UNIVERSITEIT GENT

FACULTEIT ECONOMIE EN BEDRIJFSKUNDE

GHENT UNIVERSITY

FACULTY OF ECONOMICS AND BUSINESS ADMINISTRATION

ACADEMIC YEAR 2013 – 2014

The forecast ability of the dispersion of bookmaker odds

Master thesis submitted to obtain the degree of

Master of Science in
Applied Economics: Business Engineering

Kwinten Derave

Under the guidance of

Prof. Dr. Michael Frömmel and Martien Lamers

UNIVERSITEIT GENT

FACULTEIT ECONOMIE EN BEDRIJFSKUNDE

GHENT UNIVERSITY

FACULTY OF ECONOMICS AND BUSINESS ADMINISTRATION

ACADEMIC YEAR 2013 – 2014

The forecast ability of the dispersion of bookmaker odds

Master thesis submitted to obtain the degree of

Master of Science in
Applied Economics: Business Engineering

Kwinten Derave

Under the guidance of

Prof. Dr. Michael Frömmel and Martien Lamers

PERMISSION

I declare that the contents of this thesis may be consulted and/or reproduced, provided the source is acknowledged.

Kwinten Derave

Acknowledgements

The reason for choosing “The forecast ability of the dispersion of bookmaker odds” as a topic for this thesis is two-sided. First of all, market efficiency and the related Efficient Market Hypothesis (EMH) is a trending finance topic. My first profound contact with the EMH was in the investment analysis courses, given by Dr. Prof. Frömmel. Especially the idea that it might be impossible to beat the market by adopting an active investment policy, since all information is already incorporated in the prices, has made a strong impression on me.

The second reason for choosing this topic is my genuine interest in the betting market. In the past, I sometimes made a bet with my friends on football matches. Back then, I was always wondering whether a special strategy would exist that generates a certain profit, with some fruitless attempts as result. Now, a couple of years later, I know the reason why my attempts were fruitless. Two words: Market Efficiency.

First of all, I would like to express my gratitude to my promoter, Prof. Dr. Frömmel to provide me with the opportunity to examine this interesting subject. Furthermore, I appreciate the guidance of Martien Lamers the past two years, and his quick and accurate responses to my questions.

This work was only possible thanks to the help of my father Rudi Derave, my nephew Matthias De Maertelaere, and my friends Greet Bontinck, Dominique Van der Straeten, Philippe Peelman, Bo Zenner, Fé Zenner and Rembrand Zenner. Thank you so much for proofreading my master thesis!

Table of contents

Abbreviations

EMH = Efficient Market Hypothesis
FLB = Favorite-Longshot Bias
IP = Implied Probability

List of Tables

List of Figures

Chapter 1: General introduction

In financial markets, investors could benefit from a good estimation of the future value of a company. They could invest in stocks of these companies, and accurate value estimation could yield a profit through dividend payouts and/or stock value appreciation.

If the financial stock market were to be efficient, these stock prices would fully reflect the underlying value of the company with respect to the future. However, evidence shows that it is very difficult to make accurate stock price estimations. The main problem to make accurate stock price estimations is that the future is uncertain. Therefore, inefficiencies occur in the stock market. Research has shown that the more stock analysts disagree about the future value of a company, the lower the future returns for investors will be. This is referred to as the *dispersion effect*.

However, it is difficult to measure the extent of this *dispersion effect* in the stock market. Stocks can be held infinitely long. This implies that there is no fixed termination point where feedback can be obtained about the true value of a stock. Since it is very difficult to accurately estimate the true value of a company, other markets have been sought in order to test the hypothesis of market efficiency. This is where the betting market enters the story.

The sports betting market can be seen as a similar, but simplified, version of the financial stock market. Both markets are similar to each other because they share several very important features. First, both investors and bettors have the same objective: maximize profits. Secondly, a transaction takes place between two participants, leading to a zero-sum game. Bettors can accept bets from bookmakers. The conditions of the bet are expressed by the *bookmaker odds*. These bookmaker odds can be compared to stock prices in the financial market (Levitt, 2004).

It is simplified because in contrast to the stock market, the true value of a bet is known at a well defined point in time: the date at which the sports event takes place. Thereby, quick feedback is provided every time a sport event takes place. Given that the betting market is more conducive to investigation and analysis, this paper will investigate the dispersion effect in the *tennis betting market*.

Several different bookmakers operate in the tennis betting market. These bookmakers all place odds when a tennis event takes place. Although the odds are placed for the same sports event, differences in odd placement may occur across bookmakers. This is referred to as the *bookmaker dispersion.* The analysis of the dispersion of bookmaker odds is new in the literature.

Firstly, a comparison will be made between the male and female tennis betting market. This is an extension to previous research conducted in the tennis betting market, where a comparison was made between Grand-Slam and non-Grand Slam tennis games, and between early-round and late-round tennis games. It will familiarize the reader with a very important concept in the betting market, namely the *favorite longshot bias*.

What is the forecast ability of the dispersion of bookmaker odds?

The main subject of this paper will be approached from two different angles. The first approach is based upon the expectation that the differences in placed odds across bookmakers may contain valuable information. We will investigate whether this information can be used to improve the predictability of tennis match outcomes.

The second approach establishes a link between the financial market and the betting market. Through a unique method, this paper will investigate for the betting market whether higher bookmaker dispersion will lead to lower future returns for the bettors.

Finally, a betting strategy is explained to the reader. This betting strategy is based upon new findings in the research that was conducted in this paper.

This paper consists of two main parts: Part I, review of the relevant literature and Part II, the conducted empirical research. The literature related to the subject is divided into two chapters. Chapter 2 concerns the financial market, and starts with a description of the efficient market hypothesis. Furthermore, several market inefficiencies will be described, in particular the dispersion effect. Chapter 3 describes the betting market and it starts with an enumeration of inefficiencies in this market. Moreover, this chapter contains a summary of important forecasting methods. Finally, the tennis betting market in particular will be described. Part II, the empirical research is structured as followed: In chapter 5 the data sources are described, as well as a description of the final dataset that has been used in this paper. Chapter 6 explains important core concepts that will be used throughout this paper.

Chapter 7 formulates the theoretical predictions and hypotheses. Chapter 8 analyzes the proposed hypotheses and discusses the results. In chapter 9, a betting strategy will be explained. Chapter 10 makes recommendations for further research and chapter 11 summarizes the conclusions.

PART I: LITERATURE REVIEW

Chapter 2: The financial market

2.1 The Efficient Market Hypothesis

In 1970, Eugene F. Fama wrote *efficient capital markets, a review of theory and empirical work*. "*The efficient market model is the hypothesis that security prices at any point* ***fully reflect*** *all available information*"(Fama, 1970). Fama described three forms of financial market efficiency.

The first form of market efficiency is referred to as *weak-form* market efficiency. Whenever a market is *weak-form* efficient, prices on traded assets reflect all past information. In this case, it makes no sense to base a buy-hold-sell strategy upon the analysis of historical prices since no excess returns can be achieved. The weak-form market efficiency is the easiest form to test, and therefore this form has been tested often. Statistical studies have shown mixed results about *weak-form* efficiency. Robust anomalies were found, such as the *momentum effect* and the *winner-loser reversal*. Other anomalies (i.e. *holiday effect*, *weekend effect*) tend to disappear after the discovery.

Secondly, markets could be perceived as *semi-strong* efficient. When a market is *semi-strong* efficient, all information available to the public is included in the current prices of assets. In this case, no excess returns can be achieved based on information known to the public. *Semi-strong* form market efficiency is more difficult to test than the *weak-form* market efficiency. Based on event studies, one could test the impact of stock splits, corporate merger announcements and annual reports, and information contained in publications and analyst reports.

Finally, markets can be also be perceived as *strong-form* efficient. This is the most powerful and interesting form. This type of market form implies that, beside from past and public information, private information is also incorporated in the prices. When markets are *strong-form* efficient, excess returns can never be achieved in the long run.

As stated earlier, it is very difficult to test the efficiency of the market. The following part of this paper is based upon the book of Frömmel (2011): “Portfolios and investments”, where both a theoretical and an empirical problem are discussed.
Firstly, a theoretical problem, which is known as the information paradox (Grossman and Stiglitz, 1980), occurs. In case the market is perfectly efficient, or *strong-form* efficient, rational investors would not benefit from analyzing the markets, since potential anomalies that could be profitable, are already incorporated in the prices. The paradox is that, in order to keep the markets efficient, analyzes by investors are necessary. Fama (1991) argues that: "*Prices reflect information to the point where the marginal benefits of acting on information do not exceed the marginal costs*" (p.1575). This implies that a market will become more efficient if the required information to price the assets is cheaper.
Secondly, an empirical problem, known as the *joint hypothesis* problem, is important to take into account. In order to test market efficiency, a model to value current prices is needed to calculate possible abnormal returns. However, in case abnormal returns occur, it is difficult to verify whether the valuation model is correct and whether the markets are inefficient, or vice versa. As Fama (1991) states: “*Market efficiency per se is not testable* “. (p.1575)

2.2 Inefficiency in the financial market: The Dispersion Effect

According to Miller (1977):

> “*A key assumption of the standard capital asset model is what Sharp calls homothetic expectations[…] However, it is implausible to assume that although the future is very uncertain, and forecast are very difficult to make, that somehow everyone makes identical estimates of the return and risk of every security*” (p.1151).

Using Miller’s price optimism model, we can conclude that more stock prices will be biased upwards if the disagreement about or dispersion of the current price of a stock is more significant. Investors who are optimistic about the future return of a stock will buy this stock. This will increase the stock price. On the other hand, investors who are pessimistic about the future return of a stock might not act due to high short sell costs and risks. This restricts the decline of the stock. Subsequently, the prices are overrated and this implies that these high stock prices will lead to lower future returns.

There are many different proxies to measure the disagreement among investors.

Diether, Malloy, Scherbina*, (2002)* have used the dispersion in analysts' forecasts as a proxy for differences in opinion about a stock. They provide evidence that stocks with high dispersion in analysts' forecasts earn significantly lower future returns than otherwise similar stocks.

Another tool to measure the disagreement among investors is 'the breadth of mutual fund ownership' (Chen et al. 2001). They state: "When *breadth is low-i.e., when few investors have long positions—this signals that the short sales constraint is binding tightly, and that prices are high relative to fundamentals. Thus reductions in breadth should forecast lower returns* " (p.171).

Wu (2006) generalized Tauchen and Pitts' (1983) Mixture of Distribution Hypothesis, which links asset volume and volatility in a way that derives a proxy for divergence of opinion among all individuals. Also, a higher trading volume (Lee and Swaminathan's, 2000) can be used as proxy for differences in opinion. All these proxies strongly support Millers' prediction which states that more divergence in opinion will result in upwardly biased prices and lower future returns.

Chapter 3: The betting market

3.1 Advantages and efficiency of the betting market

Betting markets contain some very attractive features to test market efficiency.

Thaler & Ziemba (1988) state:

> "*Economists have given great attention to stock markets in their efforts to test the concept of market efficiency, yet wagering markets are, in one key respect, better suited for testing efficiency and rationality. The advantage of wagering markets is that each asset (bet) has a well-defined termination point at which its value becomes certain*" (p.162).

Thus, the characteristics of a betting market clearly differ from those of a stock market, where a stock can be held infinitely long. Due to the quick and repeated feedback, the correctness of the odds can be checked immediately after the termination of the event. This well-defined termination point tempers the joint-hypothesis problem. Testing market efficiency by using betting markets has several other important advantages. Betting markets are characterized by the presence of a large number of bettors. Information sources are cheap, and widely available. Those characteristics reduce the information paradox.

Peel & Law (2002) argue: *"As a consequence, the problems that arise in evaluating future dividends are mitigated"* (p.327). Consequently, betting markets should be more efficient than financial markets.

All three forms of market efficiency (weak, semi-strong and strong) have been tested in the betting market.

For example, Hausch, Ziemba & Rubinstein (1981) have investigated weak-form market efficiency. They have successfully developed a profitable betting system for the horse race betting market, based on technical analysis, and thus not confirming weak-form efficiency.

Figlewski (1979) has tested whether betting markets are semi-strong efficient. They have incorporated predictions of professional handicappers which were known to the public. They found that they contain valuable information, but that this information cannot be used to significantly improve the forecast ability of the market odds.

Shin (1991, 1992, 1993) investigated the impact of insider trading. When markets are strong-form efficient, private or insider information should also be incorporated in the odds. According to Schnytzer, Lamers, and Makropoulou (2010): *"It is agreed that the extent of insider trading in the market is what makes bookmakers' odds deviate from winning probabilities…"* (p.537), thus refuting the strong-form efficiency hypothesis.

3.2 Inefficiencies in the betting market

3.2.1 The favorite-longshot bias (FLB)

The most robust anomaly is the favorite-longshot bias. Thaler and Ziemba (1988) defined the favorite-longshot bias as follows: "*The expected returns per dollar bet increase monotonically with the probability of the horse winning*" (p.163).

This anomaly is a strong violation of the EMH (efficient market hypothesis). Bettors tend to overestimate the winning chances of the highly quoted odds (the longshots), and equally underestimate the winning probabilities of the favorites.

