The favorite longshot bias in tennis tournaments

ERASUMUS UNIVERSITY ROTTERDAM

Erasmus School of Economics

Master Behavioral Economics Thesis

2013

Name: Julien van ‘t Leven

Student ID: 323640

Supervisor: Umut Keskin

Table of contents

[The favorite longshot bias in tennis tournaments 1](#_Toc361442749)

[Table of contents 2](#_Toc361442750)

[Introduction 3](#_Toc361442751)

[Literature review 5](#_Toc361442752)

[Early evidence for the favorite longshot bias 6](#_Toc361442753)

[Favorite longshot bias in other sports with different forecasting methods 8](#_Toc361442754)

[Market efficiency and betting strategies 11](#_Toc361442755)

[Contradicting literature 13](#_Toc361442756)

[Explaining the favorite long shot bias: 15](#_Toc361442757)

[Data 18](#_Toc361442758)

[Method 19](#_Toc361442759)

[Discover the favorite longshot bias 19](#_Toc361442760)

[Statistical test 20](#_Toc361442761)

[Results 21](#_Toc361442762)

[ATP level 21](#_Toc361442763)

[Grand Slam tournaments: 25](#_Toc361442764)

[Surfaces 27](#_Toc361442765)

[Home advantage 29](#_Toc361442766)

[Conclusion 31](#_Toc361442767)

[Discussion 32](#_Toc361442768)

[Bibliography 34](#_Toc361442769)

[Appendix 37](#_Toc361442770)

# Introduction

The first referring to the favorite longshot bias was (Griffith, 1949). Griffith was a psychologist who studied the US horse races. He looked at the ability of people to correctly interpret the winning chances of horses in several races. Griffith found that bettorswere very good at forecasting the probable winners but were not good at dealing with extremes: the bettors underestimated the chance of winning for extreme favorites and overestimated the chance of winning for extreme non-favorites.

  After this publication numerous others researched this bias. The early literature focused on race track settings, like horse, dog and boat races. Subsequently people looked for the favorite longshot bias in other sports. For example soccer (Cain, Law, & Peel, 2000), American football (Zuber, Gandar, & Bowers, 1985) and baseball (Woodland & Woodland, 1994).

  The favorite longshot bias is found in many countries. The first papers examining this bias did their research in the US (Ali, 1977) or UK (Dowie, 1976) and found the bias in both countries. After these early papers people investigated several other countries and found the favorite longshot bias in nations far apart from each other. For example from Germany (Winter & Kukuk, 2006) to New Zealand (Gandar, Zuber, & Johnson, 2001).

Literature examined the existence of the favorite longshot bias in all kind of sport betting mechanisms. The three most used sport betting mechanisms are pari-mutual betting, fixed odds betting and point spread betting. Studying these three mechanisms seemed to give different outcomes with respect to the favorite longshot bias. For example, the point spread betting mechanism did not display the favorite longshot bias. While the pari-mutual mechanism did generate a favorite longshot bias most of the times.

  So not all literature demonstrated the favorite longshot bias. For example (Woodland & Woodland, 1994) found a reverse longshot bias in the US major league baseball and the national hockey league. Also research in Asia gave no evidence for a favorite longshot bias, according to (Coleman, 2004).

In this paper research is focused on the tennis sport where available literature is really scarce. One of the few papers that did some research on tennis is (Cain, Law, & Peel, 2003), who researched 91 tennis matches on grass at the Wimbledon championship. They did not find a favorite longshot bias. Another paper investigating tennis was written by (Forrest & McHale, 2007).They examined tennis matches during three seasons at the highest level (ATP) and found the favorite longshot bias.

The aim of this paper is to investigate if the favorite longshot bias occurs in the tennis betting market in the period 2010, 2011 and 2012. Special attention is paid to Grand Slam tennis tournaments. If the favorite longshot bias indeed exists in the tennis betting market, according to (Williams & Paton, 1998), a favorite longshot bias would disappear when looking at high profile tournaments in tennis such as Grand Slam tournaments. Due to the size of these tournaments every contestant will be at the top of his game and therefore insider’s information is limited. A second feature that is taken into account is the surface tennis is played on. Thirdly, the feature of home advantage is examined. For a tennis player this means playing in the country of his or her nationality. Other literature only discussed home advantage as a team playing at their own club. For example (Schnytzer & Weinberg, 2008) found a favorite longshot bias when focusing on home advantage.

An added value of this thesis with respect to other literature is the high number of matches played in tennis between huge underdogs and heavy favorites that was taken into account. This is because of the dominance of a small number of tennis players in the world in the period 2010-2012. This results in more matches played in the highest and lowest category (these categories have the main focus). As a result there are smaller intervals per category possible, for example in (Forrest & McHale, 2007) each prior match winning probability category has an interval of 10%. This thesis exhibits intervals of 5%. Another advantage of studying tennis is that tennis is an individual sport. Therefore the betting on players is not biased by a huge fan base as in team sports. Normally this huge fan base would consequently bet on the team they are supporting, even if their team is the heavy underdog, thereby highly influencing the odds.

In this thesis the tennis matches played in the years 2010, 2011 and 2012 that were available to be betted on at Bet365 were examined. The matches investigated were played at the highest level possible in tennis, namely the ATP level. This thesis found for the period 2010-2012 a favorite longshot bias in the tennis betting market. This favorite longshot bias also occurred for each year separately. Looking at the home advantage there were no differences for the whole period. If only home games were considered, a favorite longshot bias still occurred, but the overall return was three times less negative as for the whole sample. The subsample of four different surfaces, namely clay, grass, indoor hard-court and outdoor hard-court, showed a favorite longshot bias for the surfaces clay, outdoor hard-court and indoor hard-court. Only grass gave a different outcome: in contrast to the other surfaces, the underdog category gave a positive return on grass. So for the other surfaces the underdogs were over betted, while on grass the underdogs were under betted.

**Literature review**

The literature section is split up in five parts. In the first part of this chapter the early literature about the existence of the favorite longshot bias is elaborated. The literature during and after the digital revolution is discussed next. The digital revolution marks an important moment because the upcoming online betting market made the betting on different sports possible. The third part contains the contradicting literature, as the favorite longshot bias was not found by everyone. Fourthly, market efficiency and the associated profitable betting opportunities are discussed. The final part of this literature chapter considers several explanations for the favorite longshot.

## Early evidence for the favorite longshot bias

In this section of the literature review the early literature is discussed. Early literature mainly focuses on horse racing. This kind of market is called a pari-mutual market. A pari-mutual market is a market where all the money betted in a race is pooled. From this pool of money the costs for organizing the race are covered (called track take). After the race, the remaining money is paid to the winners according to the height of their bet (bettors that played the winning horse). In the early literature the studied countries were the United Kingdom and the United States.

The first paper which mentioned the favorite longshot bias was (Griffith, 1949). As a psychologist he was interested in how people behave in a race track setting. In such a pari-mutual betting market setting the betting behavior creates the odds the people play for which is an interesting phenomena for a psychologist. Griffith was interested if the bettors at the horse race tracks in the United Kingdom could predict the exact probabilities of a horse winning the race. He researched this by comparing the odds derived by the betting procedure of the bettors and transforming them in probabilities. Then he compared this probability with the exact probability that is derived by calculating how often a horse from each odds category actually wins. His first finding was that the socially determined odds (probabilities) were very similar to the true probability of a horse from an odds category winning. So the best horse was given the highest probability and the worst horse was given the lowest probability from the bettors. There was only one deviation between this socially determined probability and true probability which occurred at the extremes. Namely there was a systematic overvaluation of the bettors of the chances of a longshot (very low probability) horse to win the race. While the horses with the highest true chance to win were systematically undervalued by the bettors.

After this first investigation of the betting behavior on a race track, (McGlothlin, 1956) was another psychologist who investigated the behavior at a race track. McGothlin wanted to look at the stability of risk-taking behavior over a series of events. Do bettors show the same betting behavior at the first race as at the last race of the day? He investigated this by looking at the expected value of a 1 dollar bet in the 8 track odds categories made. Also he distinguished between the first and last (eighth) race per day. This study included eleven race tracks from two states (New York and California) of the United States. The results concerning the relationship between the expected winning probabilities calculated in each odds group and the socially determined odds were the same as found by (Griffith, 1949). So the high probability horses showed a higher expected value than the low probability horses, again indicating an undervaluation of favorites and overvaluation of underdogs. Secondly McGothlin found that the behavior of bettors changed at the end of the day. Betting on the favorite becomes more valuable and even on average profitable when the track take is taken into account. Furthermore the heavy underdogs are less overvalued, but the average low probability bets are overvalued. So the extremes were estimated more accurately at the end of the day, in contrast to the horses with more average probability.

Following the two previous authors, (Ali, 1977) also investigated a horse racing track setting. But he investigated the existence of the favorite longshot bias for a much larger dataset. Even ten times larger than the dataset (Griffith, 1949) and (McGlothlin, 1956) used. The Sample Ali used was 20507 horse races at three different New York horse race tracks in the years 1970 to 1974. Ali had a new feature where the allocation to categories were based on. Previously Griffith based the allocation in groups purely on odds. In this way there was a chance that two competing horses were in the same group. Therefore an average is biased, because only one horse can win the race. So a representative average horse cannot be captured. Ali made sure that in each category there were no horses competing in the same race. The results showed that on every track in each year there was a favorite longshot bias. So in a pari-mutual setting on every track the bettors overvalued the chance that an extreme underdog would win the race, while the extreme favorites in the races were on average undervalued.

Another study done in the United States in a pari-mutual market was done by (Arsch & Malkiel, 1982). They investigated a dataset containing horse races from the Atlantic City racetrack. In this paper they came to the same conclusions as in (Griffith, 1949). Firstly, they found that the odds resulting from the betting behavior of the bettors are a really good indication of how the horses will finish. The other main conclusion is that on average the favorites are ‘good’ bets and the least favorites are ‘bad’ bets. Arsch & Malkiel are one of the first authors discussing the link between betting and financial markets. They noticed the similarities: in both situations one does not know what ones future earnings are with certainty. There are a lot of participants involved in these markets. There is a lot of information available and this information can come with advice of professionals. Involvement in these markets also provides knowledge of what others in the market do.

