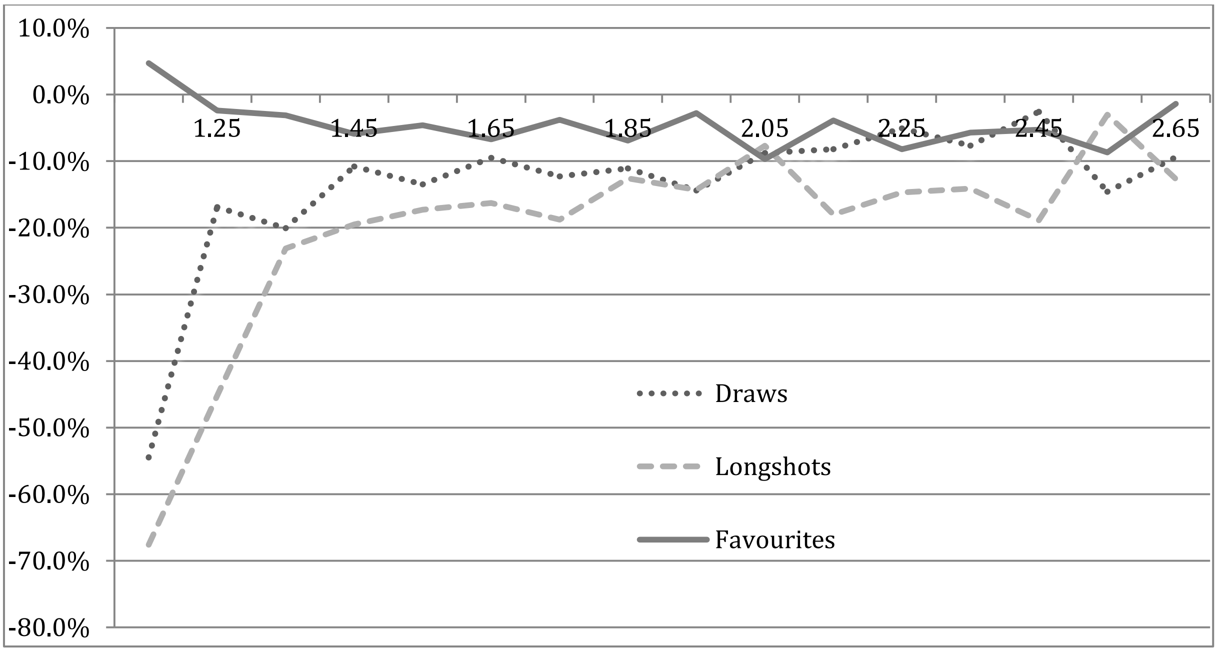

Figure 2: The favourite-longshot bias with the odds as the measure of favouritism.

The same type of analysis is done when the matches are split up according to the size of the differences in wages paid out by the clubs in a particular match. The divisions are made by using the standard deviation of the wages in that particular year and that particular competition. In this way you get matches where the difference in wages is smaller than one standard deviation, between one and two standard deviations, between two and three standard deviations and larger than three standard deviations. Table 5 and figure 3 show the results of this analysis.

	#Games	ROI Highest Wages	ROI Draws	ROI Lowest Wages
The difference in wages is smaller than 1 standard deviation	10923	-8,82% [-11,03%, -6,57%]	-9,37% [-12,04%, -6,63%]	-11,28% [-13,98%, -8,54%]
The difference in wages is between 1 and 2 st. deviations	3240	-6,88% [-10,39%, -3,32%]	-7,54% [-12,67%, -2,40%]	-17,44% [-23,06%, -11,39%]
The difference in wages is between 2 and 3 st. deviations	2017	-0,30% [-4,18%, 3,62%]	-14,80% [-21,41%, -8,07%]	-31,37% [-38,52%, -23,12%]
The difference in wages is larger than 3 standard deviations	1057	-2,12% [-6,56%, 2,51%]	-19,41% [-29,11%, -9,31%]	-28,02% [-39,90%, -13,87%]

Table 5: Favourite-longshot bias with the wages as the measure of favouritism.