In the literature, three possible explanations are described. Firstly, this anomaly could be explained in terms of risk-loving gamblers who prefer to bet on odds that have a higher return. Shin (1991,1992,1993) has an alternate explanation. He argues that bookmakers provoke the favorite-longshot bias in order to raise enough revenue from outsiders to pay the winnings of the insiders. The last explanation is that bettors overestimate winning probabilities of longshots and as a consequence, bookmakers take advantage of this psychological bias. The favorite longshot bias will be explained in greater depth throughout his paper.

3.2.2 Herding behavior

Another form of market inefficiency is herding behavior. This kind of behavior is closely related to the phenomenon of insider information.

Froot, Scharfstein & Stein (1992) have already observed this effect in the financial market. Their results display that the existence of short-term speculators can lead to herding behavior.

They call it a positive information spillover: "*As more speculators study a given piece of information, more of that disseminates into the market, and therefore, the profits from learning that information early increases*"(p.1478). Betting markets similarly have a short-trading horizon. The effect of herding behavior on these markets has also been tested.

Schnytzer and Shilony, (1995) have shown that bets to plungers (i.e. horses that exhibit large decreases in the odds against winning in the betting auction) at early odds generate superior returns to bets on non-movers.

Peel and Law (2001) found market inefficiencies as well because of insider trading and herding behavior. They state: "*Horses whose odds plunge will have lower rates of return measured at starting odds when the Shin measure of insider activity falls during the auction than when it rises* "(p.330).

3.2.3 The fixed-odds betting market

Inefficiencies were not only discovered in horse betting markets. Cain, Law & Peel (2000) found that the individual fixed-odd betting market in UK-football exhibits the same favorite-longshot bias as the favorite-longshot bias found in the horseracing betting market.
The difference with horse betting is the fact that the public can bet on football games long before the match is played (i.e. a week in advance), while bets on horse races can only be placed within a shorter time-frame, for example 30 minutes before the race takes places.

Makropoulou and Markellos (2007) have compared the odds of a traditional, offline bookmaker, with those of an online bookmaker. Unlike for horse betting, the risk for online bookmakers no longer comes from insiders, but from informed or expert bettors. They state: "*Indeed, one would expect that in sports such as football, it is somewhat unlikely that there is significant private information relative to public information given that matches are reported widely in the media and are strictly regulated*" (p.520).
Casual bettors only bet randomly, while informed bettors use information that becomes publicly available after the declaration of the odds. While the odds from online bookmakers can be adjusted after their publication, the odds from the traditional bookmaker stay fixed. Makropoulou and Markellos argue that this increased uncertainty for the traditional bookmaker evolves simultaneously with increased margins on the odds.

Vlastakis, Dotsis and Markellos (2009) have also investigated the efficiency of the European football betting market. They state the following: "*Informational efficiency requires that the market aggregates information from different sources, so that prices represent the best forecasts on the outcome of future events. This implies that no bettor or bookmaker can sustain returns that exceed transaction costs/margins" (p.428).*

However, they found that the European football betting market, which is one of the most liquid betting markets, is not weak-form efficient, because there was one arbitrage opportunity per 200 played games.

3.3 Forecasting match results

Many stakeholders could benefit from a good prediction of a sports match outcome. Fans, sponsors, team coaches and last but not least, gamblers and bookmakers. If a gambler can consistently predict the outcome of a match more accurately than his bookmaker, long-term profits can be made. Forecasting methods can be grouped in two main methods, namely expert evaluation and statistical models.

Expert evaluations are made by so-called 'tipsters'. They are experts in betting predictions, with a specific expertise in the forecasting of sports games outcomes. Tipsters are often independent specialists who publish their forecasts in specialized magazines and on websites. Despite their specialization, empirical evidence shows (Andersson et al, 2005) those tipsters hardly beat the predictions of somebody with only limited knowledge.

In the second group of forecasting methods statistical models are used. Depending on the variables that enter the statistical models, different statistical predictions exist. Some variables are related to past performance in football matches (Goddard and Asimakopoulos, 2004), while other variables relate to the information about the number of goals scored (Goddard, 2005). According to Spinn and Skiera (2008), there exist two outstanding methods, namely prediction markets and bookmaker odds. These variables aggregate the combined expectation of all the participants.

Prediction markets are online markets (websites) where participants can trade virtual stocks related to the result of sport games. The pay-off of these stocks depends on the result of the event, thus reflecting the expectation of the outcome. The second outstanding variable to predict a match outcome are odds placed by bookmakers. By reversing bookmaker odds, the implied probability is obtained, which can be seen as the chance to a specific outcome of a sports match outcome. Spann and Skiera (2008) have compared the forecast ability of the prediction market, bookmaker odds and tipster for matches in Germany's premier soccer league. They have provided evidence that both a prediction market and the odds of the betting market give the best forecasts with a hit rate[1] of approximately 54%. Tipsters had the worst hit rates. (42.6%)

3.4 The tennis betting market

The tennis betting market provides some attractive theoretical features compared to other betting markets. Firstly, the tennis betting market does not suffer from the so-called *'last race of the day'* effect. Horse and greyhound races are studied many times on its efficiency in the past (Schnytzer and Shilony, 1995; Thalor and Ziemba, 1988). Bets are often accepted by gamblers watching live in the stadium, for example the hippodrome of Ostend or Waregem in Belgium. At the end of such an event, bettors who are not pleased with their current balance state (i.e. negative), and want to go home with an acceptable profit may place their final bet on risky long-shots. This may cause a bias of the odds that will not occur in the tennis betting market. On average, 4.650 tennis matches are played every year. These matches are spread out quite well over each year. Thus, one can expect that the extent of the 'last race of the day" effect is mitigated.

Another bias that is mitigated in the tennis betting market is the sentiment bias. (Forrest and McHale, 2007). "*In contrast to team sports, individual players seldom attract committed fans whose allegiance may be reflected in betting volumes*" (p.8) .One can expect that in an individual sport such as tennis, this bias will be mitigated.

Finally, the decimal odds of the tennis betting market are widely spread, ranging from 1,01 to 60. This is due to the fact that tennis tournaments are structured in such a way that in the first round of each tournament, the well ranked player faces lower ranked players. The average WTA rank difference between the two female opponents in the first round of a Grand-Slam tennis matches amounts up to 80 points, whereas it only averages 7,5 for tournament finales. This large discrepancy in player strength is reflected by a large odd spread between the two opponents.

In order to improve the forecast ability of tennis match outcomes in particular, several variables have been taken into account. Klaassen and Magnuns (2003) proposed a model

[1] The hit rate is the percentage of correct predictions.

based upon the difference in ranking between the two opponents. Corral and Prieto-Rodriguez (2010) have further extended this model by adding three more variables, namely the players' past performance, the player's physical characteristics and match characteristics.

Forrest and McHale (2007) have published 'anyone *for tennis (betting)?*' in which they have studied the favorite-longshot bias and the bettor's attitude to risk and skewness. They found a positive bias throughout the range of odds, indicating that the favorite-longshot bias is present in the tennis betting market. A comparison was made between non Grand-Slam and Grand-Slam tournaments. They had expected that the extent of the FLB would be mitigated in the Grand-Slam subsample, but they did not find any evidence for this expectation.

They however did find some evidence of profitable trading opportunities, which indicates a violation of *weak form efficiency*. It should be noted that Forrest and McHale worked with a rather modest dataset of 8.500 tennis matches.

Andrew Zeelte (2012) conducted a master thesis study, "*Explaining the favorite-longshot bias in tennis: An endogenous Expectations Approach".*

The paper used a larger dataset of approximately 22.676 matches or 45.352 bets. Zeelte provided evidence of the presence of the FLB in the tennis betting market using 4 different methods. Similar to Forrest and McHale's (2007) investigation, Zeelte (2012) tested the extent of the favorite-longshot bias between early-round and late-round matches. Zeelte (2012) found reasonable evidence to suggest that to some extent, the FLB is mitigated in late round tennis matches.

Lahvicka (2013) discussed the causes of the FLB in the tennis betting market, using a dataset of almost 45.000 matches. He found that the FLB is much more pronounced in matches between lower-ranked player, in late-round tennis matches and in high profile tournaments.

PART II: EMPERICAL RESEARCH

Chapter 4: Introduction

Summarizing fore-going literature review, it can be said that there are several inefficiencies in the financial stock market. One of these inefficiencies is the dispersion effect. Due to a theoretical and an empirical problem, it is very difficult to measure the extent of this anomaly. The betting market provides some very attractive features to test market efficiency on. However, inefficiencies occur even in the betting market. The most robust anomaly is the favorite-longshot bias.

The empirical research conducted and discussed upon in Part II is a combination of the two main subjects described in the literature review. The extent of the dispersion effect in the financial stock market will be tested for the tennis betting market.

Part II is structured as follows. First, the data source will be described. Then, in chapter 6, some important core concepts will explain, such as odds, and bookmaker dispersion. Chapter 7 formulates three testable hypotheses. The first hypothesis will compare the extent of the FLB between male and female tennis matches. This is a continuation on previous research on odds of the tennis betting market. Although this is not the main subject of this thesis, a sufficient knowledge of the favorite-longshot bias is important for a good understanding of the remainder of this thesis. The second and third hypotheses are the main subject of this thesis, and deals with bookmaker dispersion in the tennis betting market. The second hypothesis will focuses on the information that might be comprised within this bookmaker dispersion. Therefore an explanation will be provided of the mechanisms that causes bookmaker dispersion. The third hypothesis provides the link between the financial market and the betting market with respect to the dispersion effect. Chapter 8 analyzes and interprets the results. Chapter 9 proposes a betting strategy. This betting strategy is a result of the findings made through the whole paper. Chapter 10 provides recommendations for further research and chapter 11 summarizes the findings.

Chapter 5: Sample and data sources

In this paper, data from the website < http://www.tennis-data.co.uk> will be used. All data are well-structured in .cvs format, and they can be used within standard spreadsheet applications and statistical programs.

The dataset contains tennis data of both male and female tennis matches, ranging from the year 2006 to 2013. The data includes the tennis matches of the 4 Grand-Slam tournaments, 13 Master tournaments and approximately 133 international tennis tournaments. This leads to a dataset of approximately 38.840 tennis matches or twice as many (77.680) single bets. From this dataset, 3.362 matches were excluded due to walkovers, where one of the opponents did not show up, or retirements where an injury or illness prematurely ended a started game. The decision for this exclusion is based upon Zeelte's (2012) work. He provided evidence that inclusion of non-finished tennis matches causes a bias. Furthermore, since this paper investigates the dispersion of bookmakers' odds in particular, only matches with odds of all 5 bookmakers are included in the final dataset. This led to another exclusion of 5.258 tennis matches, and brings the final sample size on 34.529 matches or 69.058 bets.

Chapter 6: Important concepts and variables

6.1 Win

Winning is the outcome of a tennis match in the perspective of the player.

The variable equals 1 if a player wins and 0 if a player loses, and is thus a binary variable.

Imagine a tennis match between Roger Federer and Raphael Nadal in the first round of the US Open. If Federer wins the match, his value for win is 1. Nadal then loses the tennis match. His value for win is 0.

win = 0/1

6.2 Odds

Odds are placed by bookmakers, and can be accepted by bettors. In this paper, odds will be expressed as decimal odds, also known as European odds, digital odds or continental odds. A decimal odd is the potential winning amount of the bettor when he places a single euro bet. When continuing our example. Imagine that Unibet, a reputable bookmaker, offers odds for the tennis match between Federer and Nadal. Unibet offers an odd ate the price of €1,60 for Federer and €2,20 for Nadal. The results are summarized in **table 1**. When a bettor accepts the bet of Federer and gambles €1,00, the better has a potential payout of €1,60. When a bettor accepts the bet of Nadal, and he bets €1,00, the potential payout for the bettor amounts €2,20

Table 1 : Example bookmakers

	Roger Federer	Raphael Nadal
UNIBET	€1,60	€2,20
BET365	€1,20	€2,10

The data used in this research contains for each tennis game the odds of five different bookmakers. Even though they cover the odds of the same player, differences may occur across the quoted odds. In this paper, *AverageOdd* is defined as the average of the 5 different odds placed by the bookmakers for the same bet.

AverageOdd = (odd1+odd2+odd3+odd4+odd5)/5

6.3 Return

The return can only be calculated after the tennis match has been played, since it depends on the outcome of the tennis match. It is the net profit or loss the bettor incurs when accepting a single euro bet. Continuing with our example: whenever a bettor accepts a €1,00 bet on Federer, and Federer wins the tennis match, his return is €1,60 -€1,00 = €0,60. When Federer loses, his return is €-1,00.

Return = Win*AverageOdd -1

6.4 Implied probability

Odds can range from 1 to theoretically infinity. A more intuitive way to express the chance of an event occurring is the implied probability. The implied probability is the reciprocal of the odd. Its range is limited between 0 and 1, and can be considered an estimation of the chances of a specific outcome to a tennis match.

ImpliedProbability = 1/AverageOdd

6.5 Bookmaker dispersion

Bookmaker dispersion is a proxy for the magnitude of the disagreement about the outcome of a tennis match between the 5 different bookmaker odds. A further expansion of the example should help to clarify this concept. As assumed, Unibet offered decimal odds of €1,60 and €2,20 for the tennis match between respectively Federer and Nadal. Now suppose that another bookmaker BET365 offers odds for the same tennis match. The odds offered by BET36 are €1,20 for Federer and €2,10 for Nadal.