(Rubinstein, 1985) investigated a financial market and found an indication for the favorite longshot bias. He investigated the put and call market. One of the main conclusions in this paper was that short- maturity out-of-the-money calls are priced significantly higher than the Black-Scholes model would predict. This means that call options that have a striking price higher than the market price are valued higher by the option buyers than the Black-Scholes model predict. In the paper of Rubinstein this finding is the only significant deviation of the valuation of the Black-Scholes model and the true value of the options. This could be evidence that the longshot in option valuation is overvalued, indicating a same finding for the longshots as in the sports betting literature and maybe the financial market exhibits the same biases as the betting market.

Rubinstein was not the only one noting similar behavior in sport betting markets and financial markets. Another paper investigating the favorite longshot bias in a financial market has been written by (Hodges, Tompkins, & Ziemba, 2003). They examined this for the S&P500 futures, the FTSE 100 futures and the British Pound/US Dollar futures in the period March 1985 to September 2002. A future is an agreement to buy or sell an asset or a financial value at some time in the future that has to be supplied by the seller to the buyer.Firstly for call options on the FTSE 100, they found a relationship between probabilities and true mean returns that is very similar to the favorite longshot bias in sport markets for call options. For these FTSE 100 3- and 1month call options a long shot bias exists if these options are deep out-of-the-money. Secondly, for the S&P 500 there was even a profit for buying deep-in-the-money calls and increasingly bigger losses as the calls were going out-of-the-money. In the options on the British pound/US Dollar trading market they found no evidence concerning the favorite longshot bias. They concluded that the patterns they observed in both S&P 500 and FTSE 100 were analogous to the favorite longshot bias in sport betting markets.

## Favorite longshot bias in other sports with different forecasting methods

After the early literature, different authors investigated if the favorite longshot bias also existed in other sport markets. There were also studies published based on data from different countries and continents. Another new feature in the later literature are the new betting methods, arising from all the new inventions in the digital world. For example, the online betting is a new booming business in the late years of 2000 resulting in new betting markets. Whereas the betting markets used to be mainly pari-mutual markets where the bettors decided what the odds previous to the race/match were, nowadays so called bookmaker markets have arisen. The main difference with the pari-mutual markets is that the odds before a match are set by a bookmaker agency and are not or just slightly influenced by bettor’s behavior. Because of all the new digital opportunities, statistical models started to play a more dominant role in predicting sport outcomes. These models are based on all kind of characteristics whereas in the past the only characteristic that played a role were the participants in a match. For example, characteristics like the previous results of other teams, the results of head to head encounters, the current ranking, the surface played on, the weather conditions, etc. Based on these characteristics the models give a probability for the chance of winning. There are also markets where an expert decides what odds a bookmaker agency sets or where expert just influence bettors with their advice. By experts are meant sport writers, editors of newspapers or sport magazines or even sports commentators. A good example of an expert who heavily influences a probability developing procedure is the newspaper expert, who shows his expected probability on coming horse races in the morning newspaper before the betting window is opened. Bettors can be biased by an opinion of an expert, for example the anchoring bias is triggered. Therefore the probability stated by the expert can influence the decision of the bettor.

In 2000 (Cain, Law, & Peel, 2000) investigated 2800 UK football matches to see if the odds proposed by the bookies were good measures for the actual outcome of the matches. So this study was done in a bookmaker market background, where the odds are predetermined by the bookmakers and the behavior of the bettors has no (or slight) impact on the odds (so called fixed odds). The first finding was that not only the bettors in a pari-mutual betting market predict the actual outcome correctly, according to (Griffith, 1949), but the odds that the bookmakers set are on average also very similar to the actual outcome. So they do a good job in predicting football match outcomes. Secondly, Cain et all searched for the existence of the favorite longshot bias in the football betting market and indeed found it. They used a simple strategy of dividing the whole sample in a few groups based on probabilities and presuming a bet of 1 dollar for every match in this group. Then the underdog group generates the lowest mean return, while the favorite group generates the highest return. After this they investigated the betting behavior for football match outcomes. This resulted in another indication for the favorite longshot bias, namely bets on a low odds outcome score are on average better than a high odds outcome score. So people like to bet on an unusual outcome.

Another study investigating a team sport is done by (Andrikogiannopoulou & Papakonstantinou, 2011). This article contains both aggregated and individual data of 10000 randomly selected major soccer matches of an online European bookmaker. They investigated match results, but also the behavior of individual bettors. Firstly, this article finds a favorite longshot bias in the wagering soccer market examined for the years 2005-2009. Therefore they suggested that the weak form of market efficiency can be rejected and the prices settled by bookmakers do not reflect all the past public information disposed. Further they found that, when looking at the exact outcome of soccer matches, even in this kind of betting there was a favorite longshot bias. Especially the winning of home games and away games of favorites are under betted. When they looked at the individual behavior of the bettors they found that only 2% constantly betted on the extreme favorites, while 6% acted the other way around and betted on the extreme underdog. So only 2% exploited the under betting of the extreme favorites consciously.

Comparable to what (Andrikogiannopoulou & Papakonstantinou, 2011) did, the authors (Vlastakis, Dotsis, & Markellos, 2008) also investigated soccer matches in Europe. They did this for European football league games for the years 2002-2004. Using the odds given by 5 large online bookmakers, who placed a possible fixed odds bet a week before every football match, Vlastakis et all found that the odds given by these bookmakers exhibited the favorite longshot bias.

The opposite applies to (Cain, Law, & Peel, 2003), who did not find a favorite longshot bias in soccer. On the other hand they did not have huge favorites (winning probability of more than 80%) in their sample. They also investigated other sports like tennis, boxing, horseraces, baseball, snooker and cricket. For cricket, boxing and horseraces a clear favorite longshot bias was found. Whereas for tennis, baseball and snooker this pattern of low returns for the underdog and high returns for the favorite did not emerge. For these three sports no reverse longshot bias was found as suggested by (Woodland & Woodland, 1994) for the baseball betting market.

(Steckler, Sendor, & Verlander, 2010) made a comparison between the three methods discussed above. They distinguished three different sources for forecasting, namely the market betting forecasts, statistical model calculations and experts. Their main research was based on answering the question which calculation method is the best to use? Others found that if the performance of these markets is measured by the amount of successful predictions, the expert in the morning newspaper did almost as good as the pari-mutual system later that day, namely 28,7% to 29,4%, according to (Stephen, 1979). Comparing the three markets yielded a different outcome: the experts and models correctly predicted the winner in a match between two teams in 60% of the cases. Only the market had a significant higher accuracy in predicting the winner. Steckler et all 2010 concluded after explaining all the findings for accuracy of sport forecasting that the betting markets in all sports develop unbiased forecasts. According to them, future research for the explanation of the favorite longshot bias should be based on the understanding of the individual’s ability to accurately process all the information available.

## Market efficiency and betting strategies

A topic linked to the favorite longshot bias is market efficiency. There are three types of market efficiency, namely Strong, Semi-strong and Weak form. The weak form market hypothesis states that all prices reflect all past publicly available information. While the Strong form market hypothesis states that even all insider information is reflected in the price. So in a horse race betting market the Weak form market hypothesis indicates that all publicly available information about the horses is translated into the odds. Eventually the odds should reflect the true probability of winning. So a profitable strategy, in the case of the favorite longshot bias betting on the favorite, would indicate that the weak form market efficiency hypothesis would not hold.

The first one testing a horse race betting market for market efficiency was (Dowie, 1976). He thought that the great advantage of a pari-mutael betting market versus a bookmaker market was that the same odds derived in the pari-mutual betting market represent the same amount of money out of the pool of money betted. So when the odds are the same in pari-mutual betting markets, the ratio of money betted on a horse and the total is also the same. Whereas in bookmaker markets this is uncertain.

Dowie worked with forecast prices and starting pricesto examine the market efficiency. A starting price is defined as the price on a horse, derived from the pari-mutual mechanism. These odds are settled at the start of the race*.* The Forecast Price is the price suggested by the ‘professionals’ of the leading newspaper in the morning prior to the betting of the race. With these two prices the authors tried to see if insider’s information was available. If this is the case then the Starting Price has to generate a higher return than the Forecast price. They researched the Starting Price and the Forecast Price on the realized probabilities separately by using regression analysis and determined the correlation and R-squared for the two Prices regarding the realized probability. The outcome showed that the correlation of the Forecast price with the realized probability was at least as big as the correlation of the Starting price and the realized probability. Thereby raising serious doubts about the existence of insider’s information.

Dowie concluded that there was an indication that the betting markets are strongly inefficient, because they do not immediately reflect new information in the prices (FP and SP behaving similarly). The authors believe that the sports betting market is reasonably weakly efficient, because SP reflects past public information (FP).

Another way to determine if markets are inefficient is checking whether a profitable betting opportunity exists. According to (Ziemba & Thaler, 1988) there are two strategies bettors can follow. Firstly there is a fundamental strategy, which is based on public information. With this information they try to find a horse with a winning probability higher than the market odds. The difference between winning probability and market odds has to be high enough to exceed the track take in order for the strategy to be profitable. Secondly there are the technical strategies; these are all the strategies people invent themselves. Most of these strategies are based on the fact that people attribute different weights to the several characteristics, for example people think the head to head is underrated.