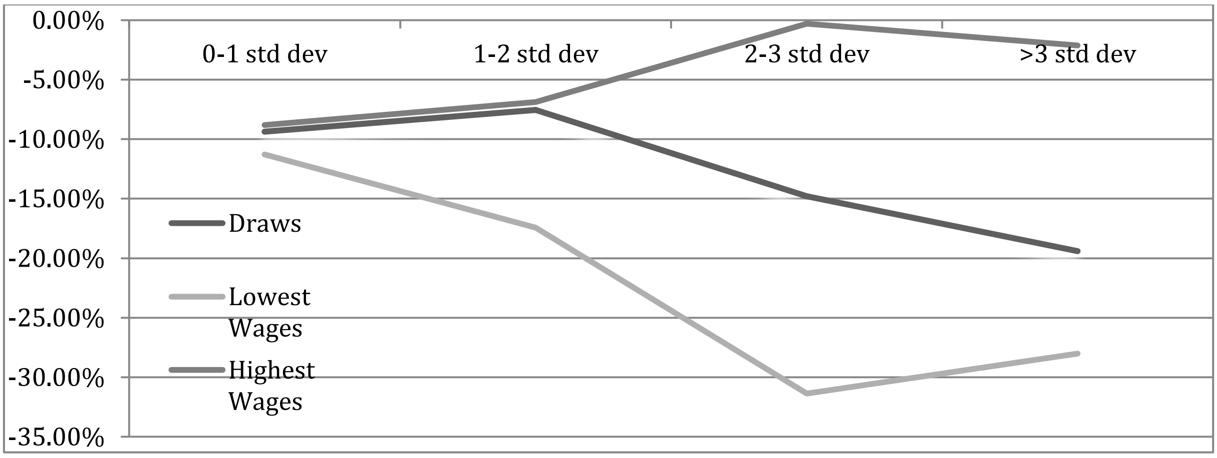

Figure 3: Favourite-longshot bias with the wages as the measure of favouritism.

The table and the figure show that for every sub-section of the dataset the return on the betting strategy of betting on the team that pays the most wages is higher than the return on the betting strategy of betting on the team that pays the least wages. For every sub-section the return on the betting strategy of betting on draws is in the middle. Except for the first sub-section, the returns are significantly different at the five percent significant level as well. Next to that, it can be seen that the differences in returns become larger when the differences in wages become larger. This can be seen most clear in figure 3. These results show that the betting strategy of betting on the team that pays the most wages leads to higher returns than the betting strategy of betting on the teams that pay the least wages and that the difference in returns grows with the size of the difference in wages paid out. This is consistent with the results from table 4 and figure 2. The clearer the favourite, the larger the bias.

6.2 The sentiment bias

First, this part discusses the results on the sentiment bias where the whole set of matches is used. One of the two measures for popularity is the average home attendance of a club in a season. This measure is available for the whole dataset of 20289 matches. Using this measure, the returns of three different betting strategies can be compared: Betting on the teams with the largest home attendance that season, betting on the teams with the smallest home attendance that season and betting on draws. Over the whole dataset, the betting strategy of betting on the team with the largest home attendance that season leads to a return of -7,3 percent. The betting strategy of betting on the team with the smallest home attendance leads to a return of -15,3 percent and the betting strategy of betting on draws leads to a return of -10,3 percent. Table 6 shows the returns of the different betting strategies with their 95% confidence intervals.

# of games	ROI Largest Attendance	ROI draw	ROI Smallest Attendance
20289	-7,32% [-8,92%, -5,70%]	-10,34% [-12,35%, -8,32%]	-15,29% [-17,39%, -13,11%]

Table 6: The sentiment bias with the average home attendance as the measure of popularity.