The odds of UNIBET and Bet365 are both summarized in **table 1** above. As can be noticed, UNIBET and BET365 almost agree in odds for Nadal. The odds are close to each other. This indicates that the bookmaker dispersion is low. The odds of Federer on the other hand, strongly differ between the two bookmakers (€1,60 and €1,20). When bookmakers place largely different odds, it means that the bookmaker dispersion is high.

In this research, the bookmaker dispersion is measured by the coefficient of variation of the odds of the 5 different bookmakers. The coefficient of variation is a normalized measure for dispersion, and is defined as the ratio of the standard deviation of the 5 bookmaker odds, to the mean of the 5 bookmaker odds.

Bookmaker Dispersion = standard deviation (odd1; odd2; odd3; odd4; odd5) / AverageOdd

Chapter 7: Theoretical predictions and hypotheses

7.1 The extent of the favorite-longshot bias between male and female tennis matches

In the betting market, the favorite-longshot bias is an observed phenomenon where on average, bettors tend to overvalue "long-shots" and undervalue favorites[2]. This is the most robust anomaly found in the betting market. The extent of the FLB in the tennis betting market has been investigated by Forrest and McHale (2007) and Zeelte (2012).

High profile tennis matches are tennis matches where the participating players are more famous, where the pricing money is higher and where the tennis matches get more media attention. Grand-Slam matches are considered more high profile than non Grand-Slam matches. Late-round tennis matches are considered more high profile than early-round tennis matches. Both Forrest and McHale (2007) and Zeelte (2012) argued that the extent of the favorite-longshot bias should be mitigated in high-profile tennis matches compared to low-profile tennis matches. This expectation is based upon two explanations that cause the favorite-longshot bias (supra p.6).

1. T*he FLB exists in the tennis betting market because bettors make bad estimations of the probabilities of a tennis match outcome.*

As a consequence, they overestimate the chances of the longshots and underestimate the chances of the favorites. Since the tennis players are considered more famous in high-profile tennis matches, one can argue that the bettors can make more accurate predictions. Furthermore, since pricing money is higher, the incentives to win are also higher. As Forrest and McHale (2007) stated:

> "*With incentives to effort are greater, there is less likelihood that a star player will be beaten because he is, for example, treating the event as practice or underperform because he does not want to risk aggravating an injury*" (p.22).

[2] Source: wikipedia

This higher predictability should decrease the extent of the FLB for high-profile tennis matches.

2. The FLB is present in the tennis betting market because of a defensive reaction of bookmakers against bettors with private information (Forrest & McHale, 2007).

One can argue that private information is less available in high-profile tennis matches. This is because the players are considered more famous, which implies they have more to lose when they leak private information or lose on purpose. Also this should mitigate the extent of the FLB in high-profile tennis matches.

Similar as for Grand-Slam tournaments and late-round tennis matches, one can argue that the extent of the FLB should be mitigated for male tennis matches compared to female tennis matches.

First, similar to Grand-Slam tournaments, the average duration of male tennis matches is longer than the duration of female tennis matches. The rules of tennis prescribe that men have to win three sets before winning the match, while women only need two sets to win the match. This best of 5 concepts for male tennis matches could also reduce the factor of luck, since male tennis players have more time to prove their best. Furthermore, it decreases the influence of random factors such as arbitration error. (Forrest & McHale, 2007). This should lead to a higher predictability of male tennis match outcomes.

Secondly, there is a discrepancy in rewards between male and female tennis matches. The top 100 male tennis players on average earned $305.345 in the year 2012, compared to an average of $249.682 earned by the top 100 female tennis players. (Chron, 2012) One can argue that that the incentive for males to win a match is bigger than for female players, leading to less private information and a more predictable tennis match outcome in male tennis matches.

Finally, there is a distinction in notoriety between male and female tennis players. The average count of fans on the social network "Facebook" of the top 10 female tennis players in April 2014 was approximately 1,8 million. This average amounted 2,5 million for the top 10 male tennis players. Furthermore, taking the viewing figures of Wimbledon's finale (Source: The Guardian, 2009) in 2009 on BBC into account, we see that the male finale between Roger Federer and Andy Roddick reached a peak of 11 million viewers, with an average of 7 million viewers. This is in strong contrast with the number of viewers of the female final between the two Williams sisters in 2009, which reached a peak of only 4 million viewers, with an average of 3,5 million viewers. Based upon these numbers, it seems convincing that male tennis players generally are more famous than female tennis players.

Because male tennis matches should be more predictable, with less private information available, one can argue that this leads to a more efficient market, were the extent of the FLB is mitigated.

Hypothesis I: The extent of the favorite-longshot bias is mitigated in the male tennis betting market

7.2. Bookmaker Dispersion: Forecasting models

In 1970, Fama stated that a market is efficient if the security prices fully reflect all available information. The counterparts for securities in the tennis betting market are the odds placed by bookmakers. If Fama's definition of an efficient market is applied to the betting market, this would imply that the bookmaker odds should fully reflect the probability of a tennis match outcome. For this, a basic model is proposed, using solely the bookmaker odds as explanatory variable. As the dataset contains odds of 5 different bookmakers, the average of the odds will be used.

However, one can expect that a variable that measures the magnitude of these differences across bookmaker odds improves the forecast ability of the basic model, since it might contain valuable additional information. To understand this, a strongly simplified description is being provided of how bookmakers price their odds and why differences may occur between different bookmakers.

Several mechanisms have an influence on the exact pricing of bookmaker odds. It is important to understand that the number one priority of a bookmaker is to maximize profits and not to make accurate forecasts. In contrast to bettors, bookmakers try not to gamble. They try to eliminate risk, and ensure that they make a certain profit no matter what the outcome of a tennis match will be. Bookmakers try to achieve this profit in two ways.

First, bookmakers take a margin (or over-round). Odds placed by bookmakers are unfair, in the sense that bookmakers should profit and bettors should lose on the long run. This bookmaker' margin is calculated through summing the *implied probabilities* of the odds. The average over-round of the tennis data in this paper is approximately 105%. However, this over-round is only a theoretical figure, and does not assure that a bookmaker will make a profit. In fact, this over-round is rarely achieved (Source: Betfair).

A simple example will clarify this structure. Imagine that 100 bettors accept a €1 bet on Federer, quoted at €1,80 while only 10 bettors accept a €1,00 bet on his opponent Nadal, quoted at € 2,20. This means that prior to the tennis match, the bookmaker collects €100,00 + €10,00 = €110. If Nadal wins the tennis match, the bookmaker has to pay back 10*€2,20 = €22,00 to the 10 winning bettors. This leaves him a profit of €110,00- €22,00 = €88,00. However, if Federer wins the tennis match, the bookmaker has a payout of 100*€1,80 = €180,00. This would yield the bookmaker a loss of €110 - €180 = €-70,00. Even though the bookmaker took an over-round of 1/1,8 + 1/2,1 = 103,17%, the bookmaker has still made a loss on this particular tennis match! That is why bookmakers try to spread the risk through attracting enough bettors on both side of a tennis match. This method is referred to as "balancing the books". Thus, a certain profit can be guaranteed, regardless the outcome of a tennis match (Source: Soccerwidow).

In order to balance the books, bookmakers have odd compilers deciding about the exact prices of bookmaker odds. Major bookmakers have a team of odd compilers to do this. The exact quotation of odds is mainly based on a combination of 4 odd calculation sources, namely the competitors 'odds, quantitative information, qualitative information and the public opinion.

First, bookmaker odds are not independent from each other. Bookmakers systematically compare their odds to the odds of the competition. For example, if 4 competing bookmakers of UNIBET decrease their odds of Federer, they probably have a good reason for doing this.

Whenever an odd of a certain bookmaker is higher than the odds of his competitors, large betting volumes can be expected since bettors always search for the most favorable bets. This would cause an unbalanced book. In our example, UNIBET risks a loss on this particular tennis match. That is also the reason why bettors can often bet only a limited amount of money on a certain odd. Bookmakers want to avoid that bettors accept favorable bets in large volumes. A bookmaker can anticipate to actions of competing bookmakers through adjusting their odds in the same direction. Comparison of competitor odds ensures that odds of different bookmakers tend to converge.

Secondly, when bookmakers place their opening odds it is important that they place the initial odds about right. If they fail to do this, they risk that too high volumes are placed on one side of the bet. Moreover, this would result in an unbalanced book and potential losses. Placing the opening odds is a dangerous affair for bookmakers because they have no competitors' odds to compare with. As a result, it will rarely be one bookmaker that places all the opening odds. Also, the margins of opening odds are often higher (Source: Soccerwidow)

When placing opening odds, odd compilers rely on a mix of both qualitative and quantitative analysis. Qualitative information ensures that odd compilers have a personal expectation about the tennis match outcome. They gain insight by reading newspapers, following upcoming events, consulting experts etc. Furthermore, opening odds are also based on quantitative analysis.

Odd compilers estimate the probability of a match outcome upon statistics. Statistical models may for example take into account player rankings, historical confrontations, weather conditions etc. However, different odd prices may occur among bookmakers. The qualitative information available across odd compilers may differ, leading to different odd placements. Furthermore, statistical probability calculations methods can lead to differences in bookmaker odds. One bookmaker might for example emphasize historical confrontations between 2 opponents, whereas others might give more weight to the opponent's rankings.

As soon as one bookmaker has placed his odds, other bookmakers quickly follow by placing similar priced odds. As from then, the odds are matched to the expectations of the public opinion. When overly large betting volumes are placed on one side of the bet compared to the other side, it means the odds are mispriced according to the public. Let us reconsider the example of Federer against Nadal. Federer may suffer a small injury or an illness a couple of days before the start of the tennis match, and this becomes known to the public. This may lead to large betting volumes on one side of the bet, namely on Nadal. Also this may cause an unbalanced book, because when Nadal wins, bookmakers risk a loss on this particular tennis match. Bookmakers anticipate on that trough adjusting the odds so that both sides are equally attractive again. In our example, the concerning bookmaker may anticipate on that risk by lowering the odds of Nadal and making the odd price more appealing to bet on Federer. Thereby, betting volumes increase for Federer and decline for Nadal, in order that both sides are balanced again and certain profits can be made.

Differences in odds between bookmakers may occur when well informed bettors place large amounts of money on certain tennis players' odds. Bettors normally don't bet on every available bookmaker, but more likely have a few preferred bookmakers. Consequently, differences in betting volumes may occur across bookmakers. This can lead to differences in the placed odds. One can expect that volatility in the market with respect to public expectations, will lead to higher bookmaker dispersion.

Even though this explanation is a strong simplification of the reality, it is clear that many factors precede the final valuation of closing odds. These factors may contain information that can be used to increase the forecast ability a of tennis match outcome. Therefore, an extended model will be proposed that also includes the explanatory variable *bookmaker dispersion*. To determine whether the variable *bookmaker dispersion* increases the forecast ability, the basic model (supra p.18) will be compared to this extended model.

Hypothesis 2: Bookmaker dispersion significantly improves a model that forecasts a tennis match outcome.

7.3. Bookmaker Dispersion: Link with the financial stock market

Miller (1977) predicted that stock prices will be biased upward if the disagreement about the current price of a stock is more significant. Subsequently, once the prices are biased upwards, these high stock prices will lead to lower future returns. This was referred to as the *dispersion effect* (supra, p4). In the literature, several proxies have been used to measure the dispersion effect. Diether, Malloy, Scherbina, (2002) have used the dispersion in analysts' forecasts as a proxy for differences in opinion about a stock. Wu (2006) linked asset volume and volatility to the disagreement of investors.

In this paper *dispersion in bookmaker odds* will be used as a proxy for differences in opinion about a tennis match outcome. Based upon the knowledge of Miller (1977), one can argue that a higher dispersion in bookmaker odds will lead to lower future returns for the bettors. This is because there are important parallels between the betting market and the stock market. As mentioned in the introduction, both stock investors and bettors have the same objective: maximizing profit. Secondly, a transaction takes place between two participants, leading to a zero-sum game. Finally, large amounts of money are at stake for both the stock market and the betting market. According to the UK gambling commission, the online betting market had an estimated turnover of £ 84 billion in 2007 (Levitt, 2004). Because of the similarities with the financial stock market, we expect that the dispersion effect is also present in the tennis betting market

Hypothesis 3: Higher bookmaker dispersion leads to significantly lower future return for bettors in the tennis betting market.

Chapter 8: Analysis and results

It is important for the reader to understand the distinction between the three proposed hypotheses while reading the analysis and conclusions. Therefore, a straightforward description of the working methods to test the 3 hypotheses is explained in the introduction of this chapter.

Hypothesis I analyzes the extent of the favorite-longshot bias between 2 subsamples, a male subsample and a female subsample.

Hypothesis II analyzes the forecast ability of bookmaker dispersion. Two forecasting models will be compared to each other:

- A basic model with only one explanatory variable: *ImpliedProbability*

-An extended model with two variables: *ImpliedProbability* and *BookmakerDispersion.*

Hypothesis III also analyzes the forecast ability of bookmaker dispersion, but uses a different approach. Two subsamples will be compared to each other:

- A subsample of bets with high bookmaker dispersion
- A subsample of bets with low bookmaker dispersion.

