Master Thesis

MARKET EFFICIENCY: AN ANALYSIS OF THE INTERNET BETTING EXCHANGE MARKET

M.Sc. Finance and Accounting (Cand.merc. FIR)

Copenhagen Business School - Institute of Accounting and Auditing

Author: Frederik Grønbek Haahr
Supervisor:
Søren Agergaard Andersen

Submission: December 2011
Pages: 71
STU: 162,332

EXECUTIVE SUMMARY

Internet betting exchange is an emerging market where people can trade odds. The stock market has been the focal for many studies. This paper uses experience from the stock market together with behaviour finance theory to investigate the efficiency of the odds market. The effective market theory exists in different forms. The objective of this paper is to confirm or reject a weak-form efficiency on the odds market. The goal is to locate circumstances where the odds differ from their fundamental value by using patterns in the odds development. That kind of study is called technical analysis.

Behaviour finance theory operates with different concepts. Primary it clarifies how people behave in different circumstance and when they are likely to misjudge the likelihood of an event. That is the foundation for testing different hypothesis.

The hypotheses are used to select specific odds that meet predefined criteria. In an efficient market a wager on all selected odds should provide zero yield. If there exist situations where this is not true, it is a violation of the weak form efficiency.

Two factors are particular noteworthy in finding violations of market efficiency and used correctly, it is possible to generate a yield. The odds at the extreme, that is below 1.2 and above 10.5, differs significantly from the fundamental value. From a buyer's perspective, the return on low odds is positive, while negative on the high odds. It is constable returns.

The other factor is volume. Low volume indicates that the odds are too low, low volumes are matched wagers below 100. Buying the odds when the matched wager is low generates a negative return. At all levels above 100 the return from buying is positive. A large part of the transactions are carried out at a low volume. Giving all transactions in the sample, over 480,000 transactions, there is relatively strong evidence that buying generally generates a negative return. The returns generated by both factors are still valid giving transaction cost. Trends before kick-off is third factor that gave some interesting results, even though it is not one where it was possible to determine a yield. A positive odds development in the time period before kick-off is more likely on a losing player. This result can be used to better predict who is going to lose the game.

There is found evidence that it is possible to generate a return on the odds market by examine historical data. This is only possible when the market violates the weak form of efficiency. The market is not efficient in a weak form.

Table of contents

1. Introduction

1.1. Preface

Across the world people try to create a fortune on the stock market. This is a game where only the best are successful in the goal of earning an abnormal return. Researches throughout the history have tried to find patterns in the stock market. Sometimes the objective has been to identify forces that drive the stock prices. Other has dedicated their research to discover trading strategies that enables them to gain an abnormal return.

From an investors point of view, such findings are much more valuable. The research in this field has been the basis for developing models that explain why different stocks provide different returns. Two of the most famous models are the Capital Asset Pricing model and Farma and French three-factor model. These models are scientifically accepted, and are fundamental in understanding asset pricing at the universities.

Not all dogmas have gained as much credibility from the scientist. Technical analysis is based on the assumption, that it is possible to predict the further stock price by analysing charts. My experience is that people generally accept that the stock market satisfy the Efficient Market Hypothesis in weak-efficient form (all public information are reflected in the stock price).

If the market is weak-efficient, there should not be any patterns in price fluctuations. It is still heavily debated if that assumption is true. In the long run I believe there is consensus concerning the difficulty in achieving an abnormal return on the stock markets.

A relative new theory could be convenient in exploring superior trading strategies or patterns.

Why use what seems like random empirical methods to find anomalies when there is a set of theory called behavioural finance that deals with human behaviour. Behavioural finance outlines how people behave when they are confronted with financial problems. It has identified a variety of situations where humans behaved irrationally. The patterns in which we behave are described by psychology phenomena's. The theory behind behavioural finance in general has introduced a new vocabulary in the financial world.

This thesis is not another analysis of the stock market, although it will find much inspiration from the research made on the stock markets. Instead behavioural finance will be the fundament to identify situations where people behave irrationally or predictably, and hopefully that can lead us to profitable trading strategies on the betting markets.

In recent time new platforms have been developed to wager on sport events. The difference is that people can enter wagers with others on the Internet in a professional environment. When the market opens, people can trade as much as they want. Equivalent to the stock market the prices on odds fluctuate, and are priced by the principles of demand and supply.

The host provides the platform and the players (comparable to traders in the financial markets) sell and buy odds at prices they determine. There is a bid price, an ask price and a spread between the two prices until they overlap and two have entered a wager.

This is a bit different compared to the traditional way of gambling on sport events. Normally you buy odds predetermined by a bookmaker. This means selling an outcome is not possible, only buying an outcome. One of the largest operators in the new sports trading industry is Betfair (www.betfair.uk).

In this thesis all data are based upon real markets transactions obtained from Betfair. The market for wagers opens roughly 24 hours before the sport event actually starts, and closes when the event has finished. When the event progresses real life trading is based upon how the event unfolds which ultimately changes the price continuous throughout the event and until a winner is determined. Obviously fluctuations are determined by the performance of the players involved in the sport events. Prices on a player or team will decline when the majority of money is being placed upon this player or team. The reason for this would most likely be that the market raises the winning potential of the player or team, which then results in lower odds on this outcome. The question of interest is efficiency. The odds should reflect the real probability of winning. E.g. if you cast an honest coin there is a 50/50 chance too hit head or tails, and the efficient odds should be 2:1 if you made a wager on an efficient market. In the odds world the ":1" is implied. When you buy you win the odds otherwise you lose one.

In the same way as with the coin, the odds should reflect the underling possibility of winning. Perhaps there are circumstances where the underling probability is not precisely reflected in the odds, hence the odds is either too high or too low.

If somehow it is possible to identify situations where the odds are not reflecting the possibility for a certain outcome, it should be possible to profit from this.
Betting on sport events are a zero sum game where no value is created. If one loses a 100 $ bet another one gains the 100 $ (excluding transaction cost). Hence the expected return is zero minus transaction cost. On Betfair the transaction costs are paid when the event is over. Only winning positions are subject to deductions trough transaction costs. The fee is 5 % (if you are trading excessively the percentage declines in your favour). Excluding transaction cost an abnormal return would be a return on everything above zero, in other words all positive returns are abnormal returns.

1.2. The scope of the investigation

There are different ways to analyse if something has the right value. The financial statement analyse contain a special technique called the fundamental approach. This method is popular when estimating the precise value of a stock/company. It is an in depth analyse where you try to estimate all future cash flows to the company. In this process aspects like the competitive environment and macro factors are important.

Another technique, referred to as technical analysis, does not try estimating the precise value of the asset. The main objective with this approach is to figure out if the stock will rise or fall over a short period of time.

When thinking about it in these terms, the fundamental approach corresponds to the exact possibility of a player or team winning the sport event. We are not interested in finding methods to estimate the precise odds. We are only interested in figuring out if the traded odds are too high or too low.

The objective is to develop a profitable trading strategy. It should be possible if the market violates the weak-efficient hypotheses. The only information important when doing technical analyses is fluctuations on the odds. The technical analyse recommends our future trading strategies based upon how the historical development of the odds has been. E.g. maybe for some reason the odds are general too high or too low at the start of an event. If we were to observe such a historical pattern, we could determine which strategy at this time is most profitable, selling or buying. Basically the objective is to find a small bias in the way people estimate the odds.

1.3. Reasons for undertaking the study

As mentioned earlier numerous studies concerning the stock market have been conducted. There has been produced plenty of inspiring material that could be useful in other circumstances, e.g. the odds market. It looks as if there is a big potential to discover new interesting knowledge by using these theories and methods on other trading markets.

Betfair was founded in 2000 and is one of the first to provide an Internet betting exchange platform (Betfair.uk). The market is relatively young and studies concerning the betting exchange market are atypical, so the potential to gain new insight is quite overwhelming. I believe this study is interesting and worth further examining. It offers the opportunity to use the economic theory in a different perspective. I will investigate if

the efficient hypothesis also applies on the sports trading exchange (Betfair). A study like this, would possible be able to deepen the understanding of the stock markets in a way which had never been realized before. Correlations might be drawn from the two markets (stock and sports trading markets), creating new insight of both markets.
Second it would be satisfying and perhaps also profitable to discover violations of the efficient market hypothesis.
It would also be interesting to investigate if the effective market hypothesis in weak form applies in a market that is much smaller compared to the stock market.
You could argue establishing an effective market comes at a price. Who is going to keep the market effective if the gain does not exceed the cost of doing the analysing?
Since the betting markets are smaller than the stock markets, this study might give us some knowledge of how small a market can afford to be before the effect of efficient markets diminishes. Many articles discuss if the effective market is a valid assumption e.g. Subrahmanyam, A (2007), Fama, E. F (1998) and consensus has yet to be reached.
In recent time more and more studies questions the effective market hypotheses and this study could either strengthen or weaken the credibility of the effective market hypotheses.

1.4. Assumptions made in the study

The main objective is to formulate profitable trading strategies. For them to work it is assumed that our trading does not have any impact on the market value of the odds. This corresponds to the assumptions behind a perfect competition market where you are the price taker and have no influence on the price. This assumption is not crucial when the aim is identifying odds that differ from their fundamental value. The problem arises in the prospective of using the strategies.
When new knowledge is integrated on the market the odds move closer to the fundamental value and the strategy could prove to no longer be effective.
This study does not consider the impact from outside events. I do not have the opportunity to investigate how external circumstances might affect the odds. External circumstance could e.g. be weather conditions, earlier confrontations, which team/player has won the majority of matches in recent time, are the player left or right handed (in tennis such different match-ups as a left-handed against a right-handed could be an advantaged or disadvantage) and so on.

1.5. Problem statement

Behavioural finance outlines psychological theory in a financial perspective. The theory provides explanations for understanding why we observe certain behaviour. There are many examples of inadequate reasoning. Inadequate reasoning often appears in situations where outcomes or the likelihood of given events are overvalued or undervalued.

Inadequate reasoning is interesting to identify on the odds market. Especially when it causes imperfect prizing. Traders or other participants could profit using this knowledge.

This thesis main goal is to clarify whether it is possible to profit on the odds market. The investigation is based on historical data. If this turns out to be possible it is a violation of the efficient market hypotheses in a weak form. The way to generate this knowledge is by testing different hypotheses.

From the above problem description following research questions is formulated to explore opportunities to profit on the odds market.

1. Are the odds that are being observed when the game is about to begin at their fundamental value?
2. Does trading volume influence market efficiency?
3. Are the odds at extreme values either too low or too high?
4. Does a period of stabilization in the odds establish an anchoring and as result cause the odds to fluctuate too little?
    - At stabilization being a period of time where the odds do not change more than a few per cent.
5. Are there any trading patterns in the period before kick-off?
    - That would show as a positive or negative trend in the period from the first traded odds to the time of kick-off.
6. Is it possible to generate a profit from the strategy of buying winners?
    - More specific, a favourite is the player who is judged to me most likely to win, that would be the player who is traded at odds below 2.0.

1.6. Synopsis of methodology

To answer the research questions it is necessary to develop adequate methodologies. An appropriate methodology is one that reveals the intended information and provides usable knowledge about the problem/area of the investigation. Detailed descriptions of the methodologies are outlined later in the thesis.

The aim of this section is to clarify, and use what I believe to be the appropriate ontology, in the design of the methodology.

In the philosophical world, knowledge is not always seen as fixed. Rather it is based on different assumptions about the reality we live in. That is with an academic word called a paradigm. I do believe there is one overall perspective that is suitable as framework in this study.

Briefly summarized there are four different paradigms: positivism, post positivism, critical theory and constructivism. I will not discus the difference or go into a deeper discussion on each of them. I would refer the reader to Cuba, E. (1990) for a full description. The focal is not to make an epistemological review. Instead I will argue why I have chosen to believe that the knowledge developed in this thesis almost fulfil a post positivism point of view.

Post positivists believe that it is possible to gain close to exact information about a topic. They are aware that it dose not make sense to assume complete objectivity in the methodology. All knowledge is somewhat influenced by the investigator. As the supervisor of this study I believe that it is possible to develop information in an almost objective perspective, and this is really the key point. The way knowledge is developed in this study is in alignment with the philosophy of post positivism and that can give us some guidance in how to develop valid methodologies.

Other paradigms assume the investigator to be only objective or subjective. Those paradigms are not in harmony with my own belief concerning what is possible in this kind of study.

In this study there is no need to use any methodologies where a subjective decision might influence the results. The results are at risk of losing their validity when this happens. You will find that the overall methodology in the positivism and post positivism is similar. The difference between positivism and post positivism is located in the degree it is believed possible to perceive reality, also called ontology.

The overall method used to gain knowledge about a topic according to the positivists and post-positivists is basically the hypothetico-deductive approach formulated by William Whewell in 1837. The hypothetico-deductive approach includes four crucial points, they most be fulfilled before the investigation satisfy the paradigms.

- Hypothesize a conception about the data you are going to collect. This is here where the theory becomes important.
- Assume a consequence of that idea. That would be a sort of prediction. Formulate experiments to see if the predicted outcome is observed.
- Gather enough data point concerning a single measurement so it has the possibility to become valid, that is typically observations about something that is unknown, unexplained, or new.
- The hypotheses is either verified, or falsified.