Ziemba earlier already found a profitable strategy written in (Hausch, Ziemba, & Rubinstein, 1981). This strategy was based on the height of a bet on ‘horse i’ in the win pool compared to the height of a bet on ‘horse i’ in the place pool. (A win bet is profitable if the betted horse gets over the finish first; a place bet is profitable when the betted horse gets first or second). The main idea is that people take action when the proportion betted on ‘horse i’ in the win pool is far bigger than the proportion betted on ‘horse i’ in the place pool. In this example the bettor places money on ‘horse i’ in the place pool. When (Hausch, Ziemba, & Rubinstein, 1981) published this strategy in their paper they thought this profitable strategy would disappear, but the same strategy still yielded a profitable return 5 years later, according to (Ziemba & Thaler, 1988).

Another possibility to develop a profitable strategy and see if the markets are efficient became available when cross betting was introduced. This meant that people at their home racetrack could bet on big races held somewhere else. If markets are efficient no difference in odds could be expected, but this was not the case. Huge differences were found and could be used in a profitable way when bettors were able to get in contact with another race area. Nowadays this occurs on the online gambling market, because it is now possible to bet with different prior match odds on a match at least 5 big online bookmakers.

Despite the strategies discussed above there are authors who believe that the search for profitable strategies is overrated. For example (Williams L. V., 1999). After he enumerated all the possible explanations for the favorite longshot bias this author concluded that depending on the explanation someone wants to accept as most likely, there is hardly any evidence that betting markets, assuming weak form information, provide chances for outsiders to earn abnormal returns.

Different literature examining the favorite longshot bias also looked at the role of home advantage. For example (Schnytzer & Weinberg, 2008) studied the existence of a home team’s advantage in the Australian Football league in the period 2002-2004. They showed that there was a significant bias in favor of teams with a home ground advantage. Another study done by (Colquitt, Godwin, & Swidler, 2004) first found a longshot bias in the baseball market. Then they researched if the longshot bias also existed by looking at games where the heavy favorites were teams not playing at home. Again a significant result at the 1% significance level appeared in favor of the favorite longshot bias. In contrast to this literature (Woodland & Woodland, 2001) showed that in the hockey betting market the opposite emerged when focusing on home teams. So they found that a reverse favorite longshot bias existed when focusing on home advantage. The data of Woodland & Woodland showed that from the 5402 matches examined 2702 matches were won by the home favorite, which is almost 50%. Therefore indicating no advantage.

## Contradicting literature

Concerning the favorite longshot bias there is a part of literature which states the total opposite of previous literature. This phenomenon is called the reverse favorite longshot bias where investigators found that the favorites are overvalued, while the underdogs are undervalued. People even claimed to have found profitable strategies to exploit the reverse favorite longshot bias.

One of the first authors that found these contradicting outcomes were (Woodland & Woodland, 1994). These two became the key authors of this part of literature. Their first research of the favorite longshot bias dates from 1994 and included the major league baseball games of the seasons 1979-1989. By processing this historical data, Woodland & Woodland tested if a favorite longshot bias existed in the major league baseball. Secondly they wanted to test if the market efficiency held. In this paper they did not find a favorite longshot bias, but a reverse one. This means that in the baseball market, the teams who get a low chance of winning beforehand are undervalued and the short odded teams are overvalued. A finding that is not in line with the previous literature. This may be a market inefficiency, because a simple strategy of betting on the underdogs yields much smaller losses than market efficiency would imply.

Another paper written by the same authors was (Woodland & Woodland, 2001). In this paper they examined a new betting market, namely the national hockey league. They researched if the national hockey league betting market is efficient and if a favorite long-shot bias exists. Woodland & Woodland found a reverse favorite longshot bias, where the short odds are over betted. This in contradictory to what is often found by others, especially in the racetrack betting research, for example (Ali, 1977) & (Griffith, 1949). On the other hand this is in line with what the authors earlier found for the baseball betting market. Their results led them to the assumption that the racetrack bettors are a unique kind. This could indicate that racetracks are not representative of market efficiency or behavior.

They even conducted a simple profitable in-sample strategy of betting on the underdog in away games. Following this strategy in the last four years of the sample period (1990-1996) could lead to a profit of 11%. They did not claim that this strategy is still used, but it is an indication that the market is not efficient.

After investigating hockey as a new sport domain, (Paul & Weinbach, 2005) studied the NBA basketball which was also relatively new. They found an over betting of the huge favorites and a more than fair bet when the money was placed on the clear home playing underdogs. This finding was in line with what Woodland & Woodland found.

Other literature that searched for the favorite longshot bias, but did not demonstrate its existence are (Dixon & Pope, 2004). In their paper they examined the fixed odds football betting market. The odds used in this paper are fixed before the game and are therefore not influenced by bettors´ betting behavior. Dixon & Dope found high odds for longshots and small odds for high probability outcomes. This is called a reverse favorite longshot bias; this is in accordance with (Woodland & Woodland, 1994). They found that a strategy based on a forecasting model using prior results can generate bets that have more-than-fair odds.

While (Woodland & Woodland, 1994) did not try to explain this reverse favorite longshot bias (Dixon & Pope, 2004) did. Because Dixon & Pope investigated a fixed betting market they searched for factors explaining the bias from the bookmaker’s perspective and discussed two possibilities.

Firstly this could be a rational response by the bookies to the cognitive biases the bettors have, in which case the bookies try to exploit this behavior. To be precise, if the bookmakers know there is a favorite longshot bias where people overestimate the chance of a longshot winning, then the bookmaker will offer less than fair odds on these longshots. This will increase the earnings for the bookmakers. This strategy implies that there is even a bigger chance to find the favorite longshot bias on the market, but (Dixon & Pope, 2004) found a reverse one.

So secondly it could be that the bookmakers themselves exhibit a cognitive bias. For example if the bookmaker is influenced by the reverse favorite longshot bias then the bookmaker will set the longshot odds more-than-fair. Thereby increasing the chance thatthe market displays a reverse favorite longshot bias.

## Explaining the favorite long shot bias

Explanations for this phenomenon come in all sorts. Firstly there is an explanation based on how people derive utility. Secondly an explanation is given with respect to the probability weighting of bettors. These first two explanations are based on biases with the bettors. Thirdly there are explanations based on informed bettors who participate in the betting markets. Fourthly literature explains the favorite longshot bias as a misuse of the data. Fifthly people suggested to look at this bias in an optimal game theory setting. Below, these explanations are further explained.

1. The Neoclassical approach states that the behavior of betting on very low odds, which are risky bets, can only be explained by risk loving utility functions. (Friedman & Savage, 1948).

In the favorite longshot bias context this means betting on a longshot can yield a high return, but with a small probability. While betting on a favorite can only yield small returns, but with a big probability. If individuals prefer riskier bets, the consequence is that the riskier bets are priced higher.

(Weitzman, 1965) examined the utility functions of bettors at a racetrack and found an indication for risk loving behavior of the bettors at a race track.

Also (Ali, 1977) tried to construct a representative Utility function for the bettors at the race track he investigated, under the assumption that bettors are sophisticated, have an equal Utility function and behave as a single race opportunity. He found that the people at the racetrack exhibited a risk loving utility function. According to Ali this could be an explanation for the favorite longshot bias.

This risk loving argument is further examined by (Quandt, 1986), who showed that if there are risk-loving bettors in the market, this has to result in a favorite longshot bias when the market reaches an equilibrium.

2. Leading psychologists (Tversky & Kahneman, 1979)developed the Prospect Theory. This was a new theory incorporating the criticism given to Expected Utility theory. A new important feature of this theory was the probability weighting function, who created a nonlinear relation between the decision weights and stated probabilities. The main expansion of this probability weighting function was the incorporation of people tending to overweigh small probabilities and to under weigh high probabilities. This could be an explanation for the favorite longshot bias, namely that the bettors cannot accurately process the odds given by the morning newspapers/ bet offices. When a small probability is given, it feels bigger to them than it actually is.

The first one being aware of a bias in estimating probabilities was the first one making notice of the favorite long shot bias, namely (Griffith, 1949). He suggested approximately the same explanation, but did not have any examples or prove.

3. Another explanation for the favorite longshot bias is the existence of well-informed bettors. First discussed by (Isaacs, 1953).  When an informed bettor knows that the predetermined odds on ‘horse i’ are lower than the actual probability, the bettor tries to exploit this. Since this takes place in a pari-mutual betting situation it is profitable to bet as low as possible. In this way the odds on ‘horse i’ will become as low as possible and the informed bettors get a larger profit in the case ‘horse i’ wins. So there is a tradeoff between the amount betting on ‘horse I’ and increasing the odds. Assuming that the favorite in the race would be of interest to informed bettors, this can result in low market odds, indicating a favorite longshot bias.

4. (Walls & Busche, 2003) came with a surprising explanation for the favorite longshot bias, namely that the research which demonstrated the bias, used wrong data. According to them, round-off odds data were used to calculate the bet volumes, whereas Walls and Busche claim that the exact data are needed to calculate the bet volumes. The difference can be found in the interpretation of the probabilities derived by the betting behavior. Early literature used round-off odds and Walls and Busche used exact odds as derived by betting behavior. The authors tested the hypothesis that using round-off data leads to a favorite longshot bias, with data from Hong Kong and Japanese horse race tracks, containing both round off and exact betting volumes. The Japanese horse track showed no signs of a favorite longshot bias when the odds derived from exact betting data were compared to the winning fraction of a ranked group. However if round-off data is used to derive the odds and is compared with the winning fraction in a group then the favorite longshot bias appears: the favorites are under betted and the longshots are over betted. Virtually the same applied to the Hong Kong horse track, where odds based on round-off data generated an under betting of the favorites.

5. (Shin, 1991) investigated what the optimal odds for a bookmaker would be, if the bookie knows the number of insiders operating in the market, but does not know who the insiders are. Shin developed a model with a bookmaker, insiders, outsiders and two horses, who played in an extensive form game. These players participated in three stages. Firstly a model sets the probabilities for the two horses.Secondly the bookmaker uses these probabilities to determine the odds on the two horses. Finally the insiders and outsiders can bet on the race. Shin investigated what the optimal strategy would be for everyone in his model. The result was that the bookmakers ideally give odds that are not in line with the true probability, but the betting odds will tend to understate the difference in winning probabilities of the two horses. So the presence of insiders resulted in odds containing the favorite longshot bias. This same problem was discussed by (Williams & Paton, 1998) for a fixed odds market. They showed that the higher the profile of a tournament or race, the lower the presence of insiders. Thereafter concluding that the presence of insiders in these kind of markets generated a positive longshot bias.