8.1 Hypothesis I

"Hypothesis I: The extent of the favorite-longshot bias is mitigated in the male tennis betting market".

First, a demonstration is given of the general presence of the favorite-longshot bias in the tennis betting market. To provide evidence of its existence, two different methods will be applied. The first method compares the average return with regard to the implied probability range in table form. The second method conducts a simple linear regression. Next, the extent of the FLB in male and female tennis matches will be compared by applying the same two methods.

8.1.1 The favorite-longshot bias in the tennis betting market

If the favorite-longshot bias (FLB) would not present in the tennis betting market, betting at any odd in the odd range would yield the same negative expected return.

Forrest and McHale (2007) stated: "*The hypothesis that there is no bias and there is strong efficiency, implies that the relationship between odds (implied probability) and yield (return) is linear with a slope of zero*." (p.13)

However, bettors tend to overestimate the winning chances of the highly quoted odds (the long-shots), and equally underestimate the winning probabilities of low quoted odds (the favorites). The presence of the FLB implies that the expected average returns per bet increase monotonically with implied probability.

The results of the first method are presented in **table 2**. The bets are divided into subgroups according to the implied probability with a 0,1 interval. For each implied probability range, the average return is reported. This is the return obtained when betting €1 on all the bets in the regarding implied probability range. An almost monotonic relationship can be noticed between implied probability and the average return. This indicates that the favorite-longshot bias is present in the tennis betting market. When accepting all 1.680 bets in the 0,0-0,1 implied probability range, an average loss of -53,66% is suffered. On the other hand, when accepting all bets on favorites, with an implied probability between 0,9 and 1,0, only a loss of -0,61% is incurred. There seems to exist only one discontinuity in this relationship between profit and the implied probability, namely in the transition of the odd ranges 0,7-0,8 and 0,8-0,9. The average loss increases from -1,69% to -2,71%. However, this discontinuity is statistically not significant[3].

Table 2: General presence of the favorite-longshot bias in the tennis betting market.

Implied Probability	Total return (€)	# Bets	Average return (%)
0,0-0,1	-901,49	1.680	-53,66
0,1-0,2	-1603,59	5.101	-31,44
0,2-0,3	-1208,02	7.475	-16,16
0,3-0,4	-848,61	8.993	-9,44
0,4-0,5	-595,13	9.123	-6,52
0,5-0,6	-452,07	7.578	-5,97
0,6-0,7	-486,28	9.563	-5,09
0,7-0,8	-147,76	8.722	**-1,69**
0,8-0,9	-180,22	6.643	**-2,71**
0,9-1,0	-25,49	4.179	-0,61

The second method to prove the existence of the FLB in the tennis betting market is a simple linear regression, with *implied probability* as independent variable and *return* as a dependent variable. A summary of the regression variables is provided in **appendix 1**[4]. The following linear regression model was employed:

[3] T-test = 1.365

[4] In order to conduct regression analyses in this paper, a subsample was created in which one player was randomly selected for every tennis match. Because every tennis match always has one winner and one loser, this results in a dependent dataset and downwardly biased standard errors (Forrest and McHale, 2007). Approximately 50% of the dataset was excluded to conduct regression analyses. This still leaves a dataset of approximately 34.500 bets.

Return = β1 + β2 (ImpliedProbability)

Since bookmakers take a profit margin, bettors on average lose money. Therefore, the constant term β1 should be negative. Further, if the FLB would not be present in the tennis betting market, the coefficient of the implied probability (β2) should not significantly differ from 0. A positive term however, would indicate that returns increase when the implied probability increases, meaning that the favorite-longshot bias is present in the tennis betting market.

Table 3 summarizes the results of the first linear regression. As expected, the constant term is significantly negative (-0,292). Thus, bettors on average lose money. Similar to the findings of Forrest and McHale (2007) and Zeelte (2012), the coefficient of implied probability significantly exceeds 0.

Table 3: The favorite-longshot bias in the whole market, N = 34.533

DV: Return	Coefficient	Standard Error
Constant	-0,292***	-0,0293
ImpliedProbability	0,372***	0,0152

*** indicate statistical significance at the 1% level

We conclude that there is evidence for the existence of the favorite-longshot bias in the tennis betting market.

8.1.2 Comparison male and female tennis matches

Now, the extent of this favorite-longshot bias will be compared between male and female tennis matches. In total, the data consist out of bets for 39.068 male matches and 29.990 bets for female matches. **Table 4** summarizes the findings in table form. Besides the average returns, the amount of bets in each implied probability range is also presented.

The table provides robust evidence for the FLB in both the male and female tennis betting market. Similar as for the complete dataset (table 2), a higher implied probability results in a higher average returns in both subsamples. If the average returns of the bets on male and female tennis matches are compared, a noticeable difference is visible for bets with an implied probability between 0,0 and 0,1. Accepting all 644 bets on female matches yields an average loss of -61,4%. The male counterpart on the other hand yields an average return of -48,9%, which is a difference of 12,53%. However, this difference is not statistically significant[5]. This is also the case for the other differences in average returns.

Table 4 : Favorite-longshot bias : male and female subsample

Implied Probability	Female		Male		
	Return (a)	**# bets**	**Return (b)**	**# bets**	**Return (a)- (b)**
0,0-0,1	**-**61,40%	644	-48,90%	1.036	**-12,53%**
0,1-0,2	-31,40%	2.225	-31,50%	2.876	**0,08%**
0,2-0,3	-16,00%	3.377	-16,30%	4.098	**0,37%**
0,3-0,4	-8,70%	3.855	-10,00%	5.138	**1,31%**
0,4-0,5	-6,60%	3.921	-6,50%	5.202	**-0,11%**
0,5-0,6	-5,50%	3.331	-6,30%	4.247	**0,81%**
0,6-0,7	-5,70%	4.116	-4,60%	5.447	**-1,14%**
0,7-0,8	-2,60%	3.792	-1,00%	4.930	**-1,52%**
0,8-0,9	-2,20%	2.955	-3,10%	3.688	**0,91%**
0,9-1,0	-0,50%	1.773	-0,70%	2.406	**0,11%**

The second method conducts a linear regression on the male and female subsample: Male subsample,

$$\text{return} = \beta 1 + \boldsymbol{\beta_m}\,(\text{ImpliedProbability})$$

Female subsample,

$$\text{return} = \beta 1 + \boldsymbol{\beta_f}\,(\text{ImpliedProbability})$$

The estimated relationships of both linear regression analyzes are presented in **Table 5.** Again, the FLB is clearly present in both betting markets. The coefficient of the implied probability for the male and female subsample significantly exceeds 0, meaning that a higher implied probability leads to significantly higher returns. On the first sight, it seems that to a certain extent, the favorite-longshot bias is mitigated in the female subsample, since the coefficient of the implied probability is slightly lower than for the male subsample. This is in contrast to the expectation of mitigation of the FLB for the male subsample.

[5] T-test = -0,797, Kolmogorov Smirnov Z = 0,283

Table 5 : The Favorite-Longshot Bias for male and female tennis matches

DV : Return	**Male**		**Female**	
	Coeficient	**Standard Error**	**Coeficient**	**Standard Error**
Constant	-0,2927***	0,0202	-0,2906***	0,0401
Implied Probability	0,3725***	0,0346	0,3721***	0,0231

*** indicate statistical significance at the 1%

To compare the coefficients of the implied probabilities more formally, the null hypothesis that $\mathbf{B_F}$ equals $\mathbf{B_M}$ was tested. B_f is the regression coefficient for the Female subsample, while B_M is the regression coefficient for the Male subsample. A dummy variable *male* is coded, and takes the value 1 for males and 0 for females. Furthermore, an interaction term *Male*ImpliedProbability* is added to the equation.

This results in the following equation:

$$\text{Return} = \beta 1 + \beta_2 \text{(male)} + \beta_3 \text{(ImpliedProbability)} + \beta_4 \text{(male*ImpliedProbability)}$$

Table 6 presents the results of the regression. It appears that both the coefficients *Male* and *Male*ImpliedProbability* are not statistically significant. *Male*ImpliedProbability* has a T-value of 0,9, and a p-value of 0,993.

Table 6: The FLB between male and female tennis matches, interaction term

DV : Return	**Coeficient**	**Standard Error**
Constant	-0,2906***	0,0231
Male	-0,0022	0,0307
ImpliedProbability	0,3721***	0,0399
Male*ImpliedProbability	0,0004	0,0528

*** indicate statistical significance at the 1% level

Evidence is provided that the FLB is present in the tennis betting market. We expected that the extent of the FLB would be mitigated in the male tennis betting market. However, no evidence was found to support this hypothesis. As we conclude that the extent of the FLB is not mitigated for the male subsample. For the remainder of the analysis, the male and female subsample will therefore be used interchangeably.

8.2 Hypothesis II

"Hypothesis 2: Bookmaker dispersion significantly improves a model that forecasts a tennis match outcome."

Two binary logistic regression models will be compared. The basic model only uses the independent variable *ImpliedProbability* to forecast a tennis match outcome. The extended model also uses the independent variable *BookmakerDispersion.* Both models have the same dependent variable with a binary response, namely *win*.

Basic model:

Win = β1 + β2 (ImpliedProbability)

Extended model:

Win = β1 + β2 (ImpliedProbability) + β3(BookmakerDispersion)

For both regression models, a binary logistic regression was conducted[6]. The coefficients of the logistic regression analyses are presented in **table 7**.

Table 7 :binary logistic regression models, N= 34.533

DV: Win	Basic Model	Extended Model
Constant	-2,6852***	-2,3843***
	(0,0343)	(0,0533)
ImpliedProbability	5,0744***	4,7379***
	(0,0605)	(0,0753)
BookmakerDispersion		-2,581***
		(0,3651)

*** indicates statistical significance at the 1% level, Standard Errors are attached in parentheses

[6] Both models are correctly specified and provide a good fit. Furthermore, no evidence of multicollinearity was found between the independent variables ImpliedProbability and BookmakerDispersion

First, the two regression models will be compared to each other through the comparison statistical measures. Afterwards, we will investigate how accurate the two models forecast in practice.

8.2.1 Statistical significance

Table 8: Comparison goodness of Fit

	Basic Model (1)	**Extended Model (2)**	**Difference (2)-(1)**
Likelihood ratio test	32166,13	32198,7	**32,575**
McFadden R²	0,198	0,199	**0,001**
BIC	32202,70	32172,13	**- 22,301**

Based on statistical criteria (**table 8**), the goodness of fit will be compared between the basic model and the extended model. When conducting a simple linear regression, the goodness of fit is generally measured by R². Unfortunately, there is no direct analogue for a logistic regression. Instead, several other methods are used to compare nested logistic regression models. These methods can basically be divided into three groups, namely the log likelihood ratio test, R² analogs, and information measures.

The Log likelihood ratio test compares the log likelihood of the basic model with the log likelihood of the extended model. It is used to assess the contribution of a variable, in this case the explanatory variable *BookmakerDispersion*. Twice this difference in log likelihood between the two models asymptotically follows a chi-square distribution. Given the large dataset of about 35.000 observations, we can assume that the log likelihood ratio approximately follows a chi-square distribution. The likelihood ratio test has a value of 32,575. This exceeds the chi-squared critical value of 5,99. Based upon the likelihood ratio test, we can conclude that the extended model significantly fits better than the basic model. Pseudo R²'s of a logistic regression give an alternative to the R² of a linear regression. McFaden pseudo-R² is perhaps the most popular one (Source: Williams, University of Notre Dame). A difference of 0,001 indicates that only a small improvement in goodness of fit is provided with the inclusion of the variable *BookmakerDispersion.*

The two first described methods are often criticized. Considering the large sample size used in this paper, it is easy to accept more variables because the chi-squared statistics are designed to detect any departure between a model and observed data. Furthermore, caution is recommended when interpreting pseudo R². The heteroscedastic nature of a logistic regression makes it impossible to compute an R² statistic with all the characteristics of R² calculated from a linear regression.

An alternative approach is the usage of information measures. These measures have some advantages compared to pseudo-R²'s and the log likelihood ratio test. First, it has a penalty for the inclusion of variables that do not significantly improve the fit. In the second place, it leads to more adequate results for large datasets.

The Bayesian Information Criterion (BIC) reduces with 22,301. This provides a very strong support that the extended model has a better fit.

We can conclude that based upon statistical measurement, the inclusion of the variable *BookmakerDispersion* significantly improves the goodness of fit with our data. Now, we will test whether this improvement of the extended model also leads to better forecasting results.

8.2.2 Economical significance

The main objective is to compare the forecast ability of the basic model with the forecast ability of the extended model. However, two additional forecasting instruments will be included in the analysis to create a more objective image of the first two forecasting models. Finally, the year 2013 is used to conduct an out-of-sample validation test.

The third instrument is a logic betting strategy. This instrument always predicts the favorite player to win. A player is considered to be favorite[7] in a tennis match, when the players' implied probability is higher than the implied probability of its opponent.

The fourth forecasting instrument is a random betting strategy. It selects for each tennis match a random player, and predicts that this player will win the tennis match. **Table 9** summarizes the in-sample results for the year 2006-2012.