The same type of analysis is performed with the number of Twitter followers as measure for team popularity. When using this measure, 20042 matches are left to analyze. Again, the returns on the two (or three, including the strategy of only betting on draws) betting strategies can be compared. Over the whole dataset, the betting strategy of betting on all the teams with the most Twitter followers leads to a return of -6.9 percent. The betting strategy of betting on the team with the least Twitter followers leads to a return of -15.5 percent. Examining the 95 percent confidence intervals it can be seen that they do not overlap and so it can be concluded that the returns of the strategies are significantly different. The returns and the confidence intervals can be seen in table 7.

# of games	ROI Most Twitter followers	ROI draw	ROI Least Twitter followers
20042	-6,90% [-8,44%, -5,24%]	-10,30% [-12,31%, -8,16%]	-15,50% [-17,62%, -13,38%]

Table 7: The sentiment bias with the number of twitter followers as the measure of popularity.

The same type of analysis is performed when the matches are split up according to the differences in average home attendance. As for the analysis with the wages, the division is made by using the standard deviations of the average home attendance per year per competition. Table 8 and figure 4 show the results of this analysis.

	# Games	ROI Largest Attendance	ROI Draw	ROI Smallest Attendance
The difference in home attendance is smaller than 1 standard deviation	11014	-9,97% [-12,22%, -7,58%]	-8,43% [-11,16%, -5,74%]	-11,38% [-13,96%, -8,68%]
The difference in home attendance is between 1 and 2 standard deviations	5958	-4,24% [-6,99%, -1,42%]	-10,96% [-14,77%, -7,31%]	-18,12% [-22,10%, -14,10%]
The difference in home attendance is between 2 and 3 standard deviations	2395	-5,80% [-9,55%, -1,88%]	-10,66% [-16,74%, -4,41%]	-23,16% [-29,85%, -15,44%]
The difference in home attendance is larger than 3 standard deviations	922	0,42% [-4,70%, 5,62%]	-28,37% [-37,59%, -18,20%]	-23,20% [-36,13%, -7,53%]

Table 8: The sentiment bias with the average home attendance as the measure of popularity

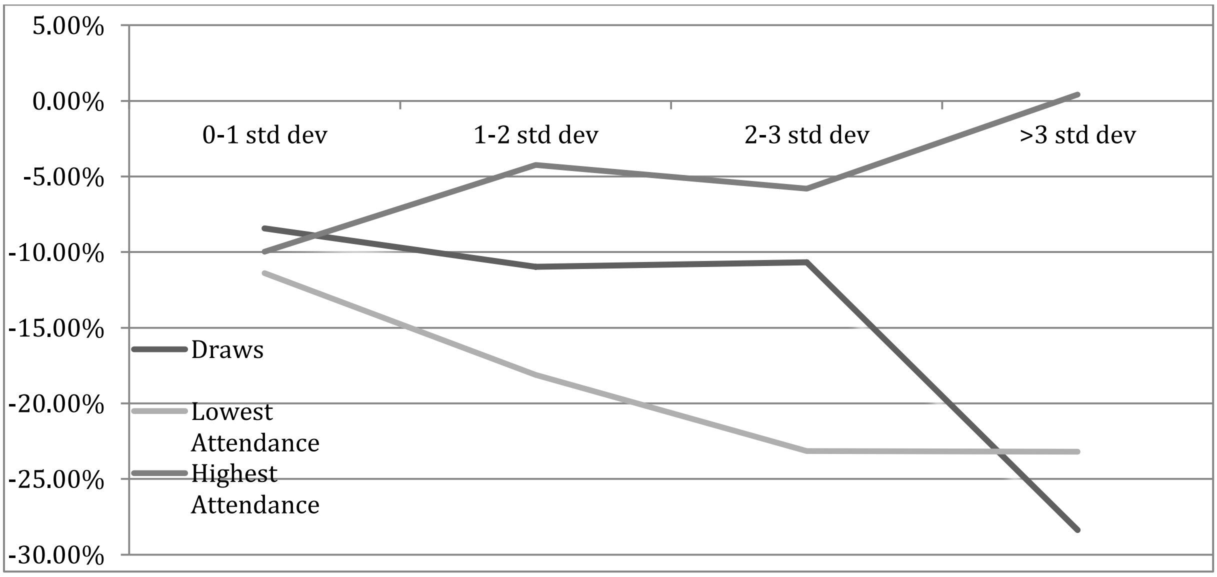

Figure 4: The sentiment bias with the average home attendance as the measure of popularity