The challenge is to secure a high degree of validity and reliability. This is crucial in development of high quality knowledge (Fuglsang, L. Hageorn-Rasmussen, P. and Olsen, P.B. 2007). Validity is a word that measures to what extent an argument measures the thing it claims to measure and the "truth" of the reasoning. (Fuglsang, L. Hageorn-Rasmussen, P. and Olsen, P.B. 2007) and (Woods A. S. and West M. A., 2010).
E.g. you could try to estimate the gender of a person by measuring the height and weight. A more precise method would be to make the conclusion based upon the characteristics of the body. The validity of the first method to estimate gender would be low. First because it is a bad reasoning, second because there is a better method to gain the information.
Reliability is a concept used to express if the specific method is free of any elements that could have an influence on the result. A high degree would mean that the experiment could be duplicated and still give the same result (Fuglsang, L. Hageorn-Rasmussen, P. and Olsen, P.B. 2007).
The above review taken into consideration, I find it appropriate to design the study as follows. The data gathering is done random when possible. The data is not selected because of inside knowledge that would lead to bias information gathering. Betfair is one of the biggest sites and the data is collected from that source. It has been necessary

to focus on one spectacular group. I have chosen women tennis players. As a consequence the results are more reliable in relation to this group. There is no saying that the results should apply on all sports betting, even though it would not be unlikely. The data is perceived as being very precise. The source collects information about the operators and their wagers on that precise moment in time. It cannot be questioned, so each observation is very reliable.

The methodology to analyse the data is described in depth under each research question. A high degree of objectivity is secured by being very critically in the approach. The research questions are inspired by the theory concerning financial behaviour. This ensures a high degree of validity.

Finally with the use of statistic tools the hypotheses is either rejected or accepted. In this paper the significances level is decided to be 5%, it is a common level used to decide if data material is valid.

1.7. Limitations inherent in the study

One of the biggest weaknesses of conducting a study in relation to finding a pattern in price fluctuations is the great risk of data mining. Data mining does not cover the concept of discovering patterns there are not valid, but rather the consequence of performing many different investigations on a data set.

When doing multiply investigations eventually some kinds of pattern will be revealed. If the hypothesis you are testing has no causality to the result, and is actually a coincidence, you make wrong conclusions about the reality.

The risk originates from the method using to conclude if a hypothesis is true or false. There is integrated a 5% error in the statistical test. All conclusions are at risk of being simply a coincidence. This is the primary limitation of the statistic tools when they are used to evaluate data.

Another limitation of the study concerns Betfair's method to register and deliver data related to the trading. I do not have the resources to verify that all the data is really a true reflection of the exact wagers. I am forced to assume that the data is reliable. I cannot find a good argument why this should not be the case. Only a technical challenge could be imagined e.g. if there is something wrong with their computer systems Betfair are using, but it seems unlikely. Given that the only sport that will be analysed is tennis, it is not certain that the conclusions are valid in other sport markets.

1.8. Report structure

The structure of the report is simple. The introduction and literature review is the basis for developing the problem area. These sections also clarify the background knowledge for this topic. After reading the introduction and literature review you should have a basic understanding of the theories, and be able to put the investigation into perspective. Now you should be able to follow the researchers mind-set in the rest of the study.

Figure 1.8 displays the structure graphically. There are six research questions which each are given a small introduction to create the context for the question. Thereafter follows a description of the methodology. At last I summarized the findings in charts and figures. When the data is neatly set, we evaluate the numbers and try to answer why we observed that kind of results.

The findings from the six research questions culminate in an overall discussion.

As it will be revealed later, the conclusions will be more convincing when perceived in a broader context. The procedure in which the study is conducted is also easier evaluated when comparing the findings. The discussion is the basis of approaching the findings critically. The findings are related to the theory and evaluated according to how good the theory had been explain the outcomes. It is also the place where we reflectively review to the overall protocol of the research.

If possible recommendations on how to profit on the odds markets will be explained in the conclusion.

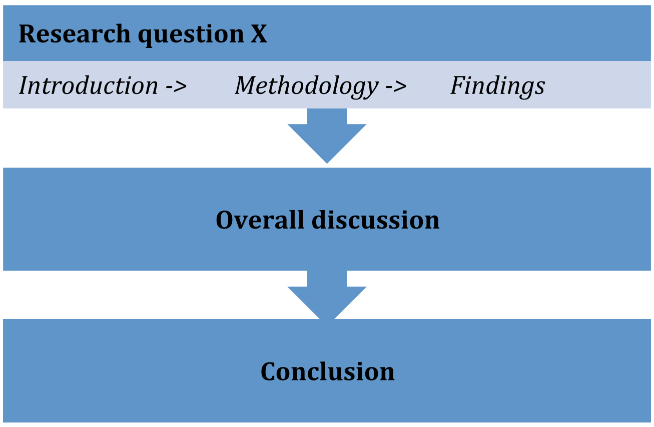


Figure 1.8 Overview of the structure in the thesis

2. Literature Review

Throughout history it has been possible to observe phenomena's on the stock market that violates the efficient market hypotheses. Real life cases are good inspiration sources and sometimes they give births to new theories, which rationalize why they happened.

Behavioural finance link psychology and finance, and many studies have showed that people do not always behave in a rational way.

This review outlines important findings related to the understanding of the human mind. It appears that we are not always as rational as we think we are.

But first there is a set of rules/observations concerning the stock market that is pleasant to have in mind when integrating the psychological aspect. It illustrates why it is difficult to formulate bullet prof statements regarding the stock market. Behavioural finance offers several explanations why we should not expect the stock market to be efficient. To use human imperfections suggested by behavioural finance can be difficult.

Some of the observations in relation to the stock market go many years back. Charles Henry Dow (1850-1902) formulated a set of rules based on his experience on the stock market. They where further developed by William Peter Hamilton. He claimed that these rules were the basic of the insight that made him able to predict the great recession in 1929, as he actually did (Tvede, L. 1999 s. 12).

The set of rules is a follows.

1. The market is ahead.

The market price reflects all available information. That includes the sum of all current and potential knowledge investors possess. All that information is often more than a single human can comprehend. Even a wise investor would still have difficulties in beating the market. That would also be true if knowing how to find ignorant participants to trade with. The reason is that they can simple observe the market price, and the market price is reflecting all available information.

2. The market is irrational

The market can be emotional. The slightest change in trends can cause people to panic. If many people traded by following the same system trends can be a self-fulfilling. The emotional aspect is what we are trying to understand with the concept of behavioural finance. In periods the impact from mass hysteria and noise traders can cause massive fluctuations. Opposed to changes in the fundamental value of the securities.

3. The environment is chaotic

The market value is influenced by macro economic factors. Forecasts of such factors are typically so vague that they become almost useless for an investor. The reason is that small changes can be of great importance. They are hard to predict and have a great impact on the estimated value. Hence, it can have widespread consequences to rely too much on forecasts. In general it is difficult to predict and integrate future events. That hurdle causes much uncertainty in a valuation.

4. Charts are self-fulfilling

If many people use the same system in stock pricing they will sell and buy on the same indicators. That kind of pressure on a stock can cause it to move in the expected direction.

Not all of these observations are equally relevant in relation to the odds market, simply because it is hard to see how they could have any influence on the price of an odds. They are merely meant to illustrate the variety of factors that can influence prices.

I think the first 2 points are important in this investigation. The others are relevant when the marketplace is more complex and the timeline of analysing is longer. Macro factors are irrelevant because of the short timeline. I have never heard that people use charts analysis on the odds market. I do not believe charts to be self-fulfilling on the odds market because there are too few people using that technique.

If the market is ahead and I believe to a great extent it is, then it is best simply to acknowledge that it is difficult to invent a system to earn money, but not impossible. The market also has is flaws and this is where the second rule and thus behavioural finance becomes relevant.

There has been carried out many experiments and surveys that help us understand beliefs and perceptions of the human mind. They can sometimes be the reason why we make unwise decisions/reasoning. I have selected the most promising ones in the pursuit to formulate profitable trading strategies on the odds market.

2.1. Framing

Framing is the concept of presenting something in such a way that it leads to a certain perception/decision. If the questions are formulated inappropriately, the responses are at risk of being biased. One has to be careful when conducting research about human

behaviour. Framing can be the reason for the behaviour/response and not the situation. This is a problem when the target is to see how people react in different situations.
*Kahnemanand and Tversky (*1981) provide many different examples of framing, some more relevant than others.
This paper is interested in developing profitable trading strategies, and framing is not directly helping us reach any conclusions.
The point of introducing framing is to illustrate the importance of being critical. The described characteristics of humans could have been influenced by the presentation of the problem. In real life there is no rules of how a problem might be formulated. The results could still by valid despite framing. Maybe the problem is likely to be presented the same way in real life.
Framing could be the reason to assume something wrong about people.
There is one example in particular that illustrate the effect of framing very precise (Kahneman, D. and Tversky, A. 1981):

> *Imagine that the U.S. is preparing for the outbreak of an unusual Asian disease, which is expected to kill 600 people. Two alternative programs to combat the disease have been proposed. Assume that the exact scientific estimates of the consequences of the program are as follows:*
> *If program A is adopted, 200 people will be saved.*
> *If program B is adopted, there is 1/3 probability that 600 people will be saved, and 2/3 probability that none will be saved.*

Compared with another formulation of the choices:

> *If problem C is adopted 400 people will die.*
> *If problem D is adopted there are 1/3 probability that nobody will die, and 2/3 probability that 600 people will die.*

When the subjects are asked using the first formulation, 72% thought best of program A. In the other situation, which offer precise the same outcomes only formulated in a different way, 22% chose program C (the consequences from choosing A and C is similar). The difference is striking, and it clearly illustrated the importance of framing.

2.2. Overconfidence

People can be overconfident in different ways. Meaning the characteristics of overconfidence behaviour is divided into several categories. The experiments used to identify overconfidence have different approaches.

Miscalibration is related to overestimating the preciseness of the knowledge an individual believes to have. People are given different quantitative questions and asked to provide confidence intervals on their answers. That kind of experiments shows that people are overconfident in the intervals they report (Alpert, M., and H. Raiffa, 1982). The interval is to narrow, thus their guesses are not so correct as they think they are. Based on the confidence intervals you expect that some of the answers lie outside the interval. The point is that there are to many answers that lie outside the confidence intervals. People's weakness for estimating distributions, have been reported in various test. What people are certain about, happens only 80% of the time, and what they think is very unlikely happens 20% of the time (Fischhoff, B. P. Slovic, and S. Lichtenstein. 1977).

The Better than average effect: when people are asked if they are better than average on some personal attribute, more then 50% think they are. But only 50% can in fact be better than average. There must be some who are too confident in their own skills and think they are better than they really are. A group of students were asked if their driving skills were better than average. 82% reported that it was (Svenson, O. 1981). Just to give one example out of many.

Excessive optimism is related to the circumstance where people think that a positive event is more likely to happen than it really is, and a negative event less likely to happen than is the case, e.g. winning in lotto and the risk of getting cancer (Ackert, L. F. and Deaves, R. 2010).

Planning fallacy is close related to excessive optimism. The concept covers the idea that we think we are able to accomplish more than what we actually end up doing. This is normal when people design a "too do" schedule with too many activities. At some point in the process we realise that we simply do not have enough time or resources to accomplish what we thought was possible (Ackert, L. F. and Deaves, R. 2010).

2.3. Representativeness heuristic

When people are asked to estimate the likelihood of a circumstance given some information about a population, they are not very good at providing a correct answer. The problem is that people rely too much on their own beliefs and ignore important information regarding the population. The explanation could be that they seek an answer based on comparable known probabilities.

The process where people use experience and fails to analyse the situation correct, is called representativeness heuristic (Kahneman, D. and Tversky, A. 1974). It is interesting that people rely so heavily on representativeness heuristic.

Normally people whose judgement is heavily influenced by stereotype perceptions are influence by representativeness heuristic.

A good example is when people are asked to judge whether a person is a lawyer or engineer. They are giving on some neutral characteristics that have no influences on the chances of the person being lawyer or engineer (natural because given no additional information there is 50 % chance of that person being a lawyer and 50% change of being a engineer). Still people are likely to think that the possibility for the person being a lawyer is the same regardless of the population consist of 30% lawyers and 70% engineers, or 70% engineers and 30% lawyers (Amos T. and Daniel K., 1974). They base their answer on the neutral information instead of using the knowledge about the population. The point is that people use their own perception instead of the facts to answer the question. That is the core of representativeness heuristic.

According to Bayes' theorem, information about the populations is crucial when one has to estimate such possibilities. In other words, the distribution of the population is of great importance. There is a big difference in the possibilities depending on whether the sample originates from a population consisting of 70% lawyers or 70% engineers. Basically people do not know enough about Bayes' theorem. If they did they would know the importance of integrating the distribution of the population in the solution. Lets take a look of Bayer's theorem and how it should be used.

Bayes' Theorem expresses the relationship between a conditional probability and its inverse. That is, it provides the relationship between P (A | B) and P (B | A) (Agresti, A. And Franklin, A. C. 2007).

Using words it sound like: given that the person is a man (B) the probability that he is also an engineer (A) is P (A | B). The inverse would be given that the person is an engineer the probability that he is also a man is (B | A).

Bayes' Theorem: (Yudkowsky S. E)

Simple form

$P\ (A\,|\,B) = P\ (B\,|\,A) \times P\ (\ A) /\ P\ (B)$

Lets continue the previous example with the layers and engineers. In this example the natural characteristic is the sex. Assume there are equalled many women and men who are lawyers and engineers. When using the logic of Bayer's theorem it becomes clear that it is wrong to ignore the distribution of the population. The answer should not depend on whether you think there are more women who are lawyers or otherwise.