[7] Further in this paper, another definition of "favorite" will be applied.

The forecast ability is measured by the *hit rate*. The hit rate is the count of the matches correctly predicted by the corresponding model, divided by the total amount of predictions It is clear that that the first three models approximately provide the same results.

Notwithstanding the statistical improvement of the extended model compared to the basic model (supra, p.30), the hit rate has not improved.

Table 9 : Forecasting instruments, year 2006-2012

Instrument	Hit Rate	Average Return
Basic Model	71,30%	-3,38%
Extented Model	71,30%	-3,34%
Logic Strategy	71,27%	-3,34%
Random Strategy	50,00%	-9,26%

Apparently, the basic model and the extended model have a hit rate of 71,30%. The extended model only gives negligible better average returns, namely -3,38% compared to -3,34%.

Furthermore, the application of a binary logistic regression models to forecast a tennis match outcome does not seem to have better results than when the logic betting strategy is applied. The logic betting strategy has a hit rate of 71,27%. This similarity in results makes sense, considering the fact that the logic betting strategy model and the basic model have predicted the same tennis match outcome for 99,56%. A random betting strategy has the worst results. It is clear that the first three models outperform the fourth. The random betting strategy has as expected a hit rate of 50%.

In order to test the accurateness of the prediction models in practice, a cross validation test was applied. Tennis data of 2006 until 2012 (59.683 bets) was used as a training sample to create the forecasting models. The year 2013 (9.376 bets) was used as the independent validation sample. The results of the year 2013 are summarized in **table 10**.

The same conclusions can be drawn as for the training sample (table 9) regarding the forecast ability of the basic model and the extended model. Both the hit rate and the rate of return are approximately the same for the validation sample 2013 (70%).

Table 10 : Out-of-sample test, 2013

	YEAR 2013 (a)		
	Basic Model	**Extended Model**	**Logic Strategy**
Total Bets	9.376	9.376	9.376
Accepted bets	4.718	4.738	4.730
Hit Rate (%)	70,26	70,27	70,21
Profit (€)	€-220	€-220	€-221
Average Return (%)	-4,67	-4,66	-4,67

(a) out-of -sample test, based on the year 2006-2012

Besides, it seems that the forecast ability has reduced for the validation sample of 2013. The training sample had a hit rate of approximately 71,30%. The validation sample has a hit rate of approximately 70,30%. A shrinkage[8] of the hit rate of approximately 1% occurs when predicting tennis match outcomes for the year 2013 based upon the training sample 2006-2012. Moreover, the average rate of returns has diminished. The average return declines from approximately -3,30% for the training sample to approximately -4,70% for the validation sample.

At first sight, it seems that both models have lost some predictive power. However, this decline in forecast ability should be put into perspective: the logic strategy method also predicted worse for the year 2013 (70,21%). Given that all three methods tend to predict the favorite to win, it appears that favorites on average performed worse in 2013 than in the preceding years 2006-2012.

We can conclude that there is evidence that the variable *BookmakerDispersion* significantly improves the fit of the basic forecasting model. However, this statistical improvement does not result in a higher hit rate or better financial results. Furthermore the predictions of the binary logistic regression models do not result in better forecasts than when a logic betting strategy, where only bets of favorites are accepted, is applied.

[8] Hit rate of training sample (71,39%) - hit rate of validation sample (70,26%)

8.3 Hypothesis III

"Hypothesis 3: Higher bookmaker dispersion leads to significantly lower future return for bettors in the tennis betting market."

Below, two methods will be described to measure the extent of the dispersion effect in the tennis betting market. The first method is successfully applied in the financial stock market, but may lead to misleading results when conducted in the betting market. However, it is still described to make a clear distinction with the second method. The second method is a new method, customized for the betting market.

Both methods are based upon the work of Diether, et al. (2002), who tested the dispersion effect for the financial stock market. In their research, the dispersion about a stock's value is measured through the differences in analysts' forecasts. If the analysts agree upon the current value of a stock, the dispersion is considered to be low. If on the other hand, all the analysts disagree upon the current value of a stock, dispersion is considered to be high. To accurately measure the impact of the dispersion effect on the returns, Diether, et al. divided the stocks into 5 subgroups according to the dispersion in analysts' forecasts. The first quintile consists of the 20% stocks with the lowest dispersion. The fifth quintile contains the 20% stocks with the highest dispersion in analysts' forecasts. So basically, in their method, two subsamples are compared to each other, one subsample with high dispersion and another subsample with low dispersion in analyst's forecasts. As Miller (1977) predicted, they found evidence that the stocks of the low dispersion quintile had better financial results than stocks in the high dispersion quintile. This is a very intuitive working method, and worked fine for the analysis of the financial stock market by Diether, et al. in 2002.

In this paper, the efficiency of the betting market is analyzed. Instead of using differences in analysts' forecast as a proxy to measure differences in opinion about a stock, this method will use *Bookmaker Dispersion* as a proxy to measure the differences in opinion about a tennis match outcome. To divide the dataset in subsamples according to bookmaker dispersion, two methods will be described in this chapter. The first method is the most intuitive method following Diether, et al. (2002)'s methodology described above. However, this method could lead to misleading results since the odds of the betting market are contaminated with the robust *favorite-longshot bias*. This can better be demonstrated by just applying the first method. The second method offers a solution for the emerged problems that occurred in the first method.

8.3.1 Method I

All 69.058 bets were divided into 5 equally large subsamples (quintiles), according to their B*ookmakerDispersion*. The bets with the lowest dispersion are assigned to D1. The bets with the highest dispersion are assigned to D5. The results are presented in the **table 11**.

The quintile of bets with the lowest dispersion has an average rate of return of - 1,5%. The bets in the highest dispersion quintile (D5) yield an average loss of approximately -26%. Based on these results, one would be tempted to conclude that higher bookmaker dispersion leads to lower returns and vice-versa.

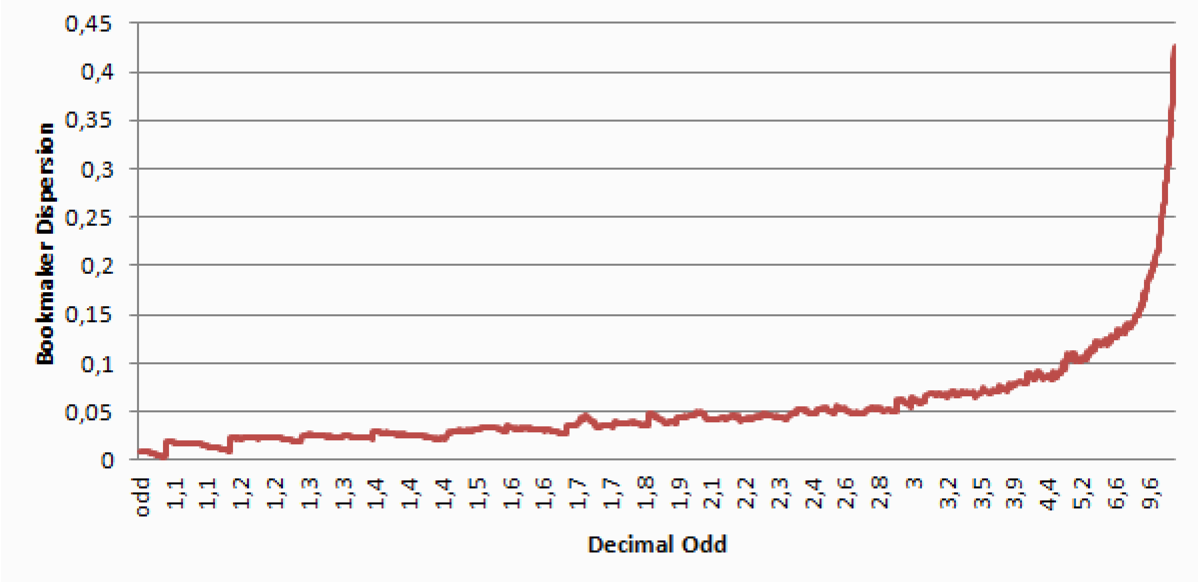
Table 11 : Dispersion quintiles, METHOD I

Dispersion Quintiles	Count	Profit (€)	Average Return (%)	Average Odd (€)
D1 (low)	13.812	-205	-1,50	1,37
D2	13.812	-542	-4,00	1,64
D3	13.813	-852	-6,20	1,99
D4	13.812	-1400	-10,10	2,72
D5 (high)	13.813	-3449	-25,97	5,94

However, it would be a premature conclusion to dedicate this finding to the effect of bookmaker dispersion. This is because the results interfere with presence of the favorite-longshot bias in the tennis betting market. **Table 11** also includes the average odds of the quintiles. The bets of the quintile with the lowest dispersion (D1) have an average odd of €1,37, while the bets of the quintile with the highest dispersion have a much higher average odd, namely €5,94. It appears that the average odds increase when bookmaker dispersion increases. This finding can be illustrated by **figure 1.**

Notwithstanding the usage of a coefficient of variation to measure bookmaker dispersion, a clear upward trend can be noticed trough the odd range.

Figure 1 : Bookmaker dispersion across the odd range



This relationship can further be demonstrated through conducting the following linear regression:

BookmakerDispersion = β1 + β2 (*AverageOdd*)

The regression results are summarized in **table 12**. The coefficient of *AverageOdd* is significantly positive, implying that a higher average odd leads to higher dispersion among the placed odds by bookmakers.

Table 12: Relationship odds and bookmaker dispersion

DV: Bookmaker Dispersion	**Coefficient**	**Standard Error**
Constant	-0,06***	0,000
AverageOdd	0,023***	0,000

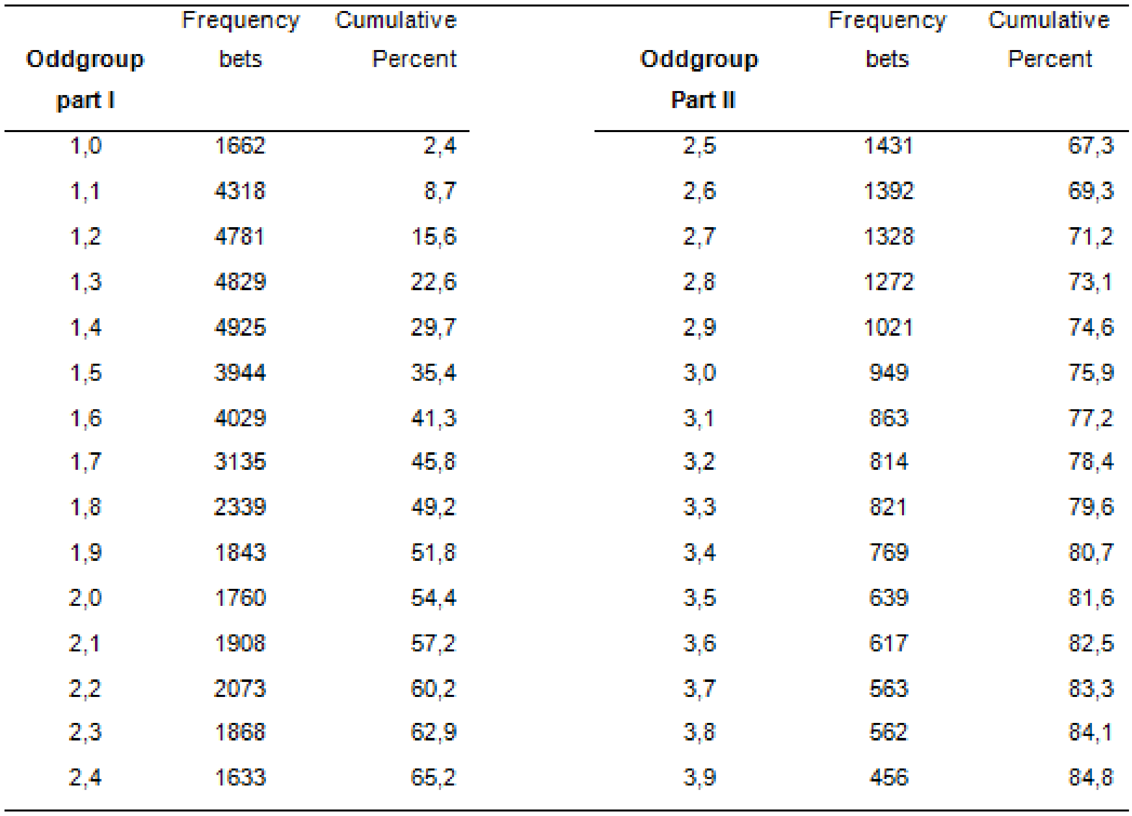
*** indicate statistical significance at the 1%

As a consequence, if one compares the subsamples D1 (low dispersion) with D5 (high dispersion) in **table 11** with respect to the average return, one basically compares the return of a subsample with favorites (low odds), with a subsample with long-shots (high odds). This means the extent of the favorite-longshot bias on returns is measured, rather than the extent of *Bookmaker Dispersion* on the returns. This is a problem, since we want to measure the effect of bookmaker dispersion in particular.