The table and the figure show that for matches where the difference in average home attendance is smaller than 1 standard deviation the betting strategy of betting on draws leads to the highest return. Followed by the betting strategy of betting on the team with the largest average home attendance. When the differences in average home attendances become larger, the differences in returns between the three betting strategies become larger as well. When the difference in home attendance is larger than one standard deviation the betting strategy of betting on the team with the largest attendance leads to higher returns than the other two betting strategies. For matches where the difference in home attendance is larger than three standard deviations, this strategy leads to a positive return. The figure shows that in the most extreme cases, where the differences in home attendance is very large, the return of the betting strategy of betting on the most popular teams leads to the highest return.

Also for the number of Twitter followers as the measure for popularity, the analysis is performed when the matches are split up based on the degree of popularity. Again, the standard deviation of the number of twitter followers per competition is used to make the division. These results are shown in table 9 and figure 5.

	# of games	ROI Most TF	ROI Draw	ROI least TF
The difference in followers is smaller than 1 standard deviation	13985	-8,14% [-10,14%, -6,07%]	-9,16% [-11,56%, -6,76%]	-12,54% [-14,93%, -10,08%]
The difference in followers is between 1 and 2 st. deviations	2181	-2,79% [-7,27%, 1,92%]	-12,45% [-18,67%, -5,96%]	-22,03% [-28,23%, -15,36%]
The difference in followers is between 2 and 3 st. deviations	2288	-3,43% [-7,36%, 0,67%]	-10,94% [-17,28%, -4,37%]	-25,32% [-32,45%, -17,50%]
The difference in followers is larger than 3 standard deviations	1588	-6,61% [-11,12%, -2,09%]	-16,43% [-23,73%, -8,87%]	-18,50% [-27,57%, -7,73%]

Table 9: The sentiment bias with the number of twitter followers as the measure of popularity.

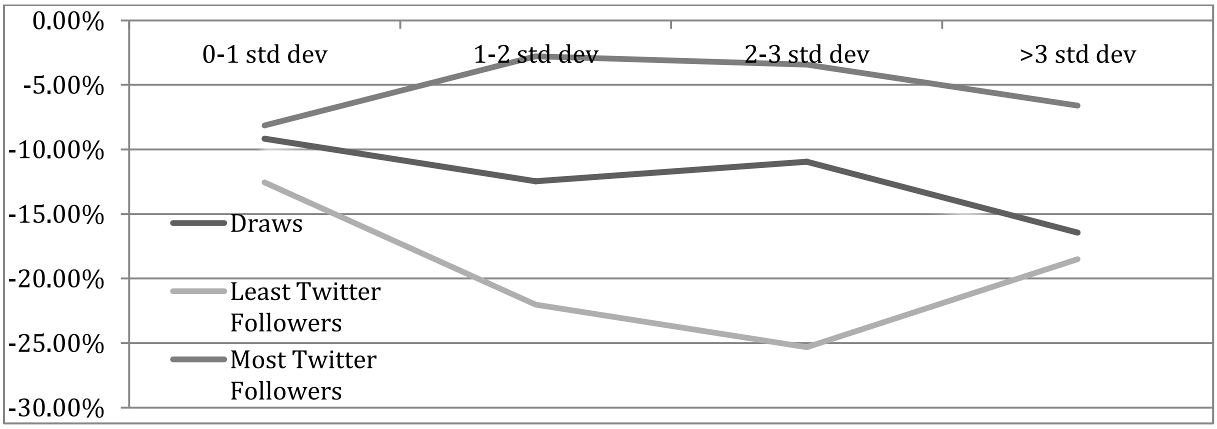

Figure 5: The sentiment bias with the number of twitter followers as the measure of popularity.