# Data

The data used in this Thesis is derived from the site: [www.tennis-data.co.uk](http://www.tennis-data.co.uk). On this site people can derive data from the top level of tennis containing all the tournaments of the ATP-tour for men and of the WTA-tour for women. For the ATP-tour data is available from 2001 to 2013 and for the WTA-tour the site provides data from 2007 to 2013. This data includes total game scores, surface, location, world rankings of players, world ranking points and pre match bet quotations of five betting agencies for all the tournament matches played. These betting agencies are Bet365, StanJames, Ladbrokes, Expekt and Pinnacles sport. So all information required to generate results for this thesis is included, except the nationality of all players. These nationalities can be found at <http://www.atpworldtour.com/Players/Player-Landing.aspx>. This is the official website of the ATP tennis organization and provides all the player profiles including their nationality.

[[1]](#footnote-1)Bet365 has a special feature, where they update the odds every fifteen minutes. In this way Bet365 can respond to betting activities against the odds prescribed by Bet365. If for example bettors put significant amounts of money on the underdog, then Bet365 could devaluate the odds of the underdog and upgrade the odds of the favorites. Such betting behavior indicates that the odds set by Bet365 are probably not very accurate and in this way they can deal with this problem. Important to point out is that bettors, who put their money in before the update, still play for the odds prescribed when they entered the betting. Only new entrances in the bet do this for the new odds. This updating indicates that the betting behavior is taking into account by the odds and the odds stated by the bookies do not deviate much from the betting behavior.

According to the data there is on average an over- round of 0.058%. This means that the bookmaker, if the proportion of betting on losers and winners is in the bookies advantage, is guaranteed to win a return of 0.069/1.069= 0.0645%. So the bookmakers of Bet365 on average put 6.45% on top of the odds they provide. In this way on average they will earn money for constructing these bets and they protect themselves against insider trading.

# Method

# In this chapter firstly the two main methods used for investigating the favorite longshot bias in tennis are discussed and the choice for the best method for this thesis is substantiated. Secondly the statistics used in this thesis is explained.

## Discovering the favorite longshot bias

Investigating the favorite longshot bias has been done in several ways. One of the main methods used by early literature is looking at the subjective and objective probability. The subjective probability is the probability an individual ascribes to a certain outcome. So this subjective probability is based on the odds given prior to the match and these subjective probabilities are divided into categories by a certain interval. Whereas the objective probability is the probability derived after the match. This objective probability is derived by dividing the number of wins in a category by the races ran or matches played. For example (Griffith, 1949) & (Ali, 1977) used this method.

This method is not suitable for this study, because of the larger number of categories used. Therefore the results will become unclear. Therefore in this thesis the method of (Cain, Law, & Peel, 2003) is used. The first thing to do with the data is to transform the odds into probabilities, this is done as follows:

**Winning Probability = 1/odds**

Due to this transformation the over round can be derived, simply by adding the probabilities of two players in a match. Another advantage is that the categories can now be derived by a constant interval. In this thesis, when the underdog category is sufficiently large, the intervals used for the categories are 5 %. So a prior chance to win the match of 0% - 5%, 5% - 10%,…., 95%-100%. In this way 20 categories are derived. This is in contrast to the early mentioned literature like (Ali, 1977), who based the categories on the horse’s rank in a race based on highest to lowest subjective probability. The advantages of this method were that no horses competing in the same race were in the same category. The number of horses in each category is almost equal, which reduces the variance. The disadvantage is that the categories do not deviate a lot in subjective categories. Therefore a horse ranked 2 and therefore assigned to category 2 can have a higher subjective probability. Another reason why the method of (Cain, Law, & Peel, 2003) is chosen.

Secondly the returns are calculated assuming a 1 euro bet on each match in a category. In this way the returns are automatically generated in percentages. After this the mean return of each category is calculated by the following formula:

Mean return of category i: ∑ returns

Ni

So the summation of all the returns is divided by the number of tennis matches played in a category. With all these mean returns the possibility emerged to compare the categories. Special focus will be given to the first and last category, the extremes. The first category from 0% - 5% is called the heavy or huge underdog category. While the last category from 95% - 100% is called the heavy or huge favorite category. If these categories are not sufficiently large, then these categories switch to 0% - 10% and 90% - 100% respectively.

## Statistical test

In order to test the statistical significance of all the mean returns found and to determine if one of the categories is significantly different from zero the T-Test is used. Therefore first the standard deviation for each probability category is determined. This standard deviation is based on all the mean returns. Due to the higher returns in the lower probability categories there will be a higher standard deviation in the lower probability categories. With this standard deviation the T-statistic is calculated in the following way:

T-statistic = Mean Return

Standard Deviation /

# 

# Results

This chapter contains all the results derived from the data. Firstly all the matches played in the period 2010 – 2012 on the ATP level are discussed. This is done for the period as a whole and thereafter for the years separately. Secondly the four most important tournaments in tennis are investigated, they are called Grand Slam tournaments. Thirdly it is studied if the difference in surfaces has an effect on the favorite longshot bias. Fourthly the effect of home advantage on the favorite longshot bias is examined.

## ATP level

This section contains all the matches played at the ATP (Associations of tennis professionals) tour from 4-1-2010 to 12-11-2012. These matches were the highest level of professional men’s tennis matches of three seasons. In these seasons 7920 matches were played, where bettors could bet on at Bet365. All of these matches are divided according to the probability bookmaker Bet365 prescribes at the start of all the ATP matches played. An important feature of this section is the size of the categories. In contrast to earlier literature (Forrest & McHale, 2007) each category contains a 5% interval. This is possible due to the larger number of heavy underdogs and heavy favorites in this period in tennis history.

In table 1 the data is divided into 20 categories. For all these categories the average returns, based on given odds, are calculated. In this way it is possible to compare the average returns in each category. These category mean returns are then tested on a significant difference of 0. If the mean return is equal for each category, as the market hypothesis would suggest, then the mean return would be the track take. So the mean return would be -0.0645.

Firstly what stands out in table 1 is the pattern that is clearly seen in the mean return column. This pattern indicates that on average the mean returns become less negative, if the odds category becomes higher. Secondly, if the players in a game are huge underdogs (meaning only about a 2.5% chance of winning according to Bet365 before the start of the game) the average return on a 1 euro bet is -0,785. This indicates that on average a bettor receives just 21 eurocents for every euro betted on an underdog. This is low, especially when taking into account that the average take by the bookmaker is 6.45 eurocent per 1 euro betted (indicating a 6.45% take). So the bookmaker sets the odds on heavy underdogs, influenced by betting behavior, differently compared to what this betting strategy would yield in reality. This is an indication for the over betting of underdogs.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | All years | | | |
| Probability category | N in category | mean return | Standard deviation | t-test |
| 0-0,05 | 135 | -0,785 | 2,487 | -3,669 |
| 0,05-0,1 | 374 | -0,324 | 2,978 | -2,101 |
| 0,1-0,15 | 553 | -0,297 | 2,276 | -3,064 |
| 0,15-0,2 | 505 | -0,143 | 2,091 | -1,533 |
| 0,2-0,25 | 783 | -0,103 | 1,832 | -1,566 |
| 0,25-0,3 | 1086 | -0,142 | 1,557 | -3,001 |
| 0,3-0,35 | 935 | -0,158 | 1,370 | -3,516 |
| 0,35-0,4 | 891 | -0,046 | 1,287 | -1,070 |
| 0,4-0,45 | 1267 | -0,086 | 1,156 | -2,658 |
| 0,45-0.5 | 697 | -0,039 | 1,068 | -0,976 |
| 0,5-0,55 | 803 | -0,097 | 0,964 | -2,849 |
| 0,55-0,6 | 600 | -0,048 | 0,873 | -1,360 |
| 0,6-0,65 | 1126 | -0,060 | 0,795 | -2,520 |
| 0,65-0,7 | 1273 | -0,071 | 0,722 | -3,519 |
| 0,7-0,75 | 1077 | -0,051 | 0,639 | -2,632 |
| 0,75-0,8 | 813 | -0,033 | 0,571 | -1,662 |
| 0,8-0,85 | 922 | -0,036 | 0,497 | -2,188 |
| 0,85-0,9 | 734 | -0,056 | 0,432 | -3,499 |
| 0,9-0,95 | 650 | -0,034 | 0,339 | -2,534 |
| 0,95-1 | 616 | -0,023 | 0,218 | -2,597 |

**Table 1: Including all the matches played in seasons 2010-2012.**

This finding is in line with most literature. For example the research done in horseracing (Ali, 1977) & (Griffith, 1949). But also the findings in the same field (Forrest & McHale, 2007) are comparable.

Thirdly there is a notable finding for the favorite’s category. Before looking at the favorites in tennis over the seasons 2010-2012 it is important to keep in mind that favorites in tennis are players with a pre-match expected winning chance of about 97.5%. Table 1 shows that betting on the category of players with the highest odds prior to the match gave the least negative odds. The mean return for these heavy favorites turned out to be -0.023. This indicates, taking the bookmaker’s take in mind, that betting on the favorite is not a profitable strategy. Meanwhile it is on average the best strategy possible and would generate a positive return if there was no bookmaker’s take. So the mean return of -0.023 indicates that the heavy favorites are under betted by the bettors or undervalued by the bookmakers. This finding is also in line with the research done in horse racing, for example (Ali, 1977) & (McGlothlin, 1956), with the side note that in horse racing huge favorites have a prior match winning probability of about 35% whereas in tennis a huge favorite’s probability is around 97.5%.