Facts about the population:

	Lawyer	Engineers	Total
Male	350	150	500
Female	350	150	500
Total	700	300	1000

Notation

P (A) Percentage of the population who are engineers 30%.

P (B) Percentage of the population who are males 50%

P (B | A) Given, that the person is an engineer, the probability is that he is a male, is 50%

Using Bayer's theorem we can calculate the possibility the person is an engineer given that we know the sex. P (A | B) = 0,3×0,5/0,5=0,3. In the opposite situation where there are 150 lawyers and 350 engineers, the probability that the person is an engineer when knowing the sex would be: P (A | B) = 0,7×0,3/0,3=0,7

It should be clear that only the distribution of the population is important. The likelihood that the person is an engineer is greater if the population consist of more engineers. It is wrong to ignore this information and base the answer on representativeness heuristic. That is you think more than 30% of men are engineers because you know more male engineers (population of 30% engineers).

2.4. Insensitivity towards sample size

A similar misinterpretation of representativeness is insensitivity towards sample size. People tend to misjudge if the sample is more likely to originate form one population or another. They rely too much on the sample ratio and compare it to the population without taking the size of the sample into consideration.

E.g. imagine an urn filled with balls of which 2/3 is of one colour and 1/3 is of another. You are then asked to assess which of the following 2 incidents would give the highest possibility to guess the right proportion of colours.

First incident, a draw that revels 4 red and 1 white ball.

Second incident, a draw that revels 12 red and 8 white balls.

Both incidents make us believe that the urn consists of 2/3 red balls and 1/3 white balls. The point is that many think that the first incident is less likely to happen in the situation where 2/3 of the balls are white (Amos T. and Daniel K., 1974), but that is wrong. Actually it is the second incident that is less likely to happen, which means that the second incident is a better representative and the conclusion should be based on that incident[1].

2.5. Misconceptions of chance

People think that a sequence is more likely to appear if it is close related to the underlying possibility. If there is a 50/50 chance on 2 different events, the possibility of a changing sequence is assessed too high (Amos T. and Daniel K., 1974).

E.g. with heads and tails people think that the sequence H-T-H-T-H is more likely compared to H-H-H-H-H. The truth is that each sequence is equally likely. In the long run it is true that you are more likely to observe an outcome with a distribution similar to the underlying possibilities, but this has nothing to do with the specific sequenced, all sequences are equally likely.

2.6. Availability biases

The risk of availability bias occurs when people use their memory/imagination to decide which of different events occur most frequently. Typically the answer is influenced by what is most retrievable from memory. What appears to happen most frequently must be most likely. This approach has limitations. The memory does not always provide us

[1] Appendix 1 shows how to make the calculation and the exact answers.

with sufficient information to assess what is most likely. The bias in the assessment is explained with the fact that not all memories are equally represented. The easier the event is to retrieve from memory, the more emphasis it gets in the judgement.

The logical explanation is that not all information is equally stored in the brain. Things we can relate to, or which make more sense, is easier to remember and more salient in our memory (Amos T. and Daniel K., 1974). This infers with the judgement of occurrences of events. The consequence is that easy memorable events are stronger represented in the process of deciding what is most likely. This phenomenon is called availability bias.

An example of availability bias appears when people have to decide what occurs most frequently? Is it words that begin with the letter r (e.g. real) or, is it words that have the letter r in the third position (war). Most people are prone to answer that most letters begin with an r. The hypothesis for this observation is that it is easier to produce words that begin with an r. If letters that begins with an R appears to happen more frequently you naturally think this is more likely. In fact most letters have an r in third position. People make the wrong conclusion because it is more difficult to produce words that have the letter in the third position (Amos T. and Daniel K., 1974).

Availability biases could play an important role when deciding the size of odds. If it influences the odds in a somewhat skewed degree there is a basis of making a profitable trading strategy. The implications could be that the player who has been exposed for most positive publicity is overestimated and as a consequence the odds is sat too low.

Economists are not sure in which degree availability biases actually happens in professional environments. Experiments shows that people have this tendency but it is not certain that an availability biased will continue. People could learn though repetition that their memory is not the perfect tool when they have to decide which events is most likely (Barberis, N. C., and Thaler. H., 2003).

Experts happen to be more immune against availability biased memory. In the long run they have to base their decisions on objective criteria.

2.7. Anchoring

When people have to estimate something quantitative, initial information can have a high influence on the response. Experiments show that people are less inclined to diverge from initial information and from the name anchoring. The reason could be that people form their estimate on the basis of the initial information and then adjust away from it. If this is true, the adjustment is not always sufficient. There are several circumstances where imperfect "adjustment" appears. An example of anchoring is when decisions are based on some incomplete computation. E.g. when the subjects had to make a calculation based on the two different sequences:

1x2x3x4x5x6x7x8

Compared to,

8x7x6x5x4x3x2x1

The beginning of the sequences had a significant impact on the responses. In the experiment, the first sequence had a median on 512. The other sequence resulted in a median on 2250, a huge difference (the right answer is 40320) (Amos T. and Daniel K., 1974). The logic is that the first part of the sequence is calculated and the rest is estimated based on the first few calculations.

Anchoring in conjunctive and disjunctive events. Maya Bar-Hillel performed a study in 1973 to test anchoring. It showed that people have a tendency to evaluate conjunctive events as most favourable, and disjunctive events as less favourable, when in fact they are not.

In the experiment people had to choose between 2 bets. In one scenario it was between a neutral bet and a conjunctive bet. In the other scenario it was between a neutral bet and a disjunctive bet. The difference in the design of the options was more important than the likelihood of wining the bet. The neutral, conjunctive and disjunctive bets were as follows:

Bet 1 (neutral bet): drawing a red ball from an urn with 50/50 red and white balls (50% chance).

Bet 2 (conjunctive bet): drawing seven red balls in a row from an urn with 90% white balls and 10% red ball. The ball is replaced after each draw. The chance of success in drawing 7 red balls in a row is 0,9^7=0,48.

Bet 3 (disjunctive bet): drawing at least once one red ball in seven attempts from an urn containing 10% red and 90% white. The ball is replaced after each draw. The chance of success is 1-(1/9)^7=0,52.

The subjects where asked to chose the most favourable bet.
In one situation they should choose between the neutral bet and the conjunctive bet. It turned out that the subjects were prone to choose the neutral over the conjunctive bet. When calculating the exact possibility it would have been best to choose the neutral bet.
In the other situation subjects should choose between the disjunctive bet and the neutral, here they chose the disjunctive bet, again the least favourable bet. Thus in both circumstances subjects in general choose the worst bet. Anchoring could explain why people choose the worst bet in both situations.
In bet 2 there is 90% chance to draw a red ball compared to only 50% in bet 1. The conjunctive bet is rated too high because of the high possibility for success regarding the elementary event.
In the other situation the likelihood of the elementary event and thus drawing one red ball is 50% versus 10%. The disjunctive is rated too low because of the low possibility of success regarding the elementary event.
When the subjects based their assessment on the elementary event the adjustment from the elementary event into a sequence is insufficient. This is because of anchoring, people do not differ enough from the starting point (Amos T. and Daniel K., 1974).
Around kick-off you know the odds on the players. Think of tennis games as a series of conjunctive events. At any given time the score in the match is dependent on what have happened in the game up to that point.
The odds around kick-off could work as an anchor. The adjustment from the starting odds is maybe not sufficient, as the match progresses and the odds change. If there is establish an anchor to the odds at kick-off the odds are not fluctuating enough.
If that happens in real life, a profitable strategy would be to sell the current favourite if she was also the worst seeded player. We expect that the odds are too high on that player. To some degree people still think she is not going to win.
This is maybe an over-interpreted version of anchoring, and it suffers from the same argument as earlier.

People who gain more experience should be less reluctant to anchor. It is impossible to say if traders on Befair are experienced or not, but they could be, and then the effect of anchoring should be minor.

The experiment above also lacks from a “how much do I want to go in depth with this problem“ constrain. Peoples choice could be influenced by anchoring because the problem is a one time assessment, it is not worth to do an in depth analysis. This is probably different if it was thought as repeated business. There an in depth analyse is more valuable. It is naturel to assume that the analysis is more precise when you are exposed to the same problem repeatedly. Again, anchoring should not be an issue in professional environments.

2.8. Belief perseverance

People tend to stick to their initial point of view although they are presented with new information that strongly contradicts that perception. People place less emphasis on empirical evidence that challenge their present belief, and seem to be more prone to remember the strength of confirming information and the weaknesses of disconfirming information.

A study perform by Lord C. G., Ross L., and Lepper M. R. 1979 strongly supports the hypothesis that new information is interpreted in order to maintain initial beliefs.

The overall methodology of the study was as follows: two groups were formed based a prior questionnaire concerning their attitude towards capital punishment and its deterrent effect on murder. The subjects where dividend into two groups: proponents who favoured capital punishment and believed in its deterrent effect, and opponents who opposed capital punishment and hence doubted its deterrent effect.

A few weeks later the two groups gathered around a table. They were asked to evaluate some information regarding this topic. They were presented with some brief statements inherited from different studies. Some of the statements were in favour of capital punishment and some against. After each statement the subjects were asked to evaluate the new information from -8=more opposed, to +8=more in favour compared to their initial belief.

Afterwards they were presented with procedural details, critiques and rebuttals, obviously to question the validity of the information. They were then asked to rate the quality of results they had received in the first place.

The results clearly showed that the initial position had a significant influence on how positive or negative the new information was perceived and subsequent rated. Proponents rated both the evidence more convincingly and the quality of the study positive when regarding pro-deterrence and negative in the light of anti-deterrence.
The opposite was true when opponents on capital punishment judge the statements and its reliability.
The experiment also showed that the subject was more polarized after they had been presented with the different statements, that is, they where more in favour or against capital punishment compared to their initial point of view. It confirms that the subject had assessed the new information unequally and in favour of the initial point of view.

The study was repeated, with one important difference. The new fictitious statements reported opposite results compared with the first set. In this situation it would seem more logical to doubt whether one's perception is really true, or if it should be re-evaluated. This was not what happened. The subjects continued to preserve the initial belief even when they received the new information, or when the information they processed was dubious.
The important observation is that to some extent people do not differ from, or question their original point of view even when, objective speaking, it would be most rationally.
If belief preference exists on the odds market, the initial assessment has an impact on the odds later on. There is too much confidence in the initial favourite even when the status of the game indicate that she is not so likely to win as she was at kick-off. The consequences would be that the demand for the best-seeded player is too high compared to the odds, and too low for the other player. The skewed demand causes a reduction in volatility.

2.9. Prospect theory

In the pursuit to profit on a speculative market it is advantageous to understand investors trading behaviour regarding losses and gains. This topic has been difficult to grasp, and for a long time there has not been any clear guidelines. There are many theories trying to describe how people behave in situations where they are confronted with gain or losses (Barberis, N. C. and Thaler. H. 2003).

The Expected utility framework is probably the best known. It deals with the risk in a final wealth perspective. In this perspective only the final positions are taken into account. The procedure to estimate which of several possibilities is the best, builds on assumptions about preferences. The total sum of money in the final stage is "translated" into utility and weighted by its respective possibility of occurrence. It is nice to assess preference in quantitative measures, that way it becomes possible to rate different outcomes and thus predict what people will do.

E.g. no bet, or a head and tails bet with 1000 on stake and an initial wealth on 10000:

No bet: U(M)=1*(10000^0,)=100

Bet: U(M)=0,5*(11000^0,5)+0,5*(9000^0,5)=99,87

In the above example with the given utility function, it is best not take the bet. Thus we predict that a person with that utility function will refuse the wager. The problem with this approach is that it is not very good at explaining the behaviour that is observed in practise (Kahneman, D. and Tversky, A. 1979).

With Prospect theory the approach is different. It is based on the decisions humans actually make in different situations.

Investigation shows it is more reasonable to conclude that gains and losses should be dealt with separately, and not in a final wealth perspective as is the case in the Expected Utility framework. Here is reviewed some of the important experiments Kahneman and Tversky conducted to establish the fundament of prospect theory (*Kahneman, D. and Tversky, A. 1979)*. They are quite interesting and good at providing an insight in the theory in general. In the end it should be clear that the experiments offer several violations on the expected utility framework.

Definitions:

(x,p; y,q), this should be read as, get outcome x with probability p, or outcome y with probability q.

Experiment 1:

In addition to whatever you own, you have been given 1000. Now choose between:

A=(1000, 0,5); B=(500, 1)

The same subjects were then asked:

In addition to whatever you own you have been given 2000. Now choose between

C=(-1000, 0,5); D=(-500, 1)

In the first situation most of the subjects thought best of outcome B. In the second situation it was outcome C. The difference was quite strong. This is a clear violation of the expected utility framework. If you think about it, you realize that the 2 situations result in the same output (2000, 0,5 ; 1500; 1). Therefore the distribution of the answers should be the same if people thought about in a final wealth perspective, but its not.

The goal with the next experiment was to determine the reduction in desirability when the outcomes are altered from certain and most likely to only probable outcomes. It illustrates that people prefer outcomes that are certain relative to outcomes that are more risky.