Examining these results it can be seen that for all sizes of differences the strategy of betting on the team with the most Twitter followers leads to a higher return than the strategy of betting on the team with the least Twitter followers. The difference in returns is smallest when the difference in the amount of Twitter followers is small. The largest difference in returns is found for matches where the difference in number of Twitter followers lies between two and three standard deviations.

6.3 The two by two design

The results so far show that betting on favourite teams leads to higher returns than betting on longshots and betting on popular teams leads to higher returns than betting on non-popular teams. The differences in returns become larger when the differences in popularity or favouritism become larger. Does this directly mean that the favourite-longshot bias and the sentiment bias are present? In both analyses there is not controlled for the other bias. It could be the case that the more popular teams are as well the better teams and are therefore favourite as well. Therefore it is not clear whether these results can be explained by the favourite-longshot bias or by the sentiment bias.

This is the reason why betting strategies with a two by two design are examined as well. The four betting strategies in this two by two design are: Betting on the teams that are the favourite according to the odds *and* have more Twitter followers than the opponent (1), betting on the teams that are the favourite according to the odds *and* have less

Twitter followers than the opponent (2), betting on teams that are not the favourite according to the odds *and* have more Twitter followers than the opponent (3), betting on teams that are not the favourite according to the odds *and* have less Twitter followers than the opponent (4).

For these four betting strategies the returns are calculated and can be compared. In table 10 the number of games where each strategy was possible, the returns for each strategy and the 95 percent confidence interval for each return can be seen.
It is important to note that a betting strategy of betting on teams that are the favourite and have the most fans is only possible when such a team is playing in a particular match. When this is not the case, for this particular match this strategy is not possible. From table 10 it can be seen that in 13134 matches the favourite has the most fans as well. Only for this part of the dataset the betting strategies of betting on the team that is the favourite and has the most fans and the strategy of betting on the team that is the longshot and has the least fans are possible. In 6908 matches the favourite has less fans than the opponent. Only for these matches the betting strategies of betting on favourites that have less fans and betting on longshots that have more fans are possible.

	Favourites (by odds)	Longshots (by odds)
Most Twitter followers	13134 matches -3,80% [-5,37%, -2,21%] Avg. odds favourite: 1.86	6908 matches -12,67% [-16,14%, -9,05%] Avg. odds favourite: 2.12
Least Twitter followers	6908 matches -9,96% [-12,46%, -7,50%] Avg. odds favourite: 2.12	13134 matches -18,54% [-21,54%, -15,38%] Avg. odds favourite: 1.86

Table 10: two by two design of betting strategies based on odds and Twitter followers.

The betting strategy of betting on teams that have the most fans and are the favourite leads to the highest return. This betting strategy has a return of -3,8 percent. The betting strategy of betting on longshots and the team with the least fans leads to a return of -18,5 percent, this is the lowest return of all four betting strategies. It is interesting to compare these returns with the other two. Betting on teams that are favourite, but do not have the most fans yields a return of -10,0 percent. This is significantly (see confidence intervals) lower at a 5 percent significance level than the -3,7 percent of betting on teams that are favourite and have the most fans. In both strategies you only bet on favourite teams, but the return of the strategy where the favourite team has the most fans as well leads to a significant higher return. This means that it matters whether a team has the most fans or not. This is the first indication of the sentiment bias. Comparing the other two betting strategies, it can be seen that the betting strategy of betting on longshots that have the most fans leads to a significant higher return (see again the confidence intervals) than betting on longshots with the least fans. In both strategies you bet on the longshots, but when the longshot has more fans than its opponent this leads to significant higher returns.