The standard deviation showed in the table is the standard deviation of the odds given by the bookies. Therefore the pattern of decreasing standard deviation as the probability is getting higher, as showed in table 1, is not surprising. Because the odds and probability are highly correlated, the low probability categories contain high odds and the high probability categories contain low odds. So if a category contains higher values, it is likely that the standard deviation is bigger.

The t-test shows that there is no significant positive mean return. Indeed the t-test column shows that for more than half of the categories it is significantly less than zero.

Given the above findings for all the matches played in these three seasons, table 2 shows the results for each year separately. Using the same method for table 1 it is possible to determine if the same conclusions apply to the results for each year separately. Firstly the pattern of less negative returns for the higher probability categories still emerges, but less pronounced than in table 1. In 2010 betting on underdogs was on average a worse strategy than betting on a favorite, because the great negative mean returns are generated by the low probability categories. This also holds for the years 2011 and 2012. So the pattern exists in all the 3 years and indicates a consistent pattern over time.

Secondly the over betting of the underdogs occurs for each separate year. In 2010 and 2011 there was no win for an underdog with a 0.0-0.5 winning probability in 42 matches. In 2012 there was only 1 win in 68 matches. Because of the low number of observations in the lowest category, the probability category of 0.05-0.1 is also taken into account for heavy underdogs[[2]](#footnote-2). But also for this category the mean returns are very negative, indicating an over betting of the underdogs.

Thirdly the heavy favorite group (category 0.95-1.0) is in every year one of the best performing categories. It is the only category that performs this well over the three years. When taking the bookies take into account, this category performs better than the take and would have generated a positive mean return. This could indicate an under betting of the bettors or too high odds prescribed by the bookies.

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | 2010 | | | | 2011 | | | | 2012 | | | |
| probability category | N | mean return | St. dev. | t-test | N | mean return | St. dev. | t-test | N | mean return | St. dev. | t-test |
| 0,0-0,05 | 21 | -1,000 | 0,000 | 0 | 21 | -1,000 | 0,000 | 0 | 68 | -0,574 | 3,491 | -1,355 |
| 0,05-0,1 | 97 | -0,515 | 2,338 | -2,171 | 115 | -0,557 | 2,382 | -2,506 | 162 | -0,043 | 3,610 | -0,152 |
| 0,1-0,15 | 199 | -0,239 | 2,293 | -1,469 | 169 | -0,275 | 2,343 | -1,526 | 185 | -0,378 | 2,193 | -2,347 |
| 0,15-0,2 | 156 | -0,192 | 2,053 | -1,170 | 188 | -0,059 | 2,166 | -0,370 | 161 | -0,193 | 2,033 | -1,202 |
| 0,2-0,25 | 274 | -0,032 | 1,892 | -0,283 | 265 | -0,176 | 1,778 | -1,614 | 244 | -0,101 | 1,819 | -0,869 |
| 0,25-0,3 | 366 | -0,162 | 1,551 | -1,997 | 376 | -0,121 | 1,572 | -1,488 | 344 | -0,144 | 1,547 | -1,722 |
| 0,3-0,35 | 310 | -0,134 | 1,388 | -1,701 | 320 | -0,227 | 1,330 | -3,053 | 305 | -0,109 | 1,391 | -1,364 |
| 0,35-0,4 | 302 | -0,082 | 1,276 | -1,121 | 308 | -0,082 | 1,277 | -1,126 | 281 | 0,032 | 1,307 | 0,410 |
| 0,4-0,45 | 449 | -0,076 | 1,158 | -1,384 | 410 | -0,028 | 1,168 | -0,491 | 408 | -0,156 | 1,138 | -2,775 |
| 0,45-0.5 | 235 | -0,043 | 1,065 | -0,622 | 235 | -0,042 | 1,067 | -0,602 | 227 | -0,033 | 1,071 | -0,465 |
| 0,5-0,55 | 288 | -0,063 | 0,965 | -1,111 | 263 | -0,187 | 0,952 | -3,189 | 252 | -0,041 | 0,968 | -0,674 |
| 0,55-0,6 | 210 | -0,089 | 0,878 | -1,462 | 198 | 0,048 | 0,855 | 0,795 | 192 | -0,104 | 0,878 | -1,649 |
| 0,6-0,65 | 394 | -0,049 | 0,791 | -1,224 | 371 | -0,074 | 0,798 | -1,782 | 361 | -0,057 | 0,796 | -1,363 |
| 0,65-0,7 | 434 | -0,080 | 0,725 | -2,303 | 427 | -0,081 | 0,723 | -2,312 | 412 | -0,052 | 0,718 | -1,466 |
| 0,7-0,75 | 360 | -0,046 | 0,637 | -1,375 | 369 | -0,036 | 0,633 | -1,090 | 348 | -0,073 | 0,647 | -2,098 |
| 0,75-0,8 | 274 | -0,011 | 0,558 | -0,319 | 285 | -0,012 | 0,559 | -0,372 | 254 | -0,081 | 0,595 | -2,172 |
| 0,8-0,85 | 281 | -0,054 | 0,514 | -1,747 | 292 | -0,026 | 0,490 | -0,914 | 349 | -0,029 | 0,487 | -1,131 |
| 0,85-0,9 | 266 | -0,065 | 0,439 | -2,402 | 266 | -0,069 | 0,443 | -2,525 | 202 | -0,027 | 0,406 | -0,955 |
| 0,9-0,95 | 219 | -0,069 | 0,378 | -2,715 | 210 | -0,033 | 0,340 | -1,391 | 221 | 0,001 | 0,289 | 0,037 |
| 0,95-1,0 | 179 | -0,015 | 0,199 | -0,973 | 207 | -0,026 | 0,220 | -1,673 | 230 | -0,027 | 0,229 | -1,762 |

**Table 2: all the ATP tennis matches played per season.**

Fourthly again the t-test showed that there are no significant positive categories available. Whereas in each year categories that generate a positive mean return occur. The probability category of 0.55-0.6 generated a mean return of 0.048 in 2011. The positive return in this probability category did not appear in the other years. Also in 2012 categories with positive mean returns can be distinguished. These were the categories 0.35-0.4 and 0.9-0.95, generating a positive return of 0.032 and 0.001 respectively. But again these categories did not generate a positive return over the other years examined.

In summary, considering all the ATP matches played per year approximately gives the same results as the three years together. So the results found are generally persistent over years. Again as discussed above these results are inter alia, in line with the results found by (Ali, 1977), (Winter & Kukuk, 2006) & (Forrest & McHale, 2007).

## Grand Slam tournaments

This section contains all the grand slam tournaments played in the seasons 2010, 2011 and 2012. The Grand slam tournaments are the four biggest and most important tournaments played every year. As a tennis player there is a big chance that you have the dream to be the number 1 in the world or to win a Grand Slam. Every tennis player is mostly remembered by his achievements in the Grand Slams. Therefore every high ranked player has these Grand Slams as the main goal in a season and the training program will be aimed at peaking at these tournaments. Therefore, chances of big upsets in these tournaments are probably smaller than in other tournaments. And when the bookies and the bettors do not keep this in mind the favorite longshot bias will be more pronounced, because the chance that a heavy favorite will be beaten by a heavy underdog is smaller. Another characteristic of a Grand Slam tournament is the direct admission of the 104 highest ranked players in the world. Therefore the chance that one of the top players meets a low ranked player is high. So the number of matches played by a huge underdog and a heavy favorite are numerous. Also as discussed in the ‘explanations’ chapter according to (Williams & Paton, 1998) the favorite longshot bias would diminish when the tournaments are more high profile, because then the insider’s information would be less. So investigating the Grand Slam tournaments would make the favorite longshot bias less pronounced.

In table 3 the results are shown. Firstly in table 3 the pattern that the lower the odds category, the higher the chance on a bigger negative return is still visible. Vice versa the chance on a less negative (or even positive) mean return is bigger for a high probability category.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| probability category | N in category | wins in category | mean return | Standard deviation | t-test |
| 0,0-0,05 | 89 | 1 | -0,674 | 3,057 | -2,081 |
| 0,05-0,1 | 165 | 6 | -0,545 | 2,355 | -2,975 |
| 0,1-0,15 | 169 | 11 | -0,470 | 2,030 | -3,013 |
| 0,15-0,2 | 139 | 14 | -0,388 | 1,832 | -2,500 |
| 0,2-0,25 | 157 | 32 | -0,047 | 1,888 | -0,311 |
| 0,25-0,3 | 198 | 52 | -0,020 | 1,646 | -0,173 |
| 0,3-0,35 | 144 | 36 | -0,235 | 1,328 | -2,121 |
| 0,35-0,4 | 108 | 37 | -0,083 | 1,271 | -0,680 |
| 0,4-0,45 | 165 | 59 | -0,148 | 1,144 | -1,662 |
| 0,45-0.5 | 97 | 38 | -0,160 | 1,047 | -1,503 |
| 0,5-0,55 | 107 | 47 | -0,152 | 0,959 | -1,641 |
| 0,55-0,6 | 75 | 45 | 0,058 | 0,864 | 0,577 |
| 0,6-0,65 | 154 | 96 | 0,007 | 0,783 | 0,108 |
| 0,65-0,7 | 172 | 110 | -0,048 | 0,716 | -0,873 |
| 0,7-0,75 | 143 | 105 | 0,010 | 0,608 | 0,201 |
| 0,75-0,8 | 132 | 98 | -0,033 | 0,570 | -0,675 |
| 0,8-0,85 | 177 | 133 | -0,083 | 0,528 | -2,098 |
| 0,85-0,9 | 165 | 142 | -0,020 | 0,395 | -0,635 |
| 0,9-0,95 | 192 | 175 | -0,012 | 0,308 | -0,557 |
| 0,95-1,0 | 294 | 284 | -0,013 | 0,186 | -1,207 |

**Table 3: All the Grand Slam matches played in the years 2010-2012.**

Secondly when looking at the heavy underdog category of 0.0-0.05, there is a mean return of -0.0674. So despite of 1 win by a heavy underdog, which would result in a return of 29 euro’s per 1 euro betted, this does not generate a positive return, but a highly negative one. Indicating over betting in Grand Slam tournaments by the bettors or/and the bookies setting the odds too low for the underdog categories.