8.3.2 Method II & Results

In *Method I,* the division in 5 subsamples according to bookmaker dispersion was applied on the complete dataset. This led to misleading results. Now, a different approach will be applied. Again, the bets will be divided into 5 subsamples according to *BookmakerDispersion*. However, when using method II, a subsample will be made for every odd to 1 digit after the comma. To demonstrate this, we refer to **table 13**.

Table 13: Oddgroups



Oddgroup part I	Frequency bets	Cumulative Percent	Oddgroup Part II	Frequency bets	Cumulative Percent
1,0	1662	2,4	2,5	1431	67,3
1,1	4318	8,7	2,6	1392	69,3
1,2	4781	15,6	2,7	1328	71,2
1,3	4829	22,6	2,8	1272	73,1
1,4	4925	29,7	2,9	1021	74,6
1,5	3944	35,4	3,0	949	75,9
1,6	4029	41,3	3,1	863	77,2
1,7	3135	45,8	3,2	814	78,4
1,8	2339	49,2	3,3	821	79,6
1,9	1843	51,8	3,4	769	80,7
2,0	1760	54,4	3,5	639	81,6
2,1	1908	57,2	3,6	617	82,5
2,2	2073	60,2	3,7	563	83,3
2,3	1868	62,9	3,8	562	84,1
2,4	1633	65,2	3,9	456	84,8

This table presents the frequency of bets when rounding the odds to 1 digit after the comma. Every bet in the dataset, with the same odd to 1 digit after the comma, belongs to the same *oddgroup*. For example, there are 4.029 bets with an average odd rounded at 1,6. This means oddgroup 1,6 contains 4.029 bets. In total, there are 30 oddgroups, ranging from

1,0 to 3,9.[9] Subsequently, the bets in each oddgroup are divided into 5 additional subgroups, only this time according to the *BookmakerDispersion*. In other words, for each oddgroup, 5 more subsamples are taken, d1, d2, d3, d4 and d5 (lowercases d) based on their bookmaker dispersion.

Finally, the 5 subsamples (d1, d2, d3, d4, d5) of all the 30 oddgroups are aggregated to each other to obtain again 5 large subsamples (D1, D2, D3, D4, D5). The results of *method II* are presented in **table 14**.

Similar to method I, the subgroup D1 includes all the bets with the 20% lowest dispersion. Only this time it includes the 20% bets with the lowest dispersion of each oddgroup, and not the bets with the lowest dispersion of the complete dataset. This is to ensure that the average odd is the same for each dispersion quintile. The subgroup D5 includes all the bets with the 20% highest dispersion of each oddgroup.

Table 14 : Dispersion quintiles, METHOD II (a)

Dispersion Quintiles	#Bets	Profit (€)	Average return (a)	Average odd (€)	Win %
D1 (low)	11.870	-441,8	-3,72%	**1,90**	58,02%
D2	11.890	-448,2	-3,77%	**1,90**	57,69%
D3	11.886	-753,6	-6,34%	**1,91**	56,47%
D4	11.890	-777,1	-6,54%	**1,91**	56,19%
D5 (high)	11.902	-1084,8	-9,11%	**1,91**	54,89%

(a) note that the quintiles are different from the quintiles presented in **table 11**

There are approximately[10] 5 equal dispersion quintiles of 11.900 bets.

Applying method II clearly results in **similar average odds of the quintiles.** The bets of every subgroup have an average decimal odd of approximately €1,91. This is in large contrast with the results of method I, where the average odd of the 5 subgroups increased when the

[9] It is noticed that the frequency of bets in the oddgroups decreases when the odds increase. There are only 456 bets of tennis players left with an average decimal odd of 3,9. Since we want to work with samples that are big enough, we limited our dataset to bets with an average odd smaller than 4,0. 58.544 bets remain, which still is approximately 85% of the whole dataset (cumulative percent in the table).

[10] The count differences are devoted to rounding errors that occurred when creating the oddgroups.

dispersion across bookmakers increased (**table 11**). It appears that method II successfully eliminated the extent of the favorite-longshot bias. The results can solely be dedicated to differences in the dispersion of bookmaker odds.

As expected by hypothesis 3, an inverse monotonic relationship can be noticed between *dispersion* and *profit*. As bookmaker dispersion increases, profit decreases and vice-versa. The average rate of return for the group with the lowest dispersion (D1) amounts -3,72%. The group with the highest dispersion (D5) has an average rate of return -9,11%. The average returns of the dispersion quintiles significantly differ from each other[11].
It appears that, the closer the prices of the 5 bookmaker odds are for a certain bet, the higher the average rate of return is for the bettors. These returns are calculated with the average of the 5 odds. Even though the average odds of the 5 bookmakers are the same, higher dispersion among the bookmaker odds yields lower returns for the bettors and vice-versa. Furthermore, it is noticed that tennis players in the lowest dispersion group have a higher chance of winning. Approximately 58 % of the tennis players of which the betting odd is situated in the lowest 20% dispersion group, have won their match. Players in the D5 quintile on the other hand, only won 54,89% of their matches.

This phenomenon was not revealed before. It appears that besides the favorite-longshot bias, there is a second anomaly in the tennis betting, namely the *bookmaker dispersion effect*. **Table 15** on the following page provides a reproduction of both the extent of the FLB and the extent of the dispersion effect in the tennis betting market. If one compares the returns of the favorites and the long-shots to one another, the extent of the FLB becomes clear. Favorites have in both dispersion quintiles a significantly higher average return compared to the return for betting on long-shots. A comparison of the two columns highlights the dispersion effect. The dispersion effect is clearly present for both the favorites as the long-shots. As one can expect, the upper left quadrant yields the best average returns (-0,84%). This represents a betting strategy where only bets are accepted of favorites with low bookmaker dispersion. The lower right quadrant has the worst results. Accepting only bets on long-shots in the high dispersion quintile yields an average loss of 22,32%.

[11] Kruskal-Wallis Chi-square = 22,96, ANOVA between groups F statistic = 6,841

Table 15 : Dispersion effect and the FLB

	Average return	
	Low dispersion (D1)	**High dispersion (D5)**
Favorites (1,0-1,4)	-0,84%	-3,94%
Longshots (3,0-3,9)	-10,23%	-22,32%

Table 16 compares the extent of the dispersion effect over the years 2006 until 2013. The presence of the dispersion effect is relatively persistent over the years. In almost every year, bets in the D1 quintile yielded higher returns than bets in the D5 quintile. The year 2009 is the sole exception to the bookmaker dispersion effect, where the opposite is true.

In 2013, bets in the lowest dispersion quintile did not even yield a loss for the bettors (0%), whereas the high dispersion quintile yielded an average loss of 10,88%.

Table 16 : The bookmaker dispersion effect, yearly overview

	average return	
YEAR	**Low dispersion (D1)**	**High dispersion (D5)**
2006	-0,55%	-11,62%
2007	-4,00%	-5,37%
2008	-1,52%	-4,91%
2009	-6,65%	-3,86%
2010	-3,07%	-12,85%
2011	-2,50%	-11,61%
2012	-4,52%	-8,98%
2013	0,00%	-10,88%

8.3.3 Possible explanation for the bookmaker dispersion effect

The large extent of these results is rather surprising. It appears that the dispersion effect is present in the tennis betting market. For the stock market, Millers (1977) argued that investors, who consider the stock prices to be overrated, are kept out of the market due to high short-sell costs. As a result, the stock had lower future returns. However, these short sell costs do not occur in the tennis betting market, which means they do not qualify as valuable argument for the existence of the bookmaker dispersion effect. Why does lower dispersion in bookmaker odds lead to significantly higher returns? Other explanations should be sought.

In chapter 6.2, it was explained how bookmakers price their odds. This explanation was mainly based on 4 pricing mechanisms, quantitative and qualitative analysis, competitor's odds and the public opinion. As mentioned, statistics and qualitative research of odd compilers are mostly important to place the opening odds. In this paper closing odds[12] are used to make the analysis. Therefore, the first two pricing mechanisms can be expected to have a rather modest influence on the dispersion among bookmakers of the final odds. The third mechanism, the comparison of competitor's odds, leads to more congruent odds instead of more dispersion. Thus, an explanation will be sought in the impact of the public opinion.

One can argue that the more the public opinion agrees about a tennis match outcome

1. The smaller the bookmaker dispersion will be
2. The higher the average return will be for the bettors.

This may explain the existence of the bookmaker dispersion effect in the tennis betting market.

As stated previously, bookmakers do not care about accuracy of their match result forecasts, since their primary objective is to make a certain profit. In order to guarantee profits, they have to attract enough bettors on both sides of the bet. This is what was called "balancing the books". However, different public opinions about the outcome of a certain tennis match can disturb this balance. Therefore,

If the public opinion agrees upon the winning probabilities of the two opponents in a tennis match, the public opinion is likely to also agree about the odds placed by the different bookmakers. The probability of the tennis match outcome is easy predictable and this results in more accurate forecast of the public opinion. The returns will be higher. Furthermore, the

[12] Closing odds are the last odds available to bet on before the start of the tennis match.

odds are placed in such way that the books are balanced. Stable betting volumes on both sides of the bet can be expected for every bookmaker. Bookmaker dispersion will be low.

If on the other hand, the public opinion does not agree upon the probabilities of a tennis match outcome, the public opinion will be more likely to consider the odds mispriced. The outcome is difficult to predict, resulting in bad forecasts and lower profits. Furthermore, different betting volumes on both sides can be expected among bookmakers. To keep the books balanced, bookmakers will adjust the odds according to the betting volumes they incur. Due to a volatile market, betting volumes may differ across the different bookmakers. Thus, bookmaker dispersion will be high.

This was partially confirmed by the analysis of two binary variables available in the dataset, namely Grand-Slam/non Grand-Slam tournaments and early-round/late-round matches.

As mentioned, Grand-Slam tournaments and late-round matches are considered to be more high-profile because the players generally are more famous, the price money is higher, and the respective matches get more media attention.

One can argue that the public opinion agrees more in high profile tennis matches, because it is easier to estimate the probabilities of the tennis match outcome compared to low-profile tennis matches. (supra, p17).

Table 17 compares Grand-Slam and non Grand-Slam tournaments with respect to the dispersion in placed bookmaker odds. An inverse monotonic relationship can be noticed between the count of Grand-Slam bets and the level of dispersion. In total, 2.457 Grand-Slam bets are situated in the low dispersion quintile (D1), whereas there are only 2.097 Grand-Slam bets in the highest dispersion quintile (D5). The differences in count are statistically significant[13]. For non Grand-Slam matches, the opposite is true. However, this is a logic consequence of the fact that every dispersion quintile contains an equal amount of bets and a tennis match is either a Grand-Slam or non Grand-Slam match.

[13] Chi-Square test = 36,4

Table 17 : Dispersion effect, Grand-Slam and non Grand-Slam tournaments

	Grand-Slam		Non Grand-Slam	
dispersion	**count**	**percentage**	**count**	**percentage**
D1 (low)	2.457	21,71%	9.413	19,56%
D2	2.337	20,65%	9.553	19,85%
D3	2.271	20,07%	9.615	19,98%
D4	2.155	19,04%	9.735	20,23%
D5 (high)	2.097	18,53%	9.805	20,38%

Stronger conclusions can be drawn from comparing early-round and late-round tennis matches. Late-round matches are matches of the quarterfinals, semifinals and finals of the tournament. Early-round matches are solely matches of the first round of a tournament. Since the matches that are not early-round or late-round are omitted, it provides stronger proof in both ways. The results are summarized **in table 18**.

Table 18 : Dispersion effect, early-round and late-round tennis matches

	Late-round		Early-round	
dispersion	**count**	**Percentage (%)**	**count**	**Percentage (%)**
D1 (low)	1.150	25,2%	5.111	18,7%
D2	1.044	22,8%	5.397	19,8%
D3	871	19,1%	5.436	19,9%
D4	842	18,4%	5.589	20,5%
D5 (high)	662	14,5%	5.758	21,1%

The same conclusion can be drawn as in **table 17**. For both the late-round and early-round tennis matches, the differences in count between the dispersion quintiles are statistically significant[14]. Late-round tournaments matches are more located in the low dispersion quintile (D1) with approximately 25,2% of all late round matches. The opposite seems to be true for early-round tennis matches. More early-round matches are located in the highest dispersion quintile (5.758) than in the lowest dispersion quintile (5.111).

These findings support our expectation that a more consenting agreement in public opinion leads to lower bookmaker dispersion and vice-versa.

[14] Late- round Chi-Square = 72,33; early-round Chi-Square = 42,46

9. Betting strategy

In 2007, Forrest & McHale proposed a betting strategy. Tennis data of the year 2003/2005 were used. Similar as this thesis, they also had a dataset with odds of 5 different bookmaker available. They suggested a betting strategy where only the odds in the 0,8-0,9 implied probability range were accepted. The strategy yielded a positive return of more than 2%, which implied a violation of weak form efficiency[15]. However, the excess returns were not significant. With the availability of a larger dataset, and the knowledge of the bookmaker dispersion effect, it was worth the gamble to try and make a profitable strategy. In this chapter a new betting strategy will be explained.