Interesting to see as well is that the return of the betting strategy of betting on the favourite team that has more fans than the opponent leads to a higher return than the strategies of betting on the favourites (without looking at fans, see table 2) and betting on the teams with the most fans (without looking at who is the favourite, see table 7). It seems that those two statistics about a certain team complement each other. Both statistics lead to a higher return when they are used in the betting strategy.

One limitation of this analysis is the fact that you compare strategies over different matches. It could be the case that these matches differ in the average odds for the favourite. When this is the case, the difference in average odds for the favourite could explain the differences in returns. As showed in table 10 the average odds for the favourites is lower (1.86) for matches where the favourite team has the most twitter followers as well than for matches where the favourite team does not have the most twitter followers (2.12). To see whether the results still hold when the average odds per subset are the same, or at least closer, the same analysis is performed for subsets of the dataset. The same two by two design is performed for matches where the favourite has an odd below 1.70, between 1.70 and 2.20 and above 2.20. By doing this, the matches that are compared are still different, but the difference in average odds is smaller. This is shown in table 11, 12 and 13.

	Favourites (by odds)	Longshots (by odds)
Most Twitter followers	4729 -4,65% [-6,68%, -2,65%] Avg. odds favourite: 1.4702	636 -17,33% [-31,83%, -1,29%] Avg. odds favourite: 1.5689
Least Twitter followers	636 -8,68% [-14,85%, -2,68%] Avg. odds favourite: 1.5689	4729 -23,62% [-29,41%, -17,14%] Avg. odds favourite: 1.4702

Table 11: two by two design for matches where the odds for the favourite are below 1.70

	Favourites (by odds)	Longshots (by odds)
Most Twitter followers	5412 -3,76% [-6,42%, -1,24%] Avg. odds favourite: 1.9262	3152 -12,82% [-18,30%, -7,29%] Avg. odds favourite: 1.9733
Least Twitter followers	3152 -9,57% [-12,89%, -6,09%] Avg. odds favourite: 1.9733	5412 -15,05% [-19,16%, -10,66%] Avg. odds favourite: 1.9262

Table 12: two by two design for matches where the odds for the favourite are between 1.70 and 2.20

	Favourites (by odds)	Longshots (by odds)
Most Twitter followers	2993 -2,55% [-6,76%, 1,59%] Avg. odds favourite: 2.3715	3120 -11,58% [-16,15%, -7,04%] Avg. odds favourite: 2.3805
Least Twitter followers	3120 -10,62% [-14,71%, -6,62%] Avg. odds favourite: 2.3805	2993 -16,84% [-21,45%, -12,11%] Avg. odds favourite: 2.3715

Table 13: two by two design for matches where the odds for the favourite are above 2.20

Looking at the average odds per subset it can be seen that the differences are smaller now. Especially when you look at table 13. The difference in average odds per subset is smaller than 0,01, the number of matches per subset is almost equal and the differences in returns are still there. Betting on favourites that have the most Twitter followers as well leads to a return of -2,6 percent, while betting on favourites that do not have the most Twitter followers leads to a return of -10,6 percent. This means that it matters whether a team has a lot of fans or not. The same can be seen for longshots. Betting on longshots that have the most Twitter followers leads to a return of -11,6 percent and betting on longshots that do not have the most Twitter followers leads to a return of -16,8 percent. In other words, for both favourites as for longshots it matters whether the team is the more popular team. This is evidence for the sentiment bias.

Betting on a team with the most Twitter followers that is the favourite as well leads to a significantly higher return than betting on a team with the most Twitter followers that is the longshot. This means that it matters for the return whether a team is the favourite or not. Betting on a team with the least Twitter followers that is the favourite leads to a higher return than betting on a team with the least Twitter followers that is a longshot. Also here, it matters whether a team is favourite or not. This is evidence for the favourite-longshot bias.