Thirdly investigating the huge favorite category the mean return of -0.013 is one of the highest, when looking at all the categories. Also taking the bookies take in mind, it would be a positive mean return. So in Grand Slam events the favorites are under betted by bettors or/and the odds derived by the bookies are too high for favorites. The second and third conclusion show that there still exists a favorite longshot bias in Grand Slam tournaments. This is not in line with the findings of (Williams & Paton, 1998), who suggested that the favorite longshot bias would diminish as the importance of tournaments would increase.

Fourthly there are some positive categories in the Grand Slam tournaments. These are the probability categories 0.55-0.6, 0.6-0.65 and 0.7-0.75. These probability categories generate a mean return of 0.058, 0.007 and 0.01 respectively. None of these categories is significantly positively different from zero. (Forrest & McHale, 2007) also investigated Grand Slam tournaments, but only found a positive return for the probability category of 0.8-0.9. So the categories that generate a positive return do not match with previous literature and this paper. On the other hand they also found a favorite longshot bias for the Grand Slam tournaments.

## Surfaces

Tennis is the only sport which is played on different surfaces throughout the year. So this is a special feature for this sport. In this paper four surfaces are distinguished, namely clay, grass, indoor hard-court and outdoor hard-court. Each of these surfaces have their own characteristics. Clay is a slow surface where the ball has a high bounce, therefore suitable for quick and defending players. While grass has the characteristic to be very fast with a low ball bounce which is better suited for players who like to attack. This while hard-court is characterized by uniform ball bounces. Ideal for players who like to hit the ball flat and hard. Most of the players have a surface they prefer and some of them are real specialists. Now the most difficult part of all these different surfaces is the switch between them. As a tennis player you have to get used to the different ball bounces and different way of movement between different surfaces. The most difficult switch is from the clay to the grass, because there is almost no time to practice on grass. This lack of time is because there are almost no grass courts in the world. So most of the surprises occur on grass. Table 4 shows if the bookies take all this into account.

Firstly when looking at the mean returns of all the matches played on a surface, playing on grass is the least negative category. Playing on grass generates on average a 50% less negative mean return than playing on clay, hard-court indoor or outdoor. Also when taking the 0.0645 take of the bookies in mind a mean return of -0.049 can be positive if the take would not exist.

Secondly the surfaces clay, indoor and outdoor hard-court generate a favorite longshot bias pattern. On all these three surfaces betting the huge favorite generates the best mean return of all categories, while vice versa the heavy underdog generates the worst mean return possible.

So the bookmakers set the odds for the heavy favorites too high and for the heavy underdogs too low. This could be caused by the bookies being biased or by the bettors being influenced by the odds and therefore over betting the underdogs and under betting the favorites. These results are in line with the previous literature, for example (Winter & Kukuk, 2006) and (Griffith, 1949).

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | grass | | | | clay | | | | Indoor hard-court | | | | Outdoor hard-court | | | |
| odds category | N | mean return | St. dev. | t-test | N | mean return | St. dev. | t-test | N | mean return | St. dev. | t-test | N | mean return | St. dev. | t-test |
| 0,0-0,1 | 72 | 0,125 | 4,460 | 0,238 | 154 | -0,649 | 2,188 | -3,683 | 89 | -0,427 | 3,821 | -1,054 | 227 | -0,577 | 2,354 | -3,693 |
| 0,1-0,2 | 129 | 0,066 | 2,539 | 0,295 | 305 | -0,241 | 2,201 | -1,912 | 176 | -0,509 | 1,683 | -4,010 | 448 | -0,182 | 2,239 | -1,720 |
| 0,2-0,3 | 213 | 0,090 | 1,822 | 0,722 | 576 | -0,204 | 1,636 | -2,998 | 375 | -0,110 | 1,690 | -1,261 | 705 | -0,134 | 1,654 | -2,151 |
| 0,3-0,4 | 221 | -0,216 | 1,274 | -2,521 | 596 | -0,087 | 1,341 | -1,577 | 357 | -0,055 | 1,360 | -0,760 | 652 | -0,107 | 1,324 | -2,056 |
| 0,4-0,5 | 200 | -0,112 | 1,114 | -1,420 | 628 | -0,083 | 1,120 | -1,851 | 355 | 0,011 | 1,140 | 0,190 | 781 | -0,085 | 1,125 | -2,119 |
| 0,5-0,6 | 137 | -0,081 | 0,929 | -1,016 | 435 | -0,079 | 0,922 | -1,790 | 277 | -0,084 | 0,929 | -1,505 | 554 | -0,069 | 0,928 | -1,747 |
| 0,6-0,7 | 258 | -0,044 | 0,751 | -0,932 | 770 | -0,057 | 0,754 | -2,091 | 431 | -0,091 | 0,761 | -2,493 | 940 | -0,068 | 0,760 | -2,728 |
| 0,7-0,8 | 206 | 0,003 | 0,591 | 0,068 | 598 | -0,039 | 0,609 | -1,561 | 394 | -0,083 | 0,626 | -2,642 | 692 | -0,039 | 0,608 | -1,673 |
| 0,8-0,9 | 199 | -0,109 | 0,511 | -3,012 | 515 | -0,037 | 0,463 | -1,806 | 317 | -0,041 | 0,470 | -1,568 | 625 | -0,032 | 0,459 | -1,759 |
| 0,9-1,0 | 171 | -0,070 | 0,340 | -2,705 | 369 | -0,015 | 0,262 | -1,101 | 160 | 0,010 | 0,248 | 0,499 | 566 | -0,035 | 0,292 | -2,866 |
| total | 1806 | -0,049 |  |  | 4946 | -0,105 |  |  | 2931 | -0,099 |  |  | 6190 | -0,099 |  |  |

**Table 4: All ATP matches played in the period 2010-2012 divided by surface played**

Thirdly, in contrast to the other surfaces, grass does not generate the favorite longshot bias. Table 4 shows that the underdogs with a prior match winning probability of 30% or lower generated a positive mean return. This suggests that underdogs playing a match on grass are under betted by the bettors or/and the odds set by the bookies are toohigh. When looking at the favorites there is no under or over betting. A mean return of -0.07 is almost equal to the bookies take. So the matches played on grass exhibit a reverse longshot bias.

Something confirmed by (Woodland & Woodland, 1994) for the baseball betting market but contradicted by the results found overall for tennis, at the Grand Slams and on surfaces clay, indoor hard-court, outdoor hard-court and previous literature as (Ali, 1977).

Fourthly all the positive mean returns found are not significantly different from zero. Also there is no other literature studying tennis matches on grass surfaces[[3]](#footnote-3).

## Home advantage

In many sports the advantage of playing at the home court is researched. This is done for sports played in teams. Literature found that the favorite longshot bias exists for home court teams, for example (Schnytzer & Weinberg, 2008). In contrast (Woodland & Woodland, 2001) found a reverse favorite longshot bias when focusing on home playing teams in the hockey betting market. Previous study was done for home teams, but tennis is a sport that is only played by individuals or a doubles team. In tennis a player only represents his country and not a club, as in team sports. Therefore when testing for home advantage in tennis, this means playing at your home country. When a tennis player plays in his home country, it is very likely that the player gets the support of the public. Can underdogs lift their game by this tremendous support or do favorites still on average generate a higher return? Table 5 examines if a subsample of tennis players playing at their home country displays the favorite longshot bias.

Firstly the pattern as described earlier, where the chance on a better mean return is higher for the higher odds categories, does not exist for the subsample of home players. The chance to find a positive mean return is as high for an underdog as for a favorite. The mean return of the total subsample is with -0.033 the least negative found in this thesis. Also four categories would have generated a positive mean return, for the 0.2-0.3 category even a 12.5% profit.

Secondly when looking at the heavy underdog category (odds of 0.0-0.1) the mean return of -0.247 is the lowest found in table 5. This indicates that either the bettors over betted the heavy home playing underdog and/or the bookies set the odds too low for this category.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| odds category | N | mean return | Standard deviation | t-test |
| 0,0-0,1 | 85 | -0,247 | 3,060 | -0,744 |
| 0,1-0,2 | 142 | -0,028 | 2,446 | -0,137 |
| 0,2-0,3 | 256 | 0,125 | 1,852 | 1,076 |
| 0,3-0,4 | 217 | -0,202 | 1,281 | -2,323 |
| 0,4-0,5 | 230 | 0,052 | 1,142 | 0,693 |
| 0,5-0,6 | 165 | -0,142 | 0,932 | -1,954 |
| 0,6-0,7 | 236 | -0,037 | 0,752 | -0,748 |
| 0,7-0,8 | 211 | -0,033 | 0,604 | -0,787 |
| 0,8-0,9 | 203 | 0,006 | 0,421 | 0,187 |
| 0,9-1,0 | 136 | 0,005 | 0,236 | 0,226 |
| total | 1881 | -0,033 |  |  |

**Table 5: A subsample of players playing at their home country.**

Thirdly the heavy favorite category is again high in comparison with the bookies take of 6.45%. But it is not shockingly high when compared to the other categories in table 5. While it generates a positive mean return it is an under betting of the home favorite and/or biased odds setting of bookies. No earlier literature researched the home advantage in tennis. In other sports, for example (Colquitt, Godwin, & Swidler, 2004), a favorite longshot appears when focusing at home teams.