The strategy is based upon two main findings described in this paper. First, the strategy will make optimal use of the knowledge of the FLB in the tennis betting market. It appeared that bets, with an implied probability range between 0,7-0,8 and between 0,9-1,0 were the best odds to accept. Secondly, it appeared that bets with low bookmaker dispersion among the different bookmakers yielded higher returns for the bettors compared to bets with high bookmaker dispersion. So basically, a good strategy would be to only accept bets of favorites with low bookmaker dispersion.

It is important that the reader can make the distinction between this chapter and the previous chapters with respect to the calculation of returns. In the previous chapters, odds were the average odds of the 5 bookmakers, and the returns were calculated with these average odds. However, in order to find a profitable betting strategy, it is better to always bet on the highest-quoted odd, instead of betting an equal amount of money on all 5 bookmakers.

This has implications for the betting strategy because a contradiction occurs. The contradiction can be illustrated with data provided in **table 19** below**.** The returns when betting on the average odds were already provided in **table 13**. The dispersion effect was clearly present, and most favorable returns are obtained in the D1 quintile, with an average

[15] Weak form efficiency is violated when excess returns are obtained through the analysis of historical data.

return of -3,72%. However, if one now only accepts bets of the highest quoted odd of the 5 bookmakers, this clearly outperforms the returns obtained through betting on the average of the 5 bookmaker odds. The D1 quintile has an average return of -1,66% when betting on the highest odds.

Table 19 : Dispersion Quintiles , acceptance highest odd and average odd

dispersion	Highest odd		Average odd	
	average return	count	average return	count
D1 (Low)	-1,66%	11.870	-3,72%	11.870
D2	-0,52%	11.890	-3,77%	11.890
D3	-2,27%	11.886	-6,34%	11.886
D4	-1,36%	11.890	-6,54%	11.890
D5 (High)	-1,18%	11.902	-9,11%	11.902

More importantly, it appears that the bookmaker dispersion effect has disappeared. There is no more an inverse relationship between average return and bookmaker dispersion. Furthermore, D2 quintile has become the most profitable quintile instead of D1 when betting in the best available odd. The D2 quintile yields a return of -0,52%. Higher bookmaker dispersion has therefore two consequences.

First of all, it implies lower average returns when taking the average odds into account. Secondly, higher bookmaker dispersion also increases the difference between the highest quoted odd and this average odd, making bets with high bookmaker dispersion just more attractive instead of less attractive.

9.1 Betting strategy requirements

Basically, the betting strategy in this paper entails three requirements.

1) Accept bets with an implied probability between 0,7-0,8 and 0,9-1,0
2) Accept bets in the D2 quintile
3) Accept the highest bet

Requirement one is self-evident. To meet requirement two, the betting strategy will decide on the basis of historical data to which dispersion quintile the bet belongs. Tennis data from the year 2006 until 2012 is used to define the boundaries of the D2 quintile. As explained earlier, the dispersion quintiles D1, D2, D3, D4 and D5 are an aggregation of the 30 oddgroups[16]. Since the participating odds are limited by *requirement ,1* only four oddgroups need to be considered, namely 1,0; 1,1 ; 1,2 ; 1,3 and 1,4.

Using data of the year 2006 until 2012, the following boundaries for the D2 quintile were obtained.

Table 20: D2 Boundaries, based on the year 2006-2012

	Bookmaker disperion: D2 Quintile	
Oddgroup	**Minimum**	**Maximum**
1,0	0,004948	0,006932419
1,1	0,010743	0,013990545
1,2	0,016128	0,020412415
1,3	0,016522	0,020895236
1,4	0,01653	0,021657431

9.2 Example

A real life example will further clarify the proposed betting strategy. On the 26st of June 2010, Wozniacki played a tennis match against Pavlyuchenkova in the third round of Wimbledon. Five bookmakers had placed odds of this tennis match, namely BET365, Centrebet, Expekt, Pinnacles Sports and Unibet.

[16] see chapter 6, oddgroups are derived from the average odds, and not the highest odds used in the betting strategy

A summary of the odds is provided in **table 21**.

Table 21 : Example betting strategy

	Bookmaker odds	
Bookmaker	**Wozniacki C.**	**Pavlyuchenkova A.**
BET 365	€1,33	€ 3,25
Centrebet	€ 1,33	€ 3,10
Expekt	€ 1,29	€ 3,50
Pinnacles Sports	€ 1,35	€ 3,56
Unibet	€ 1,33	€ 3,25

The calculations of the application of the betting strategy are provided in **table 22**.The odds of Wozniacki meets requirement 1 and 2. The implied probability of 0,75 is located in the 0,7-0,8 range. Furthermore, the bookmaker dispersion, measured by the coefficient of variation, is located in the 0,01650 - 0,02089 range of the rounded average odd of 1,3. Requirement 3 states that the bet has to be accepted from Pinnacles Sports at a decimal odd of € 1,35.

Table 22 : Calculations

Average Odd	(1.33+1.33+1.29+1.35+1.33)/5 = €1,326.
Rounded average odd to one digit after the comma	€ 1,3
Implied probability	1/1,326 = 0,75
Bookmaker dispersion	S.D. (1,33;1,33;1,29;1,35;1,33)/1.326= 0,016523

9.3 Results

In order to present the results of the proposed betting strategy in an objective manner, two other strategies will be included in the analysis, namely an advanced strategy and a basic strategy. The requirements are summarized in **table 23**.

Table 23 : Betting strategies requirements

	D2 Quintile	FLB (a)	Highest Odd	BET365 odd
Master Strategy	X	X	X	
Advanced Strategy		X	X	
Basic Strategy		X		X

(a) Only bets in the 0,7-0,8 & 0,9-1,0 implied probability range are accepted

The master strategy follows the strategy as proposed in this paper. A bettor applying this strategy only accepts bets in the D2 quintile, and the 0,7-0,8 & 0,9-1,0 implied probability range. Furthermore, he only accepts the highest odd of 5 different bookmakers.

The advanced strategy bettor accepts bets in the 0,7-0,8 & 0,9-1,0 implied probability range. Furthermore, only the highest odd of 5 different bookmakers is accepted. In the basic betting strategy, only bets in the in the 0,7-0,8 & 0,9-1,0 implied probability range are accepted. Furthermore, the bettor has only 1 betting account, namely BET365 (real data of BET365 is used). The results of these 3 betting strategies are summarized in **table 24**.

Table 24 : Results betting strategies

		Master Strategy		Advanced Strategy		Basic Strategy	
YEAR	**Total Bets**	# accepted	return	# accepted	return	# accepted	return
2006	4.811	150	1,37%	866	1,06%	866	-0,97%
2007	8.819	326	3,04%	1.776	2,05%	1.776	-1,96%
2008	8.797	349	1,00%	1.620	0,67%	1.620	-3,73%
2009	9.105	447	-1,10%	1.739	0,00%	1.739	-3,33%
2010	9.451	312	3,47%	1.712	2,44%	1.712	-2,22%
2011	9.629	370	2,82%	1.782	2,56%	1.782	-1,62%
2012	9.149	375	2,45%	1.792	-0,32%	1.792	-3,84%
2013	**9.300**	**378**	**2,97%**	**1.722**	**0,03%**	**1.722**	**-3,46%**
Total	**69.061**	**2.707**	**1,93%**	**13.009**	**1,07%**	**13.009**	**-2,74%**

Total bets are the total amount of available bets in the corresponding year. Accepted bets represents the amount of bets that satisfies the betting strategy requirements as described in **table 23**. The master strategy clearly outperforms the advanced strategy and the basic strategy. A weighted average return of 1,93% is obtained. Note that this is the average return per bet, and not the return on yearly basis. Furthermore, the average return per bet yields a positive return for almost every year.

Accept in 2009, were a loss of -1,1% is suffered. The advanced betting strategy also yields an overall positive return of 1,07%. However, the returns are lower than the returns when the master betting strategy is applied. The advanced betting strategy only outperforms the master strategy in 2009. Following the basic strategy leads to a loss of -2,74%. It must be noted that BET365 is the most expensive bookmaker in the dataset, with the lowest average odds.

As mentioned, the year 2013 is an out-of-sample validation test, to avoid problems such as over fitting. Data of the year 2006 to 2012 was used to mark the boundaries for the D2 quintile. An average return of 2,97% was obtained for the out-of-sample dataset, which is a positive return. This implies a violation of weak form efficiency.

However, similar to Forrest and McHale(2007), the returns are not significantly positive. **Table 25** summarizes the statistical bounds for the returns of the master strategy. If the proposed betting strategy is evaluated on an annual basis, the average returns do not significantly differ from 0. There is only a significant positive return when the whole dataset 2006-2013 is taken into account. This is due to the larger dataset of 2.707 bets.

Table 25 : Statistical bounds, Master Strategy

YEAR	count	Average return	p- value	Lower-bound (a)	Upper-bound
2006	150	1,37%	0,748	-7,01%	9,75%
2007	326	3,04%	0,266	-2,33%	8,42%
2008	349	1,00%	0,727	-4,55%	6,52%
2009	447	-1,10%	0,672	-6,02%	3,89%
2010	312	3,47%	0,230	-2,20%	9,14%
2011	370	2,82%	0,287	-2,38%	8,02%
2012	375	2,45%	0,328	-2,47%	7,36%
2013	378	2,97%	0,257	-2,18%	8,12%
2006 - 2013	**2.707**	**1,93%***	**0,049**	**0,01%**	**3,85%**

(a) bounds are calculated at a 95% significance level

* indicates statistical significance at the 5% level

In this chapter, a betting strategy was proposed. The strategy was based upon the knowledge of both the bookmaker dispersion effect and the favorite longshot bias. The strategy used data of the year 2006-2012 to create the betting strategy requirements. An out-of-sample test on 2013 yielded positive returns. This implies a violation of weak form efficiency. However, these positive returns are not statistically significant.

10. Discussion

The favorite-longshot bias is a well known anomaly in the betting market, and is discussed extensively in the literature (supra p.6). This paper provides evidence for a second anomaly in the betting market. It appears that the dispersion effect, well-established in the financial stock market, is also present in the tennis betting market. The afore-mentioned research finding was uncovered after applying a customized method for the betting market, as described in chapter 8.3.2. This phenomenon was previously unknown as it has not been discussed before in the literature. A possible explanation for the dispersion effect in the tennis betting market is given in this paper. However, this explanation is based on qualitative reflections, and is not fully supported by quantitative evidence.

Two points of criticism can be mentioned regarding the work performed. Firstly, in order to get the bookmaker dispersion quintiles, only average odds lower than 3,9 are considered (supra p.36). Even though these comprise approximately 85% of the dataset, all odds should preferably be taken into account to make a more complete analysis of the bookmaker dispersion effect.

Secondly, all database operations are conducted with the aid of Excel. Notwithstanding the rigorous work and practices used (e.g. extensive double-checking), the margin of error is higher when working with spreadsheet programs, because of the extensive input and manipulation of data. The use of technical computing programs to structure the database in a more efficient way is recommended for future research. To arrive to clear, well substantiated conclusions regarding the bookmaker dispersion effect, further analysis is thus required. The betting volumes and intra-day odd price changes could clarify the causes of the bookmaker dispersion effect. This data was not available in the dataset used in this paper. Furthermore, in future research, the application of the bookmaker dispersion effect could be expanded to other betting markets, such as the football betting market. Odds of 7 different bookmakers are available on the website < http://www.football-data.co.uk/>

All datasets used in this paper are available upon request. These comprise the complete dataset of the year 2006-2013 and the dataset with respect to the dispersion subsamples. Anyone interested can always contact me at <kwintenderave@hotmail.com>.

11. Conclusion

This thesis started with a description of the Efficient Market Hypothesis. One of the anomalies found in the financial stock market is the dispersion effect. There is evidence that stocks with higher dispersion in analysts' earnings forecasts yield lower future returns than otherwise similar stocks (Diether&Malloy, 2002). However, difficulties occur when testing the financial stock market on its efficiency. One reason is that there is no fixed point in time where the true value of the underlying company is known. The betting market is often used to test market efficiency. Betting markets share important properties with the stock market. Also, they possess some clear advantages with respect to efficiency testing. Investors are replaced by bettors. These bettors can accept bets against a specific odd price placed by bookmakers. After the match is played, the true value of the bet is revealed and feedback for the closing odd is received. The main topic of this master thesis was the examination of the dispersion effect in the betting market. Even though bookmakers place odds for the same tennis matches, disagreements upon odd prices often occur across these bookmakers. This is called *bookmaker dispersion.* The database used in this paper had odds of 5 different bookmakers at its disposal.

First, the extent of the most robust anomaly in the betting market, the favorite-longshot bias, was compared for female and male tennis matches. It was expected that the extent of the favorite-longshot bias would be mitigated in the male subsample, because male tennis matches are considered to be more high-profile compared to female tennis matches. However, no evidence was found for this expectation.

Secondly, Fama (1970) stated that a market is efficient if the security prices reflect all available information. The counterpart for security prices in the betting market are the odds placed by bookmakers. The betting market can be considered efficient if the odds contain all available information to forecast the probabilities of the outcome of a tennis match. A basic forecasting model was proposed, using solely the average of 5 bookmakers' odds as variable. This model was compared with an extended model that also includes the magnitude of the dispersion among the bookmakers' odds. Based on statistical measures, evidence was found that bookmaker dispersion statistically improves the basic forecasting model. However, this statistical improvement did not lead to an improved predictability of a tennis match outcome or to better financial results.