Experiment 2 (*Allais, M. 1953)*

A=(4000, 0,8); B=(3000, 1)

or,

C=(4000, 0,2); D=(3000, 0,25)

People chose B over A and C over D, but why? It should be clear that in both alternatives the only difference is a 20% decrease in probability. The inconsistency violates the expected utility framework[2]. If A maximise the utility in the first place then we would also expect people to chose C in the other situation.

The explanation should be found in the way people relate to changes in probability. A 20% decrease to 0,8 seem more striking than a 20% decrease from 0,25 to 0,20 (Barberis, N. C. and Thaler. H. 2003).

People have a tendency to be risk averse regarding gains, which are verified by the subject's choice of B. When the chance of success is low choice C seams better, and people become more risk seeking. In both situations people choose the lowest expected value. The 20% difference is maybe not perceived as significant compared to the first incident, and the prospect of receiving 4000 suddenly seams more attractive.

[2] Appendix 2 explain the mathematical logic

This kind of behaviour, where people is changing their preference when the chance of success is high, is known as the certainty effect. People like certain outcomes.
In Kahneman and Tversky's paper they outlined several other experiments where the subject should deal with uncertainty. The results all indicate the same behaviour.
Another kind of perspective is to see what people choose when the uncertainty concerns a loss. Prospect theory predicts a big difference in the how people behave. People have another way of dealing with losses compared to gains.
The following experiment show how people behave when they are facing losses.

Experiment 3
A=(-4000, 0,8); B=(-3000, 1)
or,
C=(-4000, 0,2); D=(-3000, 0,25)

If people had the same attitude towards gains and losses we would expect to see the same choices. This is not the case. When the situation is negative, choice A is preferred over B and choice D over C. People are inconsistent in their choices. They change their preference from a certain positive (3000) outcome to an uncertain negative outcome (-3000, 0,25). Their preferences are implying a risk averse behaviour regarding gains and a risk seeking with respect to losses. This is a valid trend in most of the experiments. It is not possible to reach any conclusion about risk averse behaviour based on the preferences of option C and D in experiment 2 and 3. In the positive situation C has the highest expected value but a smaller chance of success. If people are risk averse they could still choose this option and be in alignment with the risk averse hypothesis. The trade-off is sufficient to make them choose the more risky circumstance and thus they still prefer option C. In the negative situation the subjects chose option D. That does not exactly look like risk seeking behaviour since it has the highest expected value. The reason could be that there is a sufficient trade-off between higher expected value (choice D) compared to the lower possibility of a loss (option C).

The overall observation for all the experiments is that people tend to do the opposite in negative situations compared to positive situations. This is categorized as the "reflection effect" by Kahneman and Tversky (1979).

In theory there could be several implication of prospect theory that could cause the odds to differ from their fundamental value.

People who have suffered a loss will continue to keep the odds. Implying a risk seeking behaviour when they stand to lose. They will not realize the loss, but rather increase the deposit.

People who have increased the value of their odds are dealing with a profit and are more risk averse. They are more likely to realize some of the profits (it is risky to let it run). Their selling will increase the odds on the favourite (what happens is they actually buy the losing player and hedge, so no matter the outcome they will win on one of their bets). The result is that the odds on the winning player is too high, (it should have been lower as a result of the greater likelihood of winning), and the odds on the losing player is too low. In other words the balance between the demand and the supply of the winning player is not in equilibrium compared to the fundamental value distribution.

2.10. Momentum

The momentum effect is an anomaly on the stock market. Using the right technique a momentum effect inspired portfolio offers an abnormal return. The trading strategy is to buy stock that has provided a high return in the previous period. The tendency that high performance stocks continue to do well is called the momentum effect.

Jegadeesh N. and Titman S. 2001 have tested for the momentum effect in two independent periods, 1965-1989 and latest 1990-1998. Both periods offers statistically proof of the momentum effect. The overall method was to form portfolios based on past returns and keep track on how the stock performed in the next period of time.

Research shows that a portfolio composed by stocks with high returns in the previous 3-12 months was superior. The portfolio continued to outperform the market in up to 12 months before a reversal was observed. In the subsequent period of 13- 60 months the portfolio offered a worse cumulative return compared to the markets index. This is not surprising. At some point in time we do expect the impact of another well-documented tendency on the stock market.

In the long run, stocks do not continue there successfully or unsuccessfully pattern. The return on high and low performance stock will at some point in time begin to converge and both "types" of stocks will begin to perform equally well. This is called the mean-reverting effect. The logical consequences is that previously low performance stock will increase more in value compared to previously high performance stock. (Pedersen, C. V. and Plenborg, T. 2005).

Even though momentum has its limits it is still unexpected since it violates the efficient market hypothesis. The momentum effect has been met with much critique mostly because of the violation of the efficient market hypothesis.

Some thinks that the results are a consequence of data mining. The authors reject the critique because the momentum effect has been statistically documented in two separate periods. Another attack on the momentum effect is a lacking compensation for risk when comparing returns. High performance portfolios are more risky and this naturally led to a higher return in the period. It is uncertain if the accusation is valid (Jegadeesh N. and Titman S. 2001)

The momentum effect is specifically related to the stock market. The betting exchange market is a bit different in its structure. Betting is a zero sum game, and the time period of the wagers is limited. It is not possible to draw a direct parallel. If the momentum effect applies on the betting exchange market it could be observed in different ways.

First, the player who has the highest winning rate is in a good streak. If the momentum effect exists in the same way as on the stock market (a high-performance stock is undervalued), we expect it to be profitable to bet she will win[3]. She is underrated because people to do not give her high performance enough reliance.

Second, another way to interpret the momentum effect on the betting market is to explore its effect while they are playing. The first interpretation of momentum effect on the betting exchange market was based on a misjudgement before the game started. The second way will be a misjudgement as a result of the development in the match.

People will underestimate the likelihood that the player who is in a good streak will continue the good streak. Thus she has a greater opportunity of winning than the market expects. The odds are the too high. People underestimate the player with “momentum”. They think the chances of her losing are greater than what actually is the case.

[3] The data materiel do not allow for such an investigation.

2.11. Volume trading on the stock market

Financial behavioural theory it not always the main approach to find good strategies. Often it is more a question of experience. So is it with volume trading, even tough it has been connected to the psychology aspect afterwards.

On the stock market the amount/level of trading is used to indicate a trend. There is a big difference in trading activity on the stock market. The activity depends of the market conditions. Normally a bull markets is associated with higher trading activity compared to a bear markets (Tvede, L. 1999). A peak in the level of trading indicates a peak in the stock prices. When trading stagnates or decline, stocks are expected to fall within a foreseeable period of time (Tvede, L. 1999).

The concepts of overconfidence and prospect theory might be the reason why people like to trade when the market is rising. Investors like to trade when they are trading on gains (overconfidence), but they are less active if the trading involves a realisation of a loss (prospect theory). The explanation for keeping their positions is that they are risk seeking when in comes to losses. They hope their position will increase in value, so instead of selling it they do nothing.

If an investor in the past numbers of trades has been successful and has increased the value of the portfolio it is no wonder that he get more confident in his actions. More confidence has the natural consequence of increasing trading activity. The trading activity should peak when the stock market is at it highest value. When the market begins to decline people are losing money. They do not like to claim the losses, as result they reduce the amount of trading. If we transfer this behaviour to the odds market, high trading volume indicates that people are comfortable buying the odds. The demand should then increase to a level where it pushes the odds away from its fundamental value. It forces the odds to be valued too low, and the gain from buying should be negative.

2.12. Capture of behaviour finance and its assumed impact on the odds

Background theory is an important factor in the process of creating relevant hypothesis and a good problem statement. To develop a profitable trading strategy, investigation of the most plausible violations of the effective market is the most promising approach. This paper has little interest in testing if the theory is valid. Our aim is to identifying imbalances in they way the odds are priced.

Table 2.1 summarized the individual component in behavioural finance and its expected impact on the odds. The table illustrates that not all theories predict the same impact on the odds. Some of the theories expect people to underestimate and others to overestimate the odds. There are also examples of phenomena's that does not provide any conclusive trend. The theories clearly show that human's decision-making is complex. Hopefully it is possible to locate different circumstances where one effect dominates.

TABLE 2.1

ANTICIPATED IMPACT ON THE ODDS

	No clear impact, it can cause too low, as well as too high odds.	Causes too high odds for the winning player*, thus an underestimation of the chance of wining.	Causes too low odds for the winning player, thus an overestimation of the chance of wining.	Cause to little fluctuation.
Overconfidence				■
Representativeness	■			
Availability bias		■	■	
Anchoring				■
Belief preference			■	
Prospect theory				■
Momentum		■		

*The wining players is defined is the player who have had a decrease in odds, and thus a higher chance of winning, compared to the start of the game.

The overall conclusion is that there are many factors that may have or have an influence on the odds. It is interesting that Henry Dow discovered the same many years ago. In the light of all the research that has been conducted since his time it is not surprising that the market is chaotic, after all human beings influence it. Our broad objective is to created order in this chaos. If there is no chaos then maybe Internet betting exchange is more efficient than assumed.

3. Empirical study of the Internet exchange market

The investigation is conducted from a buyer's perspective. This perspective implies that wagers relating to a winner generate a return, and wagers on losers represent a cost. In an efficient market the yield and the cost should be equal. The betting strategy should not influence the yield. This analysis will assess if carefully selected odds are either too high or too low.

It is practical to outline some general remarks concerning the empirical data and the statistical tools. They will by applied and referred to several times. Instead of repeating the remarks several times, it seems appropriate to review it only once as part of the introduction to the empirical study.

The remarks can be separated into different points.

- The characteristics of odds.
- Notation and definitions
- The choices that have been made regarding the data and sample period.
- An overview of the statistic tools to test whether or not the results are significant.

3.1. Fundamentals preceding the empirical analysis

3.1.1. Characteristics of odds

Odds are closely related to the gambling world and are probably most used on sports events. Odds are another way of interpreting/specify the probability distribution for various events. Typically there are two or three different scenarios: a winner, a loser and sometimes a tie. In tennis it is not possible to play a tie and therefore only two events can occur. Table 3.1 displays the translation of odds into its implied likelihood of the event to happen.

Odds are practical because they formulate the possibility of wining intuitively.

TABLE 3.1

THE ODDS AND ITS IMPLIED CHANCE OF WINNING

Odds	Chance of succeeding
1.0	100%
2.0	50%
3.0	33%
4.0	25%
5.0	20%
6.0	17%
7.0	14%
8.0	13%
9.0	11%
10.0	10%

E.g. when playing head and tails it is logical to formulate the wager as 2-1 read as (2 to 1) often the (1) is implied. It cost one to enter the wager, if you win you get two, and if you lose you get nothing. Winning equals net +1 and losing -1. The expected value of this bet is zero, and thus it corresponds to a 50% chance of winning. Bookmakers are often sellers. You buy the odds from the bookmaker and if you win you get the promised odds. On the Internet betting exchange it is possible to buy and sell odds.

3.1.1. Definitions

The definitions mentioned in this section concerns important aspects of calculating the yield. A clearly definition of the yield is significant. The calculation of the profit is also essential, as it is part of the equation.

The yield can be interpreted differently. Should the yield originate from the total deposit, or should only the losing wagers used in calculating the yield? E.g. you know you get all you money back when you win, both the cost of the wagers and the premium. You could argue that the money used to buy the odds on the winners is not really a cost.

I think all resources used to generate the return should be perceived as a cost. Thus the cost includes entering all wagers, and not just the cost associated with lost wagers.

If the odds are at their fundamental value, the profit from placing a bet on all outcomes should be zero. Placing a bet on the winner and on the loser clarify whether or not the odds are at their fundamental value. That approach reveals the profit from the specific trading strategy.

Premium: Sum of successfully odds minus the number of associated transactions
Return: Sum of all odds relating to the winners (that includes the money necessary to enter the wager).
Cost: Sum of all transactions. The deposit on the winners is part of the cost.
Profit: Return minus cost
Yield: Profit/Cost

3.1.2. Sample period/data

It has been necessary to take some decisions regarding the period from which the data is collected. The background for the decisions made to collect the sample should give the most reliable data and thus the most unbiased results. It seems nice to cover a whole year. When the data is collected over a year all seasons are included, and any impacts from seasonal fluctuations are taken into account.

Newest data is always more relevant, thus data on women's tennis matches played in 2010 was ideal. It also gives the opportunity to benchmark against 2011 once the year is over. It turned out that it was too optimistic to make such a widespread investigation. My resources did not allow me to cover a whole year of data. The data materiel was enormous, not only because of the number of games, but also because the data needed much manual processing before it could by used in an a data processing program (each match needed to be exported to excel and then sorted with copy paste). Basically it was necessary to reduce to amount of data. Without any good objective reason I chose to investigate the data from the latest quarter in 2010. Maybe the random selection is the strongest argument concerning the validity of the data gathering.

As already mentioned the data originates from Betfair and its associated trading platform. A program named Fracsoft provides the data[4]. You get access to the program by register and making an account. The data cost money (0,21 GPD pr. Match). It was another limitation to obtain the data.

The data materiel used in this study is derived from 157 matches. The program provides data on many different aspects of the trading. In this context not all is relevant. Of the data available the following data has been selected and used in the study.

- **Timestamp:** Data and time when the wager took placed.
- **Inplay delay:** If the trading happens before or after kick-off.
- **Last price matched**: The size of the odds on which there is made an exchange.
- **Total matched**: The accumulated volume matched.