Overall, the results in this section point in the direction of the presence of the favourite-longshot bias and the sentiment bias in the odds. When the biases are examined separately it can be seen that the returns are higher for favourites than for longshots and that the returns are higher for popular teams than for less popular teams. The larger the difference in favouritism and popularity, the larger the difference in returns. When the two biases are combined in the two by two design these results are still present. Good to note as well is that a higher risk is not rewarded with a higher expected return. In other financial markets this is, in general, the case. In the football betting market apparently it is not. Bets with a small probability to win and with large payouts when you win (longshots) have smaller expected returns than bets with a larger probability to win and a smaller payout when you win.

To answer the research questions: It can be concluded that the favourite-longshot bias and the sentiment bias are present in the odds of this dataset. When the biases are examined separately it can be seen that the returns are higher for favourites than for longshots and that the returns are higher for popular teams than for less popular teams. When these two biases are combined in the two by two design, both biases are still present. In this dataset there are two betting strategies found that led to a positive return. The first one is betting on teams with odds below 1.20 and the second one is betting on teams where the average home attendance is more than three standard deviations larger than the average home attendance of the opponent. However, for these two betting strategies the confidence intervals of the returns are not entirely positive and because this kind of matches is scarce, the opportunities to exploit these potential profitable betting strategies are scarce as well.

7. Conclusion & Discussion

This paper investigates the presence of two behavioural biases in the match-odds of five big European football competitions. The behavioural biases that are examined are the favourite-longshot bias and the sentiment bias. The odds that are used are averages of the odds published by a large number of bookmakers. The competitions in the dataset are the Premier League (England), the Championship (England), The Primera División (Spain), The Serie A (Italy) and the Ligue 1 (France). In total, the dataset consists of 20289 matches. The favourite-longshot bias means that people overbet longshots (non-favourites) and underbet favourites. When this bias is present in the odds the simple betting strategy of betting on favourites leads to higher returns than the betting strategy of betting on longshots. The sentiment bias means that people bet 'too much' on the team they support and 'too little' on the other teams. When this bias is present in the odds, the odds depend on the number of fans of each team. The measures that are used as proxies for favouritism (probability to win a match) are the total amount of wages paid out per club per year and the odds given by the bookmakers. The measures that are used as proxies for popularity of a club are the average home attendance of the club in a season and the number of Twitter followers of the official Twitter account of the club. The presence of the two biases is investigated by comparing the returns of betting strategies that make use of the measures for favouritism and popularity.

The results of this paper show that both the favourite-longshot bias and the sentiment bias are present in the match-odds. Independent of which measures are used as proxies for favouritism and popularity, the results show that the betting strategy of betting on favourites leads to a significant higher return than the betting strategy of betting on longshots. The betting strategy of betting on the most popular teams leads to a significant higher return than the betting strategy of betting on less popular teams. The differences in returns between the betting strategies are largest when the difference in popularity or favouritism between the two teams is large. In other words, the biases are largest in the more extreme cases. These results hold as well when is controlled for the other bias. This means that combining these two biases leads to the highest returns. In this dataset two betting strategies are found that led to a positive return. The first one is betting on teams with odds below 1.20 and the second one is betting on teams where the average home attendance is more than three standard deviations larger than the average home attendance of the opponent. However, for these two betting strategies the 95 percent confidence intervals of the returns are not entirely positive and because this kind of matches is scarce, the opportunities to exploit these potential profitable betting strategies are scarce as well.

This paper has some limitations as well. The first one that I would like to mention is the use of the number of Twitter followers as a proxy for team popularity. The number of Twitter followers of the official Twitter account of a club in may 2015 is used as popularity proxy for the years 2001 until 2011. Here, it is assumed that popular clubs in may 2015 were popular clubs between 2001 and 2011. For most of the clubs in the dataset this assumption will hold, but probably there are some clubs of which the popularity has changed significantly over the years. Therefore, for future research, it might be nice to use (next to the average home attendance, which is accurate) a more accurate popularity measure. One possibility is, for example, the number of Google searches per club per year. Another limitation I would like to mention is that I used the average home attendance for the same season as when the matches are played. Of course this is the best proxy for team popularity, but at the moment the odds are published and one can bet on this particular match, the average home attendance for this particular season is still unknown. I do not think it would make a large difference in the results, but maybe in future research the home attendances for the last season could be used as the measure of team popularity to make the comparison of the betting strategies more 'fair'.