Third, the link with the dispersion effect in the financial stock market was made. One can expect that, similar to the financial market, higher bookmaker dispersion should yield lower future returns. In this paper, a unique method for the betting market was developed to obtain subsamples according to the dispersion of bookmaker odds. Evidence was found that the dispersion effect is also present in the tennis betting market. The results presented in this paper show that higher bookmaker dispersion leads to lower future returns for the bettors and vice-versa. This could be explained in terms of a consenting public opinion about the outcome of a tennis match. However, this explanation is based on qualitative reflections, and is not fully supported by quantitative evidence. The availability of betting volumes could provide more insights in the causes of the bookmaker dispersion effect in the tennis betting market.

Finally, a betting strategy was explained. The betting strategy uses historical data of the year 2006 to 2012 to develop the model. Positive returns were obtained for an out-of-sample year 2013. This implies violation of the weak form efficiency in the tennis betting market. However, these positive returns were not found to be significantly different from zero.

References

Betfair,2012 , URL <https://betting.betfair.com/the-art-of-bookmaking.html>

Cain, M., Law, D., & Peel, D. (2000). The Favourite-Longshot Bias and Market Efficiency in UK Football betting. *Scottish Journal of Political Economy*, *47*(1), 25-36.

Cain, M., Law, D., & Peel, D. A. (2003). Some analysis of the properties of the Harville place formulae when allowance is made for the favourite-long shot bias employing Shin Win probabilities. *Applied Economics Letters*, *10*(1), 53-57

Chen, J., Hong, H., & Stein, J. C. (2002). Breadth of ownership and stock returns. *Journal of financial Economics*, *66*(2), 171-205.

Chron, 2012, URL :<http://work.chron.com/average-salary-professional-tennis-players-6052.html>
Crafts, N. F. (1985). Some evidence of insider knowledge in horse race betting in Britain. *Economica*, *52*(207), 295-304.

del Corral, J., & Prieto-Rodriguez, J. (2010). Are differences in ranks good predictors for Grand Slam tennis matches?. *International Journal of Forecasting*, *26*(3), 551-563.

Deschamps, B., & Gergaud, O. (2012). Efficiency in betting markets: evidence from English football. *The Journal of Prediction Markets*, *1*(1), 61-73.

Diether, K. B., Malloy, C. J., & Scherbina, A. (2002). Differences of opinion and the cross section of stock returns. *The Journal of Finance*, *57*(5), 2113-2141.

Dowie, J. (1976). On the efficiency and equity of betting markets. *Economica*,*43*(170), 139-150.

Figlewski, S. (1979). Subjective information and market efficiency in a betting market. *The Journal of Political Economy*, 75-88.

Forrest, D., & McHale, I. (2007). Anyone for tennis (betting)?. *The European Journal of Finance*, *13*(8), 751-768.

Froot, K. A., Scharfstein, D. S., & Stein, J. C. (1992). Herd on the street: Informational inefficiencies in a market with short-term speculation. *The Journal of Finance*, *47*(4), 1461-1484.

Goddard, J. (2005). Regression models for forecasting goals and match results in association football. *International Journal of forecasting*, *21*(2), 331-340.

Goddard, J., & Asimakopoulos, I. (2004). Forecasting football results and the efficiency of fixed-odds betting. *Journal of Forecasting*, *23*(1), 51-66.

Hausch, D. B., Ziemba, W. T., & Rubinstein, M. (1981). Efficiency of the market for racetrack betting. *Management science*, *27*(12), 1435-1452.

Hodges, S., Lin, H., & Liu, L. (2011). Fixed odds bookmaking with stochastic betting demands. *European Financial Management*

Hong, H., & Stein, J. C. (2003). Differences of opinion, short-sales constraints, and market crashes. *Review of financial studies*, *16*(2), 487-525.

Hwang, C. Y., & Li, Y. (2008). Analysts' Incentives and the Dispersion Effect.*Available at SSRN 1361731*.

Klaassen, F. J., & Magnus, J. R. (2003). Forecasting the winner of a tennis match. *European Journal of Operational Research*, *148*(2), 257-267.

Lahvička, J. (2014). What causes the favourite-longshot bias? Further evidence from tennis. *Applied Economics Letters*, *21*(2), 90-92.

Law, D., & Peel, D. A. (2002). Insider Trading, Herding Behaviour and Market Plungers in the British Horse–race Betting Market. *Economica*, *69*(274), 327-338.

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

Makropoulou, V., & Markellos, R. N. (2011). Optimal Price Setting In Fixed-Odds Betting Markets Under Information Uncertainty. *Scottish Journal of Political Economy*, *58*(4), 519-536.

Makropoulou, V., & Markellos, R. N. (2011). Optimal Price Setting In Fixed-Odds Betting Markets Under Information Uncertainty. *Scottish Journal of Political Economy*, *58*(4), 519-536.

Malkiel, B. G., & Fama, E. F. (1970). EFFICIENT CAPITAL MARKETS: A REVIEW OF THEORY AND EMPIRICAL WORK*. *The journal of Finance*,*25*(2), 383-417.

Miller, E. M. (1977). Risk, uncertainty, and divergence of opinion. *The Journal of Finance*, *32*(4), 1151-1168.

Pope, P. F., & Peel, D. A. (1989). Information, prices and efficiency in a fixed-odds betting market. *Economica*, 323-341.

Schnytzer, A., & Shilony, Y. (1995). Inside information in a betting market. *The Economic Journal*, 963-971.

Schnytzer, A., Lamers, M., & Makropoulou, V. (2010). The impact of insider trading on forecasting in a bookmakers' horse betting market. *International Journal of Forecasting*, *26*(3), 537-542

Shin, H. S. (1991). Optimal betting odds against insider traders. *The Economic*

Journal, *101*(408), 1179-1185.

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

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

Smith, M. A., Paton, D., & Williams, L. V. (2006). Market Efficiency in Person-to-Person Betting. *Economica*, *73*(292), 673-689

Soccerwidow, 2014, URL: <http://www.soccerwidow.com/betting-advice/bookmakers/how-do-bookmakers-tick/>

Spann, M., & Skiera, B. (2009). Sports forecasting: a comparison of the forecast accuracy of prediction markets, betting odds and tipsters. *Journal of Forecasting*, *28*(1), 55-72.

Štrumbelj, E., & Šikonja, M. R. (2010). Online bookmakers' odds as forecasts: The case of European soccer leagues. *International Journal of Forecasting*,*26*(3), 482-488.

Tauchen, G. E., & Pitts, M. (1983). The price variability-volume relationship on speculative markets. *Econometrica: Journal of the Econometric Society*, 485-505.

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

The Guardian, 2009,URL : <http://www.theguardian.com/media/2009/jul/06/roger-federer-wimbledon-ratings>

Vlastakis, N., Dotsis, G., & Markellos, R. N. (2009). How efficient is the European football betting market? Evidence from arbitrage and trading strategies. *Journal of Forecasting*, *28*(5), 426-444.

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

Williams, University of Notre Dame , URL : <http://www3.nd.edu/~rwilliam/stats3/L05.pdf>
Wolfers, J., & Zitzewitz, E. (2004). *Prediction markets* (No. w10504). National Bureau of Economic Research.

Wu, J. G. (2006). Divergence of Opinion, Arbitrage Costs and Stock Returns.*Arbitrage Costs and Stock Returns*

Xuezhong, H., & Lei, S. (2012). Disagreement, correlation and asset prices.

Zeelte, A. (2012).Explaining the Favourite-Longshot Bias in Tennis:An Endogenous Expectations Approach. University of Amsterdam

Nederlandtalige samenvatting

Deze thesis begint met een beschrijving van de Efficiënte Markt Hypothese. Een van de vastgestelde onregelmatigheden in de financiële markt is "*the dispersion effect*" . Er bestaat bewijs voor de aandelenmarkt dat de toekomstige rendementen van een aandeel lager zijn naarmate het meningsverschil tussen analisten over het betreffende aandeel groter is.
Om de aandelenmarkt op zijn efficienctie te testen duiken heel wat problemen op. Eén van die problemen is dat er geen vast punt in de tijd is waarop de werkelijke waarde van de onderliggende onderneming gekend is. De gokmarkt wordt vaak gebruikt om marktefficienctie te testen. Dit komt omdat het enkele belangrijke eigenschappen met de aandelenmarkt deelt, maar bovendien een stuk eenvoudiger in elkaar zit. Zowel beleggers als gokkers hebben het zelfde doel voor ogen, namelijk winstmaximalisatie. Bovendien vindt er zich ook een transactie plaats tussen 2 deelnemers, namelijk de gokker en het gokkantoor. Het gokkantoor plaatst weddenschappen die aangegaan kunnen worden door de gokker. Het bedrag dat de gokker kan winnen wordt bepaald door de prijs van de *odd*s. *Odds* kunnen gezien worden als de voorspelling voor een bepaalde uitslag van een sportweddenschap.
De gokmarkt is ook een stuk eenvoudiger dan de aandelenmarkt. Eens de wedstrijd gespeeld is, is het resultaat van de weddenschap gekend, en kan gekeken worden of de geplaatste odds van de gokkantoren wel overeen kwamen met de werkelijkheid. Er zijn verschillende gokkantoren actief zijn. Die gokkantoren plaatsen allemaal weddenschappen aan tegen bepaalde *odds.* Ondanks het feit dat het over dezelfde weddenschappen gaat, kunnen er toch verschillen opduiken tussen de verschillende odds. De mate waarin deze geplaatste *odds* door gokkantoren verschillen, wordt in deze thesis omschreven als *bookmaker dispersion*. Er waren namelijk odds van 5 verschillende gokkantoren beschikbaar voor de analyse van dit onderzoek.

Het onderzoek begint met een beschrijving van de meest robuuste afwijking met betrekking tot efficientie in de gokmarkt, namelijk de *favorite-longshot bias.* De omvang van die afwijking wordt vergeleken tussen mannenlijke en vrouwlijke tennis matchen. Er werd verwacht dat de omvang van de *favorit-longshot bias* zou verminderen voor mannenmatchen, omdat die voorspelbaarder zouden zijn dan vrouwenmatchen. Er werd echter geen bewijs gevonden voor deze veronderstelling.

Ten tweede, Fama (1970) verklaarde dat een markt efficiënt is als de prijzen van de aandelen een weerspiegeling zijn van alle informatie die beschikbaar is. De tegenhanger voor de aandelenprijzen in de gokmarkt zijn *odds* geplaatst door de gokkantoren. De gokmarkt kan dus als efficiënt worden beschouwd als de odds alle informatie bevatten om de kansen op een uitkomst van een tenniswedstrijd te voorspellen. In deze thesis wordt een basis voorspellingsmodel opgesteld, dat enkel het gemiddelde van odds van de 5 gokkantoren gebruikt om een tennismatch te voorspellen. Dit basismodel wordt vergeleken met een uitgebreid model dat ook de grootte van de spreiding tussen die *odds* meet, namelijk *bookmaker dispersion*. Gebaseerd op statistische waarden is bewijs gevonden dat *bookmaker dispersion* statistisch gezien het basis model verbetert. Echter, deze statistische verbetering heeft niet geleid tot een betere voorspelling van de uitslag van een tennis match, of tot betere financiele resulaten voor de gokker.

Voor het onderzoek van de derde hypothese werd de brug geslagen tussen de financiële markt en de aandelenmarkt. Net zoals in de financiële aandelenmarkt werd verwacht dat een hogere dispersie tot lagere toekomstige rendementen moet leiden. Daarvoor werd een unieke methode ontwikkeld om tot subsamples te komen volgens *bookmaker dispersion*. De ontwikkelde methode was noodzakelijk omdat het onderzoek bij een gewone opdeling volgens *bookmaker dispersion* in het vaarwater komt van de *favorit-longshot bias*.
Er is bewijs gevonden dat het "*dispersion effect* " ook aanwezig is in de tennis gokmarkt. Dit een onbeschreven fenomeen in de literatuur. Uit de resultaten van dit onderzoek blijkt dat een hogere *bookmaker dispersie* leidt tot lagere toekomstige rendementen voor de gokkers en vice-versa. Het zou kunnen verklaard worden door verschillen in mening van de publieke opinie over de uitkomst van een tennismatch. Om tot eenduidelijke verklaringen te komen is er echter meer onderzoek nodig.

Tot slot werd een gokstrategie ontwikkdeld in deze thesis. Deze strategie maakte gebruik van historische gegevens van het jaar 2006-2012 om het model te ontwikkelen. Met behulp van de strategie werden positieve rendementen gerealiseerd voor een out- of- sample jaar

2013 . Dit impliceert schending van de *weak form effieciency* in de tennis gokmarkt. Echter, deze rendementen waren niet significant positief.

Appendices

Appendix 1: Summary of variables used in the regressions

Variable	**Description**
Dependent Variable	
Win	binary (0/1)
Return	(AverageOdd*Win)-1
Independent Variable	
ImpliedProbability	1/AverageOdd
Male	binary (1/0)
Male*ImpliedProbability	Male*ImpliedProbability
BookmakerDispersion	standard deviation(odd1;odd2;odd3;odd4;odd5)/AverageOdd