3.1.3. Statistical tools

In order to conclude if the findings are representative and is not just coincidences, it is essential to test their statistical significance. The statistical test follows the principles for testing population means. The findings are tested according to the null hypothesis methodology. The null hypothesis is that the betting exchange market is 100% efficient and leaves no opportunity to systematically profit. The significance test contains five steps (Agresti, A. And Franklin, A. C., 2007)

[4] http://www.fracsoft.com/

Step 1: Assumptions regarding the data

The data is randomized. Normally one of the assumptions is that the population is normal distributed. This is probably not true, but on the basis of the central limit theorem and the relative large sample size it is not crucial too meet that assumption.

Step 2: Hypotheses

The hypotheses are the same in all research questions. It is the assumption that it should not be possible to formulate a strategy that offers a positive or negative return. If just one strategy provides a different return from zero it is possible to profit on the market (transaction cost ignored).

Step 3: Test statistic

The mathematical technique to perform a t-test used to determine the level of significance:

$$t = \frac{\text{Sample men} - \text{null hypothesis mean}}{standard\ error\ of\ sample\ mean}$$

Step 4: P-value

The P-value is based on a two-sided test and a t distribution with df = n -1.

Step 5: Conclusion

The hypotheses is rejected when the P-value is less than or equal to the selected significances level, which in this study is determined to be the standard level on 0,05.

4. Research question one

4.1. Are the odds that are being observed when the game is about to begin at their fundamental value?

This question is not directly related to behaviour finance. The data is not selected after any criteria that are outlined in behaviour finance. Therefore there is no reason theoretical why the odds should departure from its fundamental value. Instead the question demonstrates its relevance, as it illustrates the overall issue in a simple way. To what degree is the market effective? It makes us able to judge the effectiveness of the market under conditions that are easily identified. This makes an assessment intuitively valid. Besides, it is relatively easy to gather all the relevant data. The data points are directly identified in the question. There is no need for additional definitions. The result also provides insights in how meticulously one has to be when formulating relevant hypotheses.

4.2. Methodology

The first step is to identity the odds traded closest to kick-off. Normally there is heavy trading around kick-off. It has not been difficult to get data close to this point in time.

Out of the extracted data two parameters become relevant, inplay delay and last price matched. Before the game starts inplay delay has the value zero. When the game is launched the value is five.

There are two criteria that must be met. The data one the last price matched has to concern a transaction. It goes without saying, but the program also provides data when this is not true. The point in time is when for the first time five is observed in inplay delay. The next step is to divide the players between winners and losers. In the data handling the winning player was assigned with a code “one”, the losing player got the code zero.

Now all the wining players and their associated odds at the beginning of the game is isolated. The sum of the odds on the winning players represents the return.

In the data materiel there is 157 games played, our cost is equal 2*157 corresponding to placing a wagers on the wining and the losing player.

4.3. Findings

The cost and the return are easily calculated. Cost corresponds to the sum of all wagers. The return equals the sum of the odds on winners. The return from the 157 matches is 323,83 and the cost was as already mentioned 314. This represent what is called a point estimate and is our "best guess". Buying odds at kick-off actually provides a positive return on 9,82. That corresponds to a yield on 3,13% relative to the total invested money (cost). This means that in general people underestimate the winning players chances of winning, and they trade the odds at a too high value. It would be profitable to buy all odds on kick-off because you know the odds are too high on the winners. You do not care about the odds on the losers they are not part the return. A point estimation do not give any information on how reliable the results are, and this is why a more in depth analyse is required. We use the statistical tools to test the significance of the result. Table 4.1 summarizes the statistic analysis.

TABLE 4.1

RESULTS OF BUYING BOTH PLAYERS WHEN THE GAME IS ABOUT TO BEGIN

	Return	Cost	Profit	Yield
Total	323.82	314	9.82	3.13%
STATISTICAL TEST				
N		157		
Average		2.063		
Standard deviation		1.094		
Standard error of sample mean		0.087		
Null Hypotheses		2.0		
t-test		0.717		
P-value		47.5		

The significance test reveal a P-value on 47.5, this is the most interesting part of the analysis. It means that there is a 47.5% possibility to observe this result if the payoff on this strategy was indeed zero. 47.5 are very high. The result is very likely if the average return on the winners indeed was 2.00 and at their fundamental value. The hypothesis is rejected because the profit is not significantly different from zero.

The overall conclusion is that this strategy does not provide positive return in total. An average odds value on 2.063 is likely. Odds traded at the beginning of the match are trading at their fundamental value. It is an indicator of an effective market.

5. Research question two

5.1. Does trading volume influence market efficiency?

The trading platform in Betfair provides information about the available volume at a specific odd. This applies to both the offered odds and the demanded odds. When somebody places a wager the amount available declines because some of the available supply has been bought. The matched volume can be small or large, and it does not have to cover the whole amount available. The demand can cover more than the entire amount available for the odd. In such situations, the rest of the demanded odd will be covered when somebody wants to sell that odd again. Alternatively the rest of the betted amount can be placed to a lower odd.

E.g. to the odd 1,5 there is total of kr. 250 for sale. If you buy 50 then immediately after there is only kr. 200 for sale. If you want to buy for 300 kr. then you have to wait until somebody will offer the rest 50 kr. to 1,5. You also can choose to place the rest on a lower odd. That could be 1,4 where there is plenty volume available. In the first instance there is a match on kr. 200 and this applies as the trading volume. The wager on the reaming 50 would apply as a new transaction. In the odds program, where the data is extracted, a kr. 200 wager corresponds to a kr. 400 incline in total matched. This is because both the buyer and the seller agree on kr. 200. Our figures are based on the change in total matched.

This is the background for understanding trading volume. It all comes down to the following: the size of the amount when entering a wager to a specific odd and specific point in time. The matched volume in this thesis will be reported according to Fracsoft.

5.2. Methodology

The required date is identified by creating a connection between the size of the trading volume and the associated odds in that transaction. The odds are only included in the analysis if the trading volume meets a predetermined criteria. In practice there is a need to do this in a number of steps.

The fist step is to select all the events where the matched volume is above the given value. Second when the volume is above the specific value the associated odds is picked out. The data processing is carried out in Excel. Every player is assigned a new column in both steps. There is a column when it is confirmed that the trading volume meets the requirement. The value in the column is set to be blank if it does not. Then there is another column in a different sheet where the odds to the transaction are registered if the trading volume column is not blank. This process is repeated until all the players have been though. The data is divided in two areas, one for the players who won the game and one for the players who lost. It makes it possible to calculate the payoff.

The data materiel is huge. For each player there can be thousands of transactions in only one match. To avoid mistakes, a good strategy is therefore to check whether there are some games where there are equally many transactions. The number can vary from a few hundred to several thousand. It is unlikely that two players have the same number of transaction associated with them. If there are incidents where the number of transactions are identical then there are probably been made an error in the data handling.

When testing if the results are valid it is necessary to associate the null hypothesis with a specific value. Remember the null hypothesis is an efficient market. For that to be true, the average odd on the sample should be precisely equal to the value that provides a zero yield. The specific value is calculated by dividing the number of unsuccessful bets by the number of successful bets plus one.

A little illustration displays the calculation.

Suppose that there are observed 20 successful and 8 unsuccessful wagers in 28 odds.

Then the average odds should be:

$$8/20+1=1,4$$

There does not have to be equally many successful and unsuccessful wagers, maybe there is more trading on the winners.

Placing 1 on each of the 28 selected odds would have resulted in a profit if the odds were any higher, and a loss if the odds were lower. If an equally large bet on all possible odds generates zero yield, then the market is efficient.

5.3. Findings

First let us deal with some simple observations regarding the data material. Table 5.1 summarized the number of transaction at the different levels. The number of trades is split into winners and losers. It is easily verified that there is more buying activity on the winners at all levels. The distribution between successful and unsuccessful wagers indicates that the buyer activity is higher when the odds is lower than two. In order to maintain an efficient market there must be a larger proportion of bets that succeed when the average odds is below 2.0. It might be tempting to think the high demand for low odds is forcing them under their fundament value. On the other hand, the sellers undertake less risk by selling low odds compared to high odds, which also means that the low odds of being more attractive to sell. The supply and demand could be in equilibrium even when the odds is lower than two. Ultimately we cannot give a good reason why there should be equally many wagers on low and high odds.

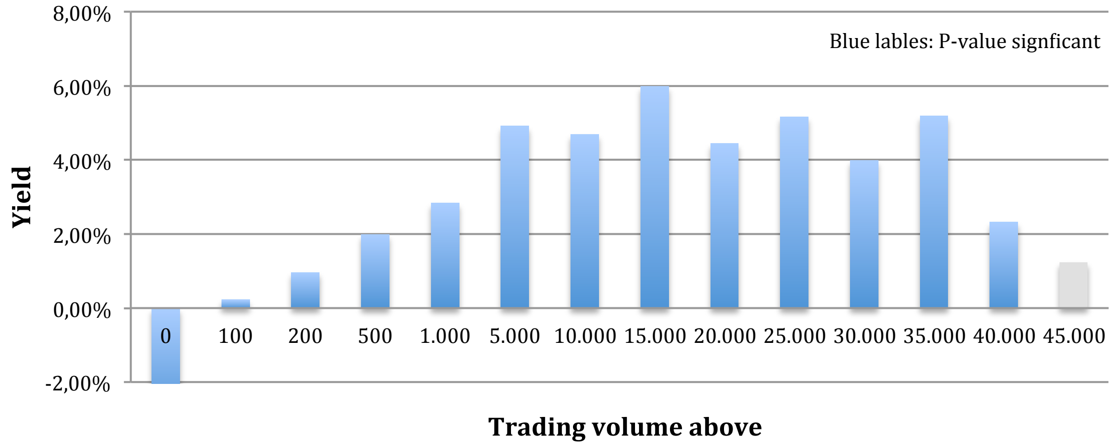
TABLE 5.1

TRANSACTIONS IN COUNTS

Trade volume above (kr.)	Winners	Losers	Total
0	297,036	187,010	484,046
100	146,601	57,227	203,828
200	110,007	38,879	148,886
500	68,947	20,975	89,922
1.000	41,791	11,078	52,869
5.000	7,905	1,380	9,285
10.000	2,959	426	3,385
15.000	1,564	180	1,744
20.000	881	108	989
25.000	594	64	658
30.000	406	42	448
35.000	313	27	340
40.000	219	22	241
45.000	163	18	181

The profitability of volume trading at each level can be found in table 5.1. It seems there is a connection between trading volume and profit. The yield increases as the low volume trades are sorted away. The peak is at the trading volume around 15.000, figure 5.1 illustrates this in a neat way. Thereafter the return declines but stays positive. It is surprising that the results are so unambiguous. A buying strategy only generates a negative return at the lowest level of trading volume (0-100).

Figure 5.1
Payoff when using tradingvolume as betting strategy



The statistical test does not acknowledge the size of the profit as a coincidence. The p-value is way above 0.05 at almost all levels. Table 5.3 summarizes the p-values. Based on these results it is not profitable to buy an odd if the trading volume is below 100. When the trading volume above it is profitable.

Transactions in the range between 15,000 and 20,000 provide the highest yield. The yield is increasing up to the point where the trading volume is above 15,000. When it begins to decrease it is implied that the yield at the higher volumes is lower, thus the highest yield must be between the trading volume 15,000 and 20,000.

TABLE 5.2
PAYOFF FROM VOLUME TRADING

Trade volume above	Return	Cost	Profit	Yield
0	469,542.69	484,046	-14,503.31	-3.00%
100	204,287.90	203,828	459.90	0.23%
200	150,325.25	148,886	1,439.25	0.97%
500	91,710.65	89,922	1,788.65	1.99%
1.000	54,369.78	52,869	1,500.78	2.84%
5.000	9,741.37	9,285	456.37	4.92%
10.000	3,543.78	3,385	158.78	4.69%
15.000	1,848.50	1,744	104.50	5.99%
20.000	1,032.96	989	43.96	4.44%
25.000	692.00	658	34.00	5.17%
30.000	465.86	448	17.86	3.99%
35.000	357.66	340	17.66	5.19%
40.000	246.60	241	5.60	2.32%
45.000	183.25	181	2.25	1.24%

From table 5.3 it is also noticed that the average odds decreases as the volume increases. The underlying mechanism to explain this correlation could be a matter of risk willingness. There are simply fewer who wants to wager large sums of money when the odds are high. When wager at low odds the seller will lose relatively less money and the buyer is more certain that he will win. The sellers Value at Risk decreases a low levels.

The results indicate that it is not a good idea to accept partly matched wagers. It is assumed that by accepting partly matched wagers many of the wagers are completed at low trading volumes.

There is a negative return associated with buying odds at low trading volume. Low volume trading should be avoided. The possibility to wager at low trading volume is higher when partly matched bets are accepted. Therefore it is best only to accept fully covered wagers.

Wagers in the interval between 15,000-20,000 provide the highest yield. In practiced it is not perfectly possible to be the one who wagers at in that interval. The high trading volume can only be observed after it has been traded. The best alternative would be to buy the odds immediately after a huge match.