# Conclusion

In this thesis the ATP tennis matches played in the seasons 2010, 2011 and 2012 are examined. The market investigated is a fixed-odds betting market, where the bookmakers set the odds during a certain period prior to a match. A special feature of the Bet365 data used is the possibility for Bet365 to adjust the odds in response to the betting behavior. In some way the betting behavior is therefore captured and the prior match odds, used in this paper, can not deviate too much from the odds suggested by the behavior of bettors (as would have been the case in a pari-mutual setting). Due to the high number of matches played between a huge underdog and a heavy favorite, the huge underdog category comprises the players with a prior match winning chance of less than 5%. Whereas the heavy favorites are the tennis players with a prior match winning chance of 95% or higher. In this thesis the results discussed above showed that for all the tennis matches played in the seasons 2010 - 2012, a favorite longshot bias occurs. Even when looking at each year separately. Finding a favorite longshot bias is in line with the early horseracing literature like (Ali, 1977) and in line with earlier investigations in tennis like (Forrest & McHale, 2007), but in contrast to other tennis related research such as (Cain, Law, & Peel, 2003). This contrasting result by (Cain, Law, & Peel, 2003) is explained by the surface the tennis matches are played on. In this thesis distinguishing the data by surface generated a favorite longshot bias for the surfaces clay, indoor hard-court and outdoor hard-court. Whereas the matches played on grass showed the opposite and even exhibited a reverse longshot bias. So the matches played on grass generate a different outcome than the other surfaces. Also betting on grass matches overall generated a - 5% return, while the other three surfaces showed an overall return per surface of -10%. The subsample of players playing in the country of their nationality generated the favorite longshot bias. Especially the longshots gave a bad mean return, but 4 of the 10 categories generated a positive return. Also when looking at the overall performance of this subsample, a – 3.3% return is displayed, which is even better than betting on grass games. The Grand Slam tournaments subsample provided evidence that even in the most important tournaments the favorite longshot bias emerged. While literature as (Williams & Paton, 1998) suggested that in the most important tournaments, where the inside information is the lowest, a favorite longshot bias should not be present. As discussed there are categories that gave a positive mean return, but none of these positive categories were significantly different from zero. Further study has to prove if this is an accident or an acceptable strategy.

# Discussion

In this paper the favorite longshot bias is found for the tennis betting market in the years 2010, 2011 and 2012. This finding is the same as found in horse racings, for example (Ali, 1977)& (Griffith, 1949) who also found a favorite longshot bias. Also in soccer the favorite longshot bias was found, according to (Andrikogiannopoulou & Papakonstantinou, 2011). For individual sports the literature is scarce, but (Forrest & McHale, 2007) examined the tennis betting market and found a favorite longshot bias. Other authors found a reverse longshot bias for other sportswhere the favorites are over betted and the underdogs under betted. For example (Woodland & Woodland, 2001) found this reverse longshot bias on the hockey betting market. Now the question is to which extend are all these findings comparable? There are several betting markets, which could be a cause for different findings. The literature showed that almost everybody who investigated a pari-mutual betting mechanism found a favorite longshot bias, while contrasting literature mainly focused on a fixed odds betting market. Whereas for example (Shin, 1991) tried to explain why a favorite longshot bias would appear in a fixed odds betting market, several authors found it was a reverse longshot bias.

Therefore different points of view are needed to identify the causes of the favorite longshot bias. Firstly, comparing different sports and even different years in each sport gives different underdogs and favorites, odds wise. For example in horseracing the difference in level between the best and worst horse that compete each other is totally different than in tennis. For example (Ali, 1977) talked about favorite horses, as horses with an objective winning probability of 35%. This is different from the favorites in this tennis related thesis, where favorites have a 99% chance of winning. Also the odds of the underdogs and favorites in hockey, examined by (Woodland & Woodland, 2001), do not match the odds of (Ali, 1977) nor this thesis. Woodland & Woodland searched for the favorite longshot bias by dividing their data in just three equal subgroups, so the underdogs group exhibited matches of odds lower than 1.4 (probability of 0.714)[[4]](#footnote-4). This stands in no relation to the odds used for an underdog in this thesis or the horseracing literature. (Cain, Law, & Peel, 2003) investigated if 8 different sports betting markets exhibited the favorite longshot bias. They showed that horse races, baseball and soccer do not have matches with a team or player with a winning chance of higher than 80%. Baseball did not even have underdogs with a winning chance smaller than 20%. Again in baseball a reverse favorite longshot bias is found by (Woodland & Woodland, 1994). When looking at the underdogs, an explanation for the over betting is the risk loving utility function of bettors, according to (Friedman & Savage, 1948) & (Quandt, 1986). But when is betting on an underdog risky? In this thesis it relates to a 2.5% chance of winning 14 times your bet. In contrast to (Woodland & Woodland, 1994) where betting an underdog relates to 20% chance of winning 5 times your bet. Sometimes these differences for odds given to favorites and underdogs relate to the existence of dominant teams or individuals. For example in this thesis there are more heavy underdog and heavy favorite matches as in contrast to (Forrest & McHale, 2007). So the teams or individuals dominating their sport influence the numbers in a category. Secondly this thesis showed evidence that matches played on the grass surface generated positive returns for the underdog categories. This surface has very special characteristics, because the possibility to train on it is limited, due to the limited availability of grass courts. Also the grass season only lasts for three weeks, so immediate form is required and experience crucial. According to the results of this thesis the bookies and bettors do not accurately process the uncertainties involved with the grass surface in the odds. The appendix contains a table for all the 2013 played grass matches and again betting on this surface generates a high overall return of -1.8%. The heavy underdog category in seasons 2010 till 2012 generated a 12.5% return, whereas in 2013 this would have generated a 163.6% return. Absurdly high. The two other underdog categories did not generate a positive return in 2013, as they did in 2010-2012. Further research on this would be very interesting and might be profitable.

# Bibliography

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

Andrikogiannopoulou, A., & Papakonstantinou, F. (2011). Market Efficiency and Behavioral Biases in the Sports Betting Market.

Arsch, P., & Malkiel, B. G. (1982). Racetrack Betting and Informed Behavior. *Journal of Financial economics*, 187-194.

Cain, M., Law, D., & Peel, D. (2000). The Favorite-longshot Bias and Market Effiency in UK Football Betting. *Scottich journal of Political Economy*, 25-36.

Coleman, L. (2004). New light on the longshot bias. *Applied Economics*, 315-326.

Colquitt, L. L., Godwin, N. H., & Swidler, S. (2004). Betting on Long shots in NCAA basketball games and implications for skew loving behavior. *Finance research letters*, 119-124.

Dixon, M. J., & Pope, P. (2004). The value of statistical forecasts in the UK association. *international journal of forecasting*, 697-711.

Dowie, J. (1976). On the Efficiency and Equity of Betting Market. *Economica*, 139-150.

Forrest, D., & Mchale, I. (2007). Anyone for Tennis (Betting)? *The European Journal of Finance*, 751-768.

Friedman, M., & Savage, L. (1948). The Utility Analysis of Choices Involving Risk. *Journal of Political Economy*, Milton Friedman and L. J. Savage , Vol. 56, No. 4 (Aug., 1948), pp. 279-304.

Gandar, J. M., Zuber, R. A., & Johnson, R. S. (2001). Searching for the favourite- longshot bias down under: an examination of the New Zealand pari-mutuel betting market. *Applied Economics*, 1621-1629.

Griffith, R. (1949). Odds adjustment by american horse-race betters. *The American Journal of Psychology*, 290-294.

Hausch, D. B., Ziemba, W. T., & Rubinstein, M. (1981). Efficiency of the Market for Racetrack Betting. *Management Science* , 1435-1452.

Hodges, S., Tompkins, R., & Ziemba, W. T. (2003). The favorite/Long-Shot bias in S&P 500 and Ftse 100 index futures options: The return to bets and the cost of insurance. *EFA Annual conference paper*, 1-20.

Isaacs, R. (1953). Optimal Horse Race Bets. *The American Mathematical Monthly*, 310-315 .

Law, D., Cain, M., & Peel, D. (2003). The favourite-longshot bias, bookmaker margins and insider trading in a variety of betting markets. *Bulletin of Economic Research*, 263-273.

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

Paul, R. ,., & Weinbach, A. P. (2005). Bettor Misperceptions in the NBA : The Overbetting of Large Favorites and the ''Hot Hand''. *Journal of Sports Economics*, 390-400.

Quandt, R. E. (1986). Betting and Equilibrium. *The Quarterly Journal of Economics*, 201-208.

Rubinstein, M. (1985). Nonparametric Tests of Alternative Option Pricing Models Using All Reported Trades and Quotes on the 30 Most Active CBOE Option Classes from August 23, 1976 through August 31, 1978. *the journal of finance*, 455-480.

Schnytzer, A., & Weinberg, G. (2008). Football Fixed-Odds and Point Spread Betting Markets Testing for Home Team and Favorite Biases in the Australian Rules. *Journal of Sports Economics* , 173-190.

Shin, H. S. (1991). Optimal Betting Odds Against Insider Traders. *The Economic Journal*, 1179-1185 .

Steckler, H., Sendor, D., & Verlander, R. (2010). Issues in sports forecasting. *The journal of forecasting*, 606-621.

Stephen, F. (1979). Subjective Information and Market Efficiency in a Betting Market. *Journal of Political Economy*, 75-88.

Tversky, A., & Kahneman, d. (1979). Prospect Theory: An Analysis of Decision under Risk. *Econometrica*, 263-292.

Vlastakis, N., Dotsis, G., & Markellos, R. N. (2008). How efficient is the European football beting market? Evidence from arbitrage and trading strategies. *Journal of Forecasting*, 426–444.

Walls, W. D., & Busche, K. (2003). Broken odds and the favourite-longshot bias in parimutuel betting: a direct test. *Applied Economics*, 311-314.

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

Williams, L. V. (1999). Information efficiency in betting markets: A Survey. *Bulletin of Economic Research*, 1-30.