References

Ali, M. M. (1977). Probability and Utility Estimates for Racetrack Bettors. *Journal of Political Economy , 85* (4), 803-815.

Avery, C., & Chevalier, J. (1999). Identifying Investor Sentiment from Price Paths: The Case of Football Betting. *Journal of Business , 72* (4).

Babad, E., & Katz, Y. (1991). Wishful Thinking - Against All Odds. *Journal of Applied Social Psychology , 21* (23), 1921-1938.

Braun, S., & Kvasnicka, M. (2013). National Sentiment and Economic Behavior: Evidence From Online Betting on European Football. *Journal of Sports Economics , 14* (1), 45-64.

Cain, M., Law, D., & Peel, D. (2000). The Favourite-Longshot Bias And Market Efficiency In UK Footbal Betting. *Scottish Journal of Political Economy , 47* (1), 25-36.

Dixon, M. J., & Pope, P. F. (2004). The value of statistical forecasts in the UK association football betting market. *International Journal of Forecasting , 20* (4), 697-711.

Feddersen, A., Humphreys, B. R., & Soebbing, B. P. (2013). Sentiment Bias in National Basketball Association. *Working Paper , 13* (3).

Forrest, D., & Simmons, R. (2008). Sentiment in the betting market on Spanish football. *Applied Economics , 40* (1), 119-126.

Griffith, R. M. (1949). Odds Adjustments by American Horse-Race Bettors. *The American Journal of Psychology , 62* (2), 290-294.

Humphreys, B. (2010). Point Spread Shading and Behavioral Biases in NBA betting markets. *Rivista di Diritto ed Economia dello Sport , 6* (1), 13-26.

McGlothlin, W. H. (1956). Stability of Choices among Uncertain Alternatives. *The American Journal of Psychology , 69* (4), 604-615.

Ottaviani, M., & Sorensen, P. N. (2010). Noise, Information, and the Favorite-Longshot Bias in Parimutuel Predictions. *American Economic Journal: Microeconomics , 2* (1), 58-85.

Peeters, T., & Szymanski, S. (2014). Financial Fair Play in European Football. *Economic Policy , 29* (78), 343-390.

Quiggin, J. (1982). A theory of Anticipated Utility. *Journal of Economic Behaviour and Organization , 3* (4), 323-343.

Thaler, R. H., & Ziemba, W. T. (1988). Anomalies: Parimutuel Betting Markets: Racetracks and Lotteries. *The Journal of Economic Perspectives , 2* (2), 161-174.

Tversky, A., & Kahneman, D. (1992). Advances in Prospect Theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty , 5* (4), 297-323.

Weitzman, M. (1965). Utility Analysis and Group Behavior: An Empirical Study. *Journal of Political Economy , 73* (1), 18-26.

William, L. W., & Paton, D. (1998). Why are some favourite-longshot biases positive and other negative? *Applied Economics , 30* (11), 1505-1510.

Woodland, L. M., & M., W. B. (2001). Market Efficiency and Profitable Wagering in the National Hockey League: Can Bettors Score on Longshots? *Southern Economic Journal , 67* (4), 983-995.

Woodland, L. M., & Woodland, B. M. (1994). Market Efficiency and the Favorite-Longshot Bias: The Baseball Betting Market. *The Journal of Finance , 49* (1), 269-279.

Woodland, L. M., & Woodland, B. M. (2010). The Reverse Favorite-Longshot Bias in the National Hockey League: Do Bettors Still Score on Longshots? *Journal of Sports Economics* , 1-12.