TABLE 5.3

STATISTICAL TEST

Trade volume above	Average odds	Standard deviation	Standard Error	Null Hypotheses	T-test	P-value
0	1.581	0.877	1.609E-03	1.630	30.345	4.465E-202
100	1.393	0.433	1.130E-03	1.390	2.776	5.508E-03
200	1.367	0.380	1.146E-03	1.353	11.413	3.703E-30
500	1.330	0.322	1.227E-03	1.304	21.150	4.867E-99
1.000	1.301	0.288	1.406E-03	1.265	25.535	5.953E-143
5.000	1.232	0.227	2.553E-03	1.175	22.615	2.805E-110
10.000	1.198	0.214	3.931E-03	1.144	13.650	2.449E-41
15.000	1.182	0.210	5.301E-03	1.115	12.605	6.298E-35
20.000	1.172	0.210	7.083E-03	1.123	7.045	3.469E-12
25.000	1.165	0.205	8.409E-03	1.108	6.807	2.243E-11
30.000	1.147	0.183	9.083E-03	1.103	4.843	1.764E-06
35.000	1.143	0.179	1.014E-02	1.086	5.563	5,366E-08
40.000	1.126	0.160	1.079E-02	1.100	2.371	0.019
45.000	1.124	0.161	1.265E-02	1.110	1.092	0.276

6. Research question three

6.1. Are the odds at extreme situations either too low or too high?

The question implies a definition, what is regarded as extreme situations? In the literature review we talked about overconfidence and referred to a study. The study showed that something people where 100% sure about, happens only about 80% of the time. In the other extreme, what people thought would never happen did happen 20% of the time.

The methodology of this investigation is as far as possible in alignment with the 80/20 rule. Extreme situations are thus odds that imply a higher or a lower chance of winning than respectively 80% and 20%.

There are more spread on high odds compared to low odds, in other words, the difference between the values are greater at high odds. Thus it is not practical to conduct at 100% symmetrical examination. It does not make sense to use the same interval to examine the odds values. Furthermore low odds are characterised as being traded with more decimals that affects the classification of the intervals.

6.2. Methodology

The data necessary is the Last Matched Price. As we simply want to investigate if the odds at selected levels reflect the appropriate distribution between winners and losers.

The procedure is to test the distribution between winners and losers at different odds values. At each value the odds are split between winners and losers. The distribution should be similar to the one implied by the odds, if not it is a violation of the efficient market hypothesis. The LPM is placed in one sheet. In another sheet we set op the criteria. The odds only appear in this sheet if it is equal to the chosen value. Two areas spilt the values depending on whether it is an odd on a winner or a loser. Finally, it is easy to count the number of cases at each odd value.

The most difficult task is to determine the criteria and thus the odds values to be included in the analysis. There could emerge a critic of the objectivity in the selection of the odds values. Is the selection valid and does it give us a real perception of reality?

In fact there are not any values that are more right than others. They should all provide zero payoffs in an efficient market, but if the market is not efficient the selection could result in erroneous conclusions.

Giving the asymmetry of the odds the selection approach is a little ambiguous. Low odds are trades with two decimals and there is a relatively rich amount of data at those levels. It makes sense to include all values that lie between 1 and 1.2. A player traded at odds 1.2 is predicted to have a winning rate of 83%. This is close to the predefined level where only odds that have a possibility for success higher than 80% is regarded as an extreme. As for the high odds it is not as simply. It is not practical to select all values above 6.0. Odds above 6.0 correspond to a possibility of success lower than 17%. There would be too many different odds values to investigate. Besides most of values would not provide any trading activity because at higher levels the odds are mostly traded with one or no decimals. From experience I know that at higher levels transactions almost only appears at 0.5 intervals, e.g. 6.5 and 7.0 and so on. At very high levels the transactions appears on even more sporadic values. I have chosen to examine all odds on the basis of a 0,5 interval in the range between 6 and 20. Above 20 the selected values are partly arbitrary. The general approach is to choose values where there is most activity. This is necessary as the spread dilute the data, and the result could easily lose their significance.

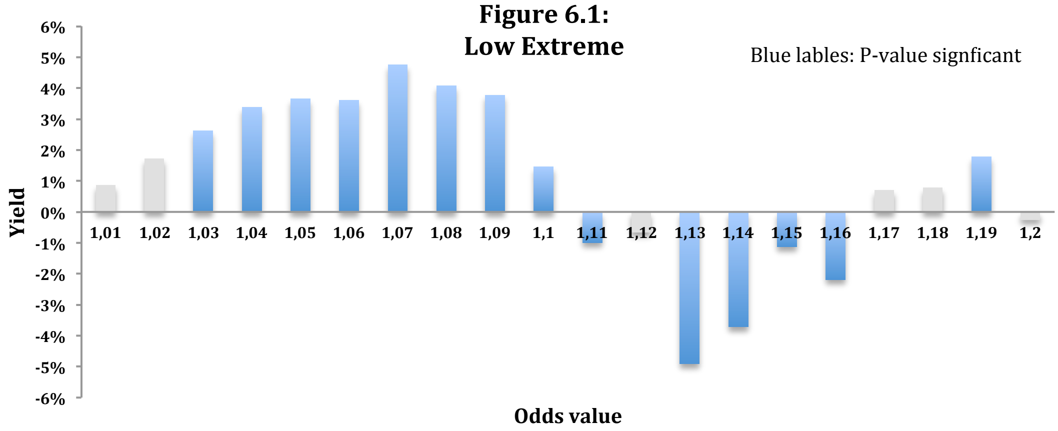
6.3. Findings

The yield at the different odds values is summarized in two charts. One displays the yield at the low extreme and one at the high extreme.

There are several requirements to be satisfied before the results can be characterised as significant. Most of them are already mentioned but in this situation it is necessary to include an additional criteria. In a binominal statistical test it is vital to have at lest 15 successes and failures (Agresti, A. and Franklin A. C. 2007). If that is not fulfilled the results are not significant no matter how many of observations there are included.

At the two odds values 1.01 and 1.02, the results are not significant, even when in total there are several thousand observations. The problem is that there are not enough failures. See appendix 3 for a detailed data reporting. The odds between 1.03 and 1.10 provide a positive return. Between 1.11 and 1.2 the results are more inconclusive. There are examples of positive yields negative yields and results not significantly different form a zero. In general low odds yields a positive return, and by this it is implied that the likelihood on success is estimated too low.

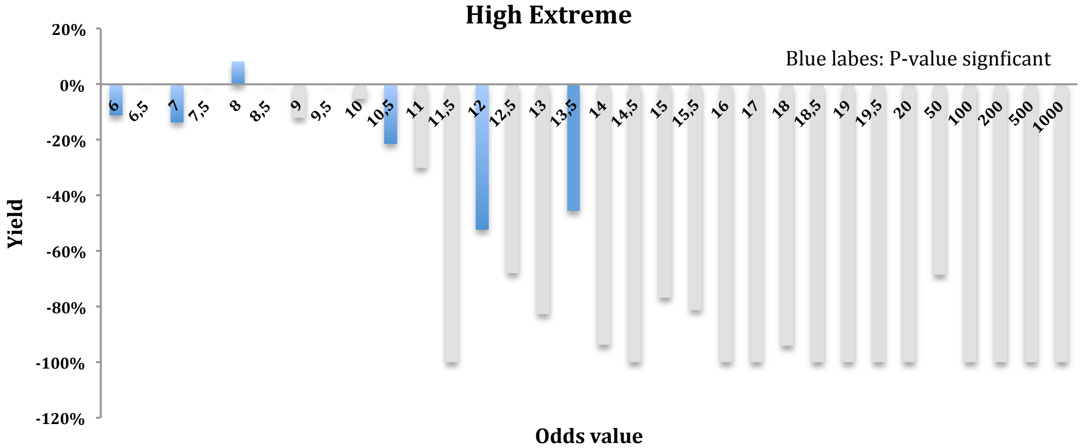
Figure 6.1:
Low Extreme



According to behaviour finance theory people should not believe that a player with high possibility of winning should be losing, as result they should be willing to buy the player with low odds even when the odds are in fact too low. In this perspective the yield on the extreme low odds should be negative, because in theory, people should overestimate the chance of wining. The results are different from what is expected from the theory. Actually they show quite the opposite. Low odds are set too high and it is a good strategy to buy odds at the low extreme.

There are relatively few significant results at the high extreme. See figure 6.2. The yield is mostly negative. There is only one incident where the yield is a positive. The reason why minus 100% yields are observed is because there are no winners on those odds. In many incidents the data materiel is weak, and it is not possible to reach any conclusions. The problem is that the spread dilutes the data and there is simply not enough success satisfying the requirements of 15 success and 15 failures.

Figure 6.2
High Extreme



There is another way to interpret the data. Assume that all wagers between 10,0 and 15,0 have been made at odds 15. In this way the data is manipulated in such a way that the requirements of 15 success and 15 failures is fulfilled. The assumption also leads to the highest possible number of failures. Thus the assumption favours the likelihood of loosing.

The total number of winners and losers between 10.0 and 15.0 are respectively 305 and 5,822. This corresponds to a share of 4.89% winners. Assuming all wagers where made at odd 15.0 there should be 6,67% that won. The winning rate is significantly too low[5], even when it is underestimated. The point being, in the high end of the scale there is too few winners. Thus the odds are too low. This is not a surprise. If the odds at the low end of the scale are too high, then it is natural to expect that the odds are too low at the high end of the scale.

According to theory the odds should be too high in the high extreme. People should underestimate the chances that the player will win, when the possibility of wining is less than 20%. Thus the wagers should happen at higher values, actually so high that we expected a profit from buying odds at the high extreme.

Both in the low and high extreme the result contradicts our hypotheses. This could be a indicating of a superseding effect. An effect that causes the odds at the low extreme to be traded at too high values and conversely at the high extreme. We will disuses the possibility and shape of such an effect in the overall discussion.

The investigation of odds at the extremes reports a mixed clutter of significant results. In the lower end of the scale we find many significant results but in the high end only six out of 32 provides us with significant results.

There should be a certain degree of scepticism about the results in the investigation. When testing many cases, as in this case the results could be a consequence of data mining. Therefore we are wary of being too heavy-handed in the conclusions.

[5] P-value equal $1,32*10^{-9}$

7. Research question four

7.1. Does a period of stabilization in the odds establish an anchoring and as result cause the odds to fluctuate too little?

There should be no doubt that the most significant factor in estimating the size of an odd is the performance of one player relatively to the other. That said, behavioural finance theory has lead us to understand that people sometimes form their expectation or opinion based on earlier observation. This effect is called anchoring. The reason why it becomes relevant is if people use the observable odds in their own estimation.

If the odds has stabilized around a specific value there could be formed a belief of what is a fair odds. That belief could work as an anchor and cause the traded odds to fluctuate less than it should.

There is a challenge in defining the characteristics of "stable odds". It is not difficult to outline the factors, but there is no valid methodology to decide the value of the parameters. The two factors that are intuitively right are time and the change in the odds over that time. In a situation where you can question the researchers objectivity, the strength of the investigation is dependent on the validity of the subjective decision. I have found it appropriate to look at intervals on respectively 3 and 5 minutes. Two different periods is better than one. If the conclusion is the same giving two different procedures then we are more inclined to believe in the results. Regarding the change in the odds, it should be as low as possible but still gives enough data to make valid statistical assessments.

I have chosen a max 5% change. If the effect of anchoring occurs at even lower levels there is a risk of including to many data points. That could dilute the data material. Hopefully the value is low enough, and even if the data materiel is diluted to some degree, it should still be possible to make valid conclusions.

Odds trading also take place before kick-off. Right now only the odds that are traded doing the match are included in the analysis. There is a separate research question that deals with that period before kick-off.

Besides as it will be reviled later, the odds before the match are very stable. There is not much new information to alter the odds.

This is quite different compared to after kick-off. After kick-off the flow of new information is massive, and the odds fluctuate tremendously. The aim is to determine if anchoring causes people to react inappropriately to new information.

7.2. Methodology

The data concerning Inplay delay, Timestamp and LPM are necessary to conduct this investigation. The Inplay delay takes two values depending on the state in which the odds are being traded. The two states are respectively before or after kick-off. If it is before kick-off, Inplay delay takes the value five otherwise zero. The date is sorted so only relevant odds are included. That is when the Inplay delay is five and the transaction has happened during the match.

The first step is to identify the odds at every three and five minute. It is practical to separate the data into two different sheets in Excel. The three and five minuet interval is calculated on the basis of the difference in Timestamp. Ideally the odds should be selected at every 3 and 5 minute. In practice there is some reasons why this cannot be done perfectly. There is no rule saying that there should be a transaction precisely ever 3 or 5 minuet. It is only possible to get as close to the interval as the data material allows. It turns out, that this is not a big problem. The trading activity is so heavy that there is a transaction at almost every second.

If there is an incident where there is a little break in the trading the methodology result in a divergence. The problem is not believed to have a significant impact on the findings.

There are some technical hurdles in the process of selecting odds that satisfy the time interval criteria. Our resources are limited to Excel. The way that Excel works leaves us with some problems in the process of identifying the odds every 3 and 5 minute. The best way to use Excel is as follows.

The playing time is calculated by summering delta timestamp. It happens in chronological order. Each time the playing time ad up to 3 or 5 minutes the total playing time is subtracted with respectively 3 or 5 minutes. This “resets” the total time played. The odd associated with the transaction that triggers the subtracting is used in the next step of the analysis.

As mention before, a transaction does not necessary take place precisely at every 3 or 5 minute. The exact number is subtracted each time, and thus the next odd is influence by the foregoing approximation. The problem is not believed to have a huge impact. There are many transactions, and the bias is corrected repeatedly.