Williams, L. V., & Paton, D. (1998). Why are some favourite-longshot biases positive and others negative? *Applied Economics*, Leighton Vaughan Williams & David Paton (1998) Why negative?, Applied Economics, 30:11, 1505-1510.

Winter, S., & Kukuk, M. (2006). Risk Love and the Favorite-Longshot Bias: evidence from German harness horse racing. *Schmalenbach Business Review*, Stefan Winter/Martin Kukuk\* Risk Love and the FavoRite-Longshot Bias: evidence FRom geRman haRness hoRse Racing 349-364.

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

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

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

Zuber, R. A., Gandar, J. M., & Bowers, B. D. (1985). Beating the Spread: Testing the Efficiency of the Gambling Market for National Football League Games. *Journal of Political Economy*, 800-806.

# Appendix

* In table 6 all the tennis matches played at the ATP level are divided among the year they are played in and in which prior match probability category the matches belonged to. This is equal to the table 2, except the probability interval is bigger. So table 6 has the same format as showed in (Forrest & Mchale, Anyone for Tennis (Betting)?, 2007). In (Forrest & Mchale, Anyone for Tennis (Betting)?, 2007) as in this thesis the favorite longshot bias is found for every year investigated.
* The thesis outcomes for matches played on grass exhibited no favorite longshot bias. As the grass season of 2013 has come to an end, the matches played are investigated according to the method used in this thesis. In table 7 the results show that the heavy underdog category generates a positive 160 % return. This is higher than the mean return over the period 2010-2012. The categories 10% - 20% and 20% - 30% did generate a positive return when the period 2010 – 2012 was used, but when looking at table 7 these returns are not positive for the year 2013. A strategy, based on the period 2010 – 2012, of betting the lowest three categories would generate a positive return of 9% in 2010 – 2012. Performing this strategy for the period 2013 would generate a positive 12.7% return. The heavy underdog category is even significantly different from zero, indicating this is not just based on luck. So interesting to continue to follow the tennis matches played on grass.
* In tables 8 and 9 the results are shown for the ladies tennis matches played in the period 2010 – 2012[[5]](#footnote-5). These women’s tennis matches are played at the WTA level. The WTA stands for the Women’s Tennis Association and represents the highest level of competing in women’s tennis. So all the women’s tennis matches are processed according to the method discussed in the ‘Method’ chapter. When the heavy underdog category (0% - 5% probability of winning) of table 8 is investigated, the small amount of observations and the high return stands out. The return in the heavy underdog category is 119% and based on this return a favorite longshot bias is absent in women’s tennis. The 21 observations are in contrast to the 135 observations of heavy underdogs matches played in the men’s tennis sport in the period 2010 – 2012. Therefore table 9 has been made with bigger 10% intervals for each category. When in table 9 the same data is investigated and focused on the heavy underdog category (0% - 10%), there is a – 56% return. This leads again to the question in the ‘Discussion’ chapter, namely how to divide the categories? Furthermore the favorite longshot bias does exists in table 9. Where the favorites are under and the underdogs over betted.

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | 2010 | | | | 2011 | | | | 2012 | | | |
| odds category | N | mean probability | St. dev. | t-test | N | mean probability | St. dev. | t-test | N | mean probability | St. dev. | t-test |
| 0,0-0,1 | 118 | -0,602 | 2,128 | -3,072 | 136 | -0,625 | 2,023 | -3,603 | 230 | -0,200 | 3,583 | -0,847 |
| 0,1-0,2 | 355 | -0,218 | 2,190 | -1,878 | 357 | -0,161 | 2,254 | -1,350 | 346 | -0,292 | 2,122 | -2,559 |
| 0,2-0,3 | 640 | -0,106 | 1,707 | -1,578 | 641 | -0,144 | 1,660 | -2,190 | 588 | -0,126 | 1,666 | -1,835 |
| 0,3-0,4 | 612 | -0,109 | 1,334 | -2,013 | 628 | -0,156 | 1,306 | -2,989 | 586 | -0,041 | 1,353 | -0,738 |
| 0,4-0,5 | 684 | -0,065 | 1,127 | -1,497 | 645 | -0,033 | 1,132 | -0,747 | 635 | -0,112 | 1,116 | -2,534 |
| 0,5-0,6 | 498 | -0,074 | 0,929 | -1,775 | 461 | -0,086 | 0,919 | -2,011 | 444 | -0,068 | 0,930 | -1,551 |
| 0,6-0,7 | 828 | -0,065 | 0,757 | -2,477 | 798 | -0,078 | 0,759 | -2,889 | 773 | -0,054 | 0,755 | -1,999 |
| 0,7-0,8 | 634 | -0,031 | 0,604 | -1,286 | 654 | -0,026 | 0,602 | -1,089 | 602 | -0,076 | 0,626 | -2,991 |
| 0,8-0,9 | 547 | -0,059 | 0,479 | -2,878 | 558 | -0,046 | 0,469 | -2,339 | 551 | -0,029 | 0,459 | -1,467 |
| 0,9-1,0 | 398 | -0,045 | 0,312 | -2,859 | 417 | -0,029 | 0,287 | -2,075 | 451 | -0,013 | 0,261 | -1,076 |

**Table 6: All the ATP matches played in the years 2010, 2011 and 2012 with a probability interval of 10%.**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| odds category | N | mean probability | Standard deviation | t-test |
| 0,0-0,1 | 22 | 1,636 | 2,863 | 2,681 |
| 0,1-0,2 | 47 | -0,223 | 2,191 | -0,699 |
| 0,2-0,3 | 52 | -0,195 | 1,678 | -0,838 |
| 0,3-0,4 | 65 | -0,000 | 1,332 | -0,003 |
| 0,4-0,5 | 58 | -0,120 | 1,126 | -0,814 |
| 0,5-0,6 | 44 | -0,07 | 0,926 | -0,501 |
| 0,6-0,7 | 63 | -0,064 | 0,757 | -0,674 |
| 0,7-0,8 | 70 | -0,106 | 0,611 | -1,454 |
| 0,8-0,9 | 51 | -0,049 | 0,469 | -0,740 |
| 0,9-1,0 | 56 | -0,014 | 0,286 | -0,360 |
| total | 528 | -0,018 |  |  |

**Table 7: the matches played at the ATP level for all the tennis matches played on grass for 2013**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| odds category | N | mean return | Standard deviation | t-test |
| 0-0,05 | 21 | 1,190 | 6,752 | 0,808 |
| 0,05-0,1 | 247 | -0,709 | 1,863 | -5,978 |
| 0,1-0,15 | 525 | -0,128 | 2,497 | -1,171 |
| 0,15-0,2 | 455 | -0,175 | 2,065 | -1,805 |
| 0,2-0,25 | 731 | -0,019 | 1,887 | -0,274 |
| 0,25-0,3 | 1049 | -0,110 | 1,581 | -2,262 |
| 0,3-0,35 | 832 | 0,000 | 1,448 | -0,006 |
| 0,35-0,4 | 889 | -0,088 | 1,275 | -2,063 |
| 0,4-0,45 | 1169 | -0,061 | 1,159 | -1,797 |
| 0,45-0.5 | 634 | -0,084 | 1,061 | -1,984 |
| 0,5-0,55 | 964 | -0,088 | 0,957 | -2,841 |
| 0,55-0,6 | 548 | -0,052 | 0,865 | -1,416 |
| 0,6-0,65 | 1042 | -0,063 | 0,796 | -2,536 |
| 0,65-0,7 | 1134 | -0,085 | 0,725 | -3,942 |
| 0,7-0,75 | 995 | -0,068 | 0,648 | -3,334 |
| 0,75-0,8 | 799 | -0,046 | 0,579 | -2,266 |
| 0,8-0,85 | 907 | -0,063 | 0,516 | -3,665 |
| 0,85-0,9 | 666 | -0,065 | 0,441 | -3,806 |
| 0,9-0,95 | 613 | -0,053 | 0,362 | -3,648 |
| 0,95-1 | 376 | -0,006 | 0,196 | -0,642 |
| total | 14596 | -0,078 |  |  |

**Table 8: All WTA matches played in the years 2010, 2011 and 2012 with a probability interval of 5%**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| odds category | N | mean return | Standard deviation | t-test |
| 0,0-0,1 | 268 | -0,560 | 2,651 | -3,456 |
| 0,1-0,2 | 980 | -0,149 | 2,306 | -2,029 |
| 0,2-0,3 | 1780 | -0,073 | 1,714 | -1,795 |
| 0,3-0,4 | 1721 | -0,046 | 1,362 | -1,392 |
| 0,4-0,5 | 1803 | -0,069 | 1,126 | -2,599 |
| 0,5-0,6 | 1512 | -0,075 | 0,925 | -3,144 |
| 0,6-0,7 | 2176 | -0,074 | 0,760 | -4,553 |
| 0,7-0,8 | 1794 | -0,059 | 0,618 | -4,018 |
| 0,8-0,9 | 1573 | -0,064 | 0,486 | -5,205 |
| 0,9-1,0 | 989 | -0,036 | 0,310 | -3,600 |
| total | 14596 | -0,078 |  |  |

**Table 9: All WTA matches played in the years 2010, 2011 and 2012 with a probability interval of 10%.**

1. http://www.bet365.com/extra/en/promotions/horse-racing/best-odds-guaranteed/ [↑](#footnote-ref-1)
2. The Appendix contains a table where the three years are showed with a 0.1 probability category interval. In this table the amount of observations per category are divided more equally. [↑](#footnote-ref-2)
3. Therefore in the appendix is table 7 added for all the grass matches played for the season 2013. [↑](#footnote-ref-3)
4. Different definitions of what an underdog is can give different results. An example is shown in the ‘Appendix’ for the WTA matches played in Women’s tennis. [↑](#footnote-ref-4)
5. For more results in ladies tennis please contact [julien\_vantleven@hotmail.nl](mailto:julien_vantleven@hotmail.nl). [↑](#footnote-ref-5)