For example there could be a gap between two odds such that the time interval are 3 minutes and 45 seconds instead of 3 minuets. The next odds will as result be selected approximately 2 minutes and 15 seconds after the previously.

That is the overall problem, but afterwards and thus the time between the third and forth odds should be approximately 3 minutes again. For each player the odds traded at 3 and 5 minutes are identified as foretold. This is not the only criterion that has to be fulfilled.

The next step is to sort the odds depending on their degree of fluctuation. In technical terms; the odd is only included if there has been a decline in the odd less then 5% compared to the last odd traded odds 3 or 5 minutes ago. These sorting leave us with only the odds that have fluctuated between -5% and 0%. The sigh is negative because only a decline in the odds is relevant. We want to examine if anchoring appears at the time where the odds is *not* raising and not falling too much. If also an increase were included then we could not be sure of observing anchoring. Anchoring can happen in both direction and if it does, the effects could offset each other. Maybe the odds do not raise and fall enough. It is only possible to separate the effects if the selected odds are based on a change in only one direction.

The point is that we believe in too little fluctuation on the down side (that is when the odds declines) and that causes the future odds to be too low. We want to test if this phenomenon takes place.

Because of anchoring maybe an odd on 1.07 is observed while the fundamental value is 1.10, thus anchoring causes buyers to buy too low odds. As a final criterion we also test whether it has any impact if the “stable” odds is below 1.2. The reason to include this extra criterion is because of the behaviour called overconfidence. In the low end of the scale anchoring and overconfidence could reinforce each other. If people are overconfident that the favourite end up wining, they are willing to buy the odds even if it is too low. That would be the similar to the result we would observe with anchoring.

7.3. Findings

First lets us engage in the analysis of the 3-minute interval. The results are showed in table 7.1. As always the findings are followed up by a statistical significance test. The results of the statistical test are to be found in table 7.2.

When all odds are included there is a slightly negative yield. This is not unlikely. The P-value is 0.44 far above 0.05, which is the highest accepted value. Thus the result is not significant and the null hypothesis is not rejected.

TABLE 7.1

PAYOFF AND (COUNTS)

3-MIN INTERVAL. CHANGE IN ODDS BETWEEN -5% AND 0%.

	Return	Cost	Profit	Yield
All	2,618.4 (1,794)	2,651	-32.6	-1.23%
Odds below 1,2	1,068.93 (1,009)	1,076	-7.1	-0.66%

When looking only at the odds that are below 1.2 there are weak signs of anchoring, or at least it could be the result of anchoring. The yield is with statistical certainty different from the fundamental value and negative just as we predicted it to be. Knowing this about the odds market makes us able to generate a profit. When also introducing risk and time it is probably not worthwhile the effort.

If we think of the return as a measurement of the market infectivity then the relatively small value should be perceived as a sign of only small violations of the a effective market hypotheses.

TABLE 7.2

STATISTICAL TEST

3-MIN INTERVAL. CHANGE IN ODDS BETWEEN -5% AND 0%.

	Average odds	Std. Div.	Std. error	Null Hypotheses	T-test	P-value
All	1.46	1.00	2.37E-02	1.48	-0.77	0.44
Odds below 1,2	1.06	0.05	1.55E-03	1.07	-4.51	7.15E-06

Table 7.3 reports the payoff from the 5-minute interval strategy. The conclusion is very much the same as with the 3-minute interval. When all odds are included there is a small positive yield but again, the P-value is so high that it is not unlikely from a random sample. The yield is also negative when the odds are below 1.2.

TABLE 7.3

PAYOFF AND (COUNTS)

5-MIN INTERVAL. CHANGE IN ODDS BETWEEN -5% AND 0%.

	Return	Cost	Profit	Yield
All	1.251.3	1,248	3.3	0.26%
Odds below 1,2	554.5	557	-2.5	-0.45%

The negative yield when only odds below 1.2 are included is a bit higher in the five-minute interval. However the results are less significant compared to the three-minute interval situation. The p-value indicates that anchoring is more likely to happen at a 3-minute interval. It is perhaps not so surprising that we observe a higher p-value at the 3-minute interval taken into account that there are more statistics to support the results for the 3-minute interval.

TABLE 7.4

STATISTICAL TEST

5-MIN INTERVAL. CHANGE IN ODDS BETWEEN -5% AND 0%.

	Average odds	Std. Div.	Std. error	Null Hypotheses	T-test	P-value
All	1.43	1.03	3.47E-02	1.42	0.11	0.91
Odds below 1,2	1.06	0.05	2.15E-03	1.06	-2.20	2.86E-02

The overall conclusion is that it is not a good strategy to buy odds if it has been low for a while. You could argue to sell instead, but the yield is insignificant and it dose not really make sense to pursuit that strategy. There are small signs of anchoring at both intervals when the odds are below odds 1.2.

8. Research question five

8.1. Are there any trading patterns in the period before kick-off?

This question is formulated as result of user experience on Betfair. Users claim to have noticed that the odds decline in the period from where the market opens and until the match begins. I have not been able to get special insight into the conditions under which such trend might appear. I think it is worth investigating if there is a period before kick-off where the odds approximate fundamental value. Therefore I have chosen three different scenarios where it may be possible to identify a trend.

The data are dividend into different groups. 1) A group for the winners. 2) A group for the losers. 3) One where all data is included.
In the situations where winners and losers are handled separately the results cannot be directly used to formulate a trading strategy. Post game analysis on a particular group cannot be used in the pre game decision-making. It would mean that we know who is going to win. In real life this is obviously unknown.

Assume that people uses other people's assessment in the process of estimating odds. Gradually information available on the market would be integrated in the odds. That could cause the number of transactions to have an influence on the odds. If so it would push the odds closer to its fundamental value. Giving the odds was not at its fundamental value in the first place.
If the above reasoning to some degree occurs there could be a connection between trading and the preciseness of the odds. The point is that the odds might change a little in the period after the market opens.
A correlation between the odds value and the number of transactions is the basis for a trend. In a coordinate system the odds would apply as y-values and the number of transaction as x-values. The correlation could either be positive or negative. If the odds is too high to begin with then the odds should decrease, but it could just as easily be too low and then instead increase. The objective is to test if there is a dominating trend. The null hypothesis is that such does not exist.

8.2. Methodology

The necessary data is the odds development from all the players in the time before kick-off. The data program assigns a code to the transaction. It is zero when the transaction is before kick-off. After kick-off it is five. In Excel all the odds, which is assigned with the right code is used in the further analysis. The odds values correspond to the y-value in a linear regression. The x values are found by calculating which number the odds had in the transaction order. A wager could e.g. be transaction number 60 and that would be the x value. This is the basis of the regression (y=ax+b). The a-parameter and thus the slope is the target. The slope determines whether there is a trend in any direction for the given player. The aim is to confirm or reject overall trends regarding the three different scenarios, which was set earlier. The most important value is the slope. It is tested statically if the slope is different from zero.

The numbers of transactions very considerably, and when there are many transactions it can easily dilute the data and cause the slope to approach zero.

To address this problem we also use the slope and the number of transaction for each player to identify the total chance in the odds before kick-off. This provides another basis for understanding the odds development. Using the method where the slope is multiplied with the number of transactions we find the total change between the first and the last traded odds. The total difference is not sensitive to a dilution in the slope.

The reason to use the slope to calculate the difference between the first and the last traded odds is that the odds have a tendency to fluctuate a bit. We want the trend and not the fluctuations because of the difference in the odds. By using the slope these fluctuations do not have an impact on the results. Thus it is a safer way to determine the total change in the odds.

8.3. Findings

To begin with let us examine the slope in the different scenarios. Table 8.1 shows the average slope in the three scenarios.

The winners have an average slope close to zero, and a high p-value. The average slope of the losers is higher, but still not so high that is we expect it to be different from zero.

A small slope implies that the change in the odds is small. That is when looking at two transactions in chronological order. It is not surprising that the change between two transactions is insignificantly. After all a big change would mean that the first an last traded odds would differ a lot. That would imply a violation on the market efficiency. The information flow in this time period is limited and the odds should no differ very much.

TABLE 8.1

TREND BEFORE KICK-OFF

	Winners	Losers	All
Average Slope	-4.88E-05	7.07E-03	3.51E-03
Std. Div.	0.0030	0.0470	0.0335
Std. Error	2.380E-04	3.753E-03	1.888E-03
Null Hypotheses	0	0	0
T-test	-0.205	1.884	1.859
P-value	0.838	0.061	0.065

Table 8.2 displays the average number of transactions. There are a greater number of transactions on the winners compared to the losers. That could be one reason why the slope is closer to zero on the winners. If the participants are fast in finding equilibrium a trend is actually only observable in a short time. When the odds have reach its fundamental value all the relevant information is implanted in the valuation.
Afterwards the odds should be stable. Subsequent transactions cause the slope to approach zero, thus the high number of transactions dilutes the data.
The p-value for the slope is above 0,05 in all three cases. It could be a sign of a quick stabilization, or that the odds are at their fundamental value from the beginning. The high p-values imply that the market reacts, as an efficient market should. When our data do not provide us with significant results, there are not any reasons to discuss them further.
The results are more interesting when looking at the total change in the odds. Table 8.2 shows the results for this part of the analysis.
The numerical difference between the last and first traded odds before kick-off is bigger for the losers compared to the winners. The p-value for the winners is above 0.05, but for the losers the p-value is below. The null hypothesis is rejected for the losers. We believe there is a difference between the traded odds in the period before the kick-off.
Actually knowing that losers tend to have a positive slope can be used to take better decisions when trading. The fact that the losers tend to have positive slope can be used as an indicator to predict who is going to lose. Thus a trading strategy would be to sell a player with a positive slop, or buy the player other player.

TABLE 8.2

DIFFERENCE FROM FIRST TO LAST TRANSACTION BASED ON THE SLOPE

	Winners	Losers	All
Average number of transactions	202.6	155.8	179.2
Difference between the first and last odds	-0.027	0.238	0.105
Std. Div.	0.2403	1.3647	0.9872
Std. Error	1.918E-02	1.089E-01	5.571E-02
Null Hypotheses	0	0	0
T-test	-1.416	2.183	1.890
P-value	0.159	0.030	0.061

Table 8.2 provides the evidence for this reasoning. Take all the positive slopes from the 314 players and divide them between winners and losers. If the slope does not indicate

who is going to win, half of the players with positive slopes should lose. Knowing the value of the slope should be irrelevant and it should also be zero. This is not what the data indicate. The results are showing that the players with a positive slope are likely to lose.

TABLE 8.3
INVESTIGATION OF THE POSITIVE SLOPE

Winners	Losers	Total
59	90	149

Statistical test	
Point estimate	0.60
Standard error	4.007E-02
Null Hypotheses	0.5
Z-test	2.60
P-value	0.010

It does to some degree make sense that the slope is positive for the losers. If the number of transactions affects the accuracy of the odds, it is logical that the odds on players who actually end up losing are increasing.

There are not found evidence of significant results concerning the overall trend. The null hypotheses are not rejected. In total the odds does not increase or decrease in the time period before kick-off. On average the odds are going to be traded at the same value in the period before kick-off.

People who claim that the odds tend to differ are influenced by wishful thinking. It is true that there is a positive trend in the odds development on the losers, but it is not known who is going to lose at that time. It is not possible to differentiate the players between winners and losers before the game is over. A timing strategy is not profitable, that is to buy when the market opens or at kick-off, simply because the total change is not significantly different from zero.

The trading strategy has to work the other way around. Our findings can by used to give some idea of who is going to win the game. If it is observed that the odds on a player have increased in the time period up to kick-off, it is believed she will lose.

9. Research question six

9.1. Is it possible to generate a profit from the strategy of buying winners?

The inspiration to test this strategy is derived from the momentum effect on the stock market. High performance stocks have proven to yield an abnormal return in a subsequently period. This strategy requires some adjustment in order to use it on the odds market. First, what is a winner?

One way to interpret a winner is simply to define a winner as the player who has the greatest likelihood of winning. In other words it is the present favourite. Translated into odds, the favourite is the player with an odd traded below 2.0.

That definition implies a closely related methodology to one used in a previously research question. When we investigate extreme odds the main parameter was also odds categorized on a scale. The main difference is that this question includes a broader spectre.

From the earlier investigation we discovered that the odds seem to be too high in the low end of the scale. That tendency will give an advantage to a strategy where you buy all odds below 2.0. We are prone to believe that it would be a good strategy to buy the winners. The impact of momentum strategy on stock market is well documented. I think it is important to test the momentum effect in a full range perspective. The momentum effect could exist in all high performance situations. This way of thinking leads us to another way of defining a "winner".

Unlike a favourite, a high performance player could be one where the odds have depreciated. When the odds decline on a player that player must has prospered. A lower odd implies a greater likelihood of winning compared to before the decline in the odd. This also corresponds to a better performance. Thus alternatively to the first way of defining a winner, a winner could be the player who has performed best in resent time.

The first approach is a fundamental way of perceiving a winner the second is relatively. You could argue that both ways meet the requirements necessary to observe a momentum effect. Hence we test them both. The stock market has no time constrain, we think of the companies as going concerns. In a tennis match we know that we are closer to an ending each time a point has been played. The momentum effect on the stock market is not permanent, so what about the odds market?

Fortunately when you buy an odd it does not chance when the odd of the of the player change, you still have it. The payoff is fixed no matter what happens afterwards. Time limitations on the odds market can be ignored. If the odds where too high when it was bought, it is too high permanently.

9.2. Methodology

The Last Price Matched is the key data. The odds only have to satisfy one criterion, it has to be below 2.0. There is no reason to distinguish between wagers carried out before or after kick off. Odds traded before kick-off is as likely to be influenced by the momentum effect.

We select the odds on the basis of its value. Odds below 2.0 are selected and copied to another sheet. The result is a sheet in Excel where only the traded odds below 2.0 are represented. From there the odds is divided into winners and losers. As always it is tested if the trading strategy, where you place an amount on all odds, provides a positive or negative return.

The research question was two sided, and there is a different methodology attached to the second part. The odds are selected on the bases of the change from the previously traded odds. Odds are only relevant if they have declined. Hence for each player we follow the development of the odds, if there is a decline in the value, it is picked out and put in another sheet. If the odds is the same or have inclined nothing happens. In the end all odds traded at lower values than the previously traded odds are divided between winners and losers.

9.3. Findings

A winner in this context is not related to who wins the game. The term is used to describe odds which meets some clearly defined criteria. It is inherited from the way a winner stock is perceived in a momentum effect perspective. Since that is the background I find it worthwhile to use the same terminology. The data that meet the first definition on winner odds is seen in table 9.1. The first observation is that there are a higher number of transactions on the favourites, almost 3 times as many compared to the decline in the odds. It is consistent to the average odds, which is also in the low end of the scale. There should be more successful transactions on low odds.

The average odds and the Null hypotheses are close to each other. The difference is on third decimal. The great amount of data reduces the Std. error to such a level that our finding is significant. The p-value is very low and almost zero. The selected odds, under the given criteria, are different from what we should observe in an efficient market.

The null hypothesis is rejected, but the yield we can expect from this trading strategy is not very high, our data reports a 0.49% yield. The momentum effect is not at level where we give it much importance. In chaotic markets it is not surprising to find small violations on the effective market hypothesis.

TABLE 9.1
MOMENTUM EFFECT WHEN A "WINNER" IS DEFINED AS ODDS BELOW 2,0
PAYOFF AND (COUNTS)

Return	Cost	Profit	Yield
345,045.2 (256,000)	343,348	1,697.2	0.49%

Statistical test	
Average odds	1.348
Std. Div.	0.256
Std. Error	5.05E-04
Null Hypotheses	1.34
T-test	13.13
P-value	2.37601E-39

Lets look at the results when a winner is perceived as an odd that is traded at a lower value than last it was traded.

The average odd is higher in this situation. This methodology includes all odds no matter how high there are, unlike above where only low odds are included. The yield from this selection method is negative and almost -2.5%.

There are fewer data points compared to above. The logic explanation is that many of the transactions are traded to the same value. Hence there must be fewer points half is sorted away and out of the rest not all are included.

The p-value is extreme low, and there is no doubt the results is significant. Obviously there is a big difference in how to perceive a "winner". According to the momentum effect a "winner" is the indicator of a good stock. Our conclusion is that there is no momentum effect when a winner odd is defined as a decline in the odd, actually quite the opposite.

On the other hand there is observed small signs of momentum effect when a winner odd is one below 2.0. This is a dilemma. Is it more correct to use the first results or the second results? Both interpretations are ways of perceiving a winner odd.

Given the discussion, it seems more right to give most credit to the first interpretation. It is the one that provide a positive return, and therefore it is the trading strategy one should follow.

TABLE 9.2
MOMENTUM EFFECT WHEN A WINNERS IS SEEN AS AN IMPROVEMENT IN THE UNDERLYING CHANCE OF WINNING
PAYOFF AND (COUNTS)

Return	Cost	Profit	Yield
120,965 (75,640)	124,047	-3082.03	-2.48%

Statistical test	
Average odds	1.599
Std. Div.	0.256
Std. Error	9.30E-04
Null Hypotheses	1.64
T-test	-43.81
P-value	0.00E+00

The results from table 9.2 leave us with an interesting question. If you get a negative return when buying odds that had decreased in value, does that leave us with a positive return if we buy the odds that had increased? To explore this question we use the same methodology as before, except we select odds that have increased in value from the previously traded odds.

This part of the research question is not connected to the momentum effect. It is not a way to perceive winner odds in that sense. It is a logic consequence from the, to some extent, large negative result we saw in table 9.2. The results of the new investigation are showed in table 9.3.

The conclusion is similar to the one we got when we look at the odds that had decreased. The number of data points and p-value are in the same range. The yield is a bit lower, thus this trading strategy is even worse. Our overall conclusion is that there is no positive effect from buying odds on the background of the previously traded odds.

TABLE 9.3
RETURN FORM BUYING ODDS THAT HAS INCREASED IN SIZE COMPARED TO THE PREVIOUSLY TRADED ONE
PAYOFF AND (COUNTS)

Return	Cost	Profit	Yield
122,528.64 (73,235)	125,816	-3,287.36	-2.61%

Statistical test	
Average odds	1.673
Std. Div.	0.256
Std. Error	9.45E-04
Null Hypotheses	1.72
T-test	-47.49
P-value	0.00E+00

It is a bit surprising that buying based on a change generate negative yield. This means that it is in general bad to be the buyer. If there is a negative return for buyers then there must be a positive return for sellers.

That does not really makes sense, but maybe when looking at the overall perspective there is a logic explanation.

In an earlier study (trading on the basis of volume) we saw that the total return if you bought all odds where negative. Trading without a strategy the data does provide the evidence that it is best to be a seller. That is the reason the results are as they are. To buy odds based on the change from the previously traded is just a random strategy.

10. Discussion

A thorough analysis of the odds market reveals small signs of skewed human decision-making. We have documented several violations of the efficient market hypotheses. It is dubious if the reasons only should be found in the advanced version of financial behaviour. The majority of the themes mentioned in the literature review emphasize the advanced version. To some degree the violations could simply be a matter of risk avoidance. We have also seen imbalances that are not directly explained by financial behaviour.

Our aim is to formulate good trading strategies in order to profit. To succeed it is necessary to identify the characteristics of central factors. Central factors reveal profitable opportunities in the market. They also contribute to the understanding of human behaviour. Knowing which forces that affect the price structure of the odds market could be useful in other markets.

It is time to sum up the findings. There are several good trading strategies and observations that are insightful for the active trader. There are plenty of violations and they can be use to personal benefit. The major findings are summarized in table 10.1.

TABLE 10.1
PROFITABLE TRADING STRATEGIES AND OBSERVATIONS

	Yield
The odds at kick-off is at its fundamental value	0.00%
Buy when matched volume is around 15.000	5.99%
Sell it the trading volume is below 100	Minimum 3.0%
Buy low odds (1-1,1) highest yield	5.00%
Sell when the odds is high (above 10,5)	61.07%
A stabile odds indicate a fair price	0.00%
There is no trend before kick-off, it does not matter if you buy or sell as soon as the market opens	0.00%
A positive odds development in the time before kick-off is more likely to happen on the loser of the game compared to the winner of the game.	n/a
Buy odds under 2,0	0.49%
If the odds change in value sell, it dose not matter in which direction. This finding is likely the result of the low volume strategy. Remember that over 50% of the transactions happens when the trading volume is below 100, and where the yield from buying is negative. Hence in general it is best to be a seller.	2.50%

Two main factors differentiate different trading strategies, trading volume and the odds value. When the results are separated into these two categories we find that many of the strategies overlap and are natural consequences of each other. Lets take an example.

If winner odds are defined as odds below 2.0 then you should buy the winners, it provides a yield of 0.49%. But the yield from buying the odds between 1.0 and 1.1 is even higher. The odds that are between 1.0 and 1.1 are of course also included in the odds below 2.0. Thus they are part of the reason why all odds below 2.0 provides a positive return. That leaves one to think that the positive return from odds below 2.0 is caused by the positive return on the very low odds.

A way to refine the “buy below 2.0” strategy is to stick to the low odds because they provide a bigger yield. Besides they could be the reason the yield on odds below 2.0 is positive in the first place.

Our investigation shows that odds higher than 10.5 provides a negative return, it is best to sell when the odds are high.

Following a trading strategy of selling odds at high values is the most profitable trading strategy. The yield is at a level where it should be questioning if something is neglected. Maybe our data is not significant but in the retrospect we have still not be able to find errors in the procedure.

The return from odds at the extremes differs from their fundamental value. The research questions that revealed those biases were inspired by overconfidence and anchoring.

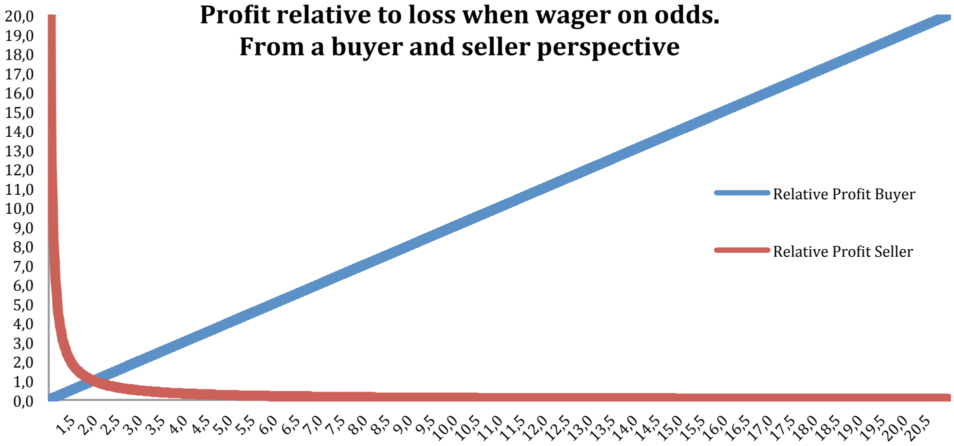
In this stage where it is possible to get a broader look at the results, there is an explanation of the violations that appears to be more straightforward. Risk aversion can explain why the odds differ from their fundamental value under extreme conditions. Think about what happens when two people wager at odds 1.01. With that wager you can get 0.01 in profit and lose 1.0. This is a highly skewed distribution. There is a high chance of a small profit and a low chance of a high loss. A risk averse person would not make such a wager without any expected profit.

A risk averse person would demand a risk premium to wager when the distribution is skewed, and that is exactly what we see.

Take a look at figure 10.1 from there it is easily to see the premium relative to the loss at different odds values. Seen from a buyer’s perspective the potential premium is relatively low compared to the loss when odds are low. In the same situation, the seller has the potential for a high premium relatively to the potential loss. The opposite is true when the odds are high.

These observations regarding the premium on low and high odds are consistent with what we expect from risk averse people. From figure 10.1 it also becomes clear that odds around 2.0 should not differ from their fundamental value because of risk aversion. The risk associated from a wager in this area is similar for both buyer and seller.

Figure 10.1
Profit relative to loss when wager on odds.
From a buyer and seller perspective



The other category of indicators is the size of the trading volume. There are several findings that show a relationship between trading volume and the degree of deviation form the fundamental value. When the trading volume is high there is a high demand and supply of the odds, and it would be natural to think that it should press the odds closer to its fundamental value. This is not the case.

At high trading volumes, it is good to buy and if the trading volume is low it is best to sell. The yield for a buyer peaks when the trading volume is around 15.000. That is a bit random, there is no good explanation why the odds value should differ from the fundamental value at this level.

Behaviour finance theory does not provide any good explanations for this observation, and neither does observations on the stock market. On the stock market it is usual best to be a seller when the trading activity is high (Tvede, L. 1999). We do not like a result that we can't explain and which appears to be random. Maybe the result is caused by a bias in the data.

More interesting is it to see that odds traded at low volumes provide a negative return. Previously we concluded that the reason could be a result of how the trading system works.

The trading platform gives you different options. You can agree not to get the entire wager matched at once. By accepting that, your wages are at risk of being matched gradually. This setting is default. You have to actively choose that you will only involve in the wager if it fully matched. My theory is that the acceptance of a partly matched wager causes many transactions at low volume.

The "delay" causes the odds to be "not up to data". The remaining part of the deposit is bought at odds that are generally too low. Odds are too low at low trading volume because the odds are surplus offers from wagers that were partly matched.

You might think that it would also work the other way around. Partly matched wagers could sell at odds that are too high. Following this logic the total effect of partly mated wagers should be neutral. But who would sell odds that are too high? Besides somebody else will provide a better deal and the partly matched odds will be forgotten. The conclusion is that it is best to avoid making partly matched wagers, because you end up buying odds that are too low.

The skewedness in the odds, at low volume, is from this point of view caused by the design of the trading system and how it is used. Not because of some underlying psychological dysfunction. If that is true, there is not any causality between the negative return from buying odds at low trading volume and the trading activity.

Buying based on a change in the odds generates a negative return. That observation is most likely linked together with another finding. Low volume trading generates a negative return. Over 50% of the transactions happened at a trading volume below 100. Using change in the odds as trading strategy is likely to coherence with low volume trading. It does not seem appropriated to credit another reason for this bias.

The negative yield from buying based on change in the odds is not considered as a very important finding. The trading strategy is considered random. We already know that it is best to be a seller when trading without a system.

10.1. Are 157 matches enough data to provide us with valid results?

This is a data technical issue. We know some odds have a very low possibility for success attached to them. The question is if 157 matches are enough to be represent the extreme outcomes?

With 157 matches there are odds related to 314 players, on those players there is a total of approximately 480,000 transactions. There is no doubt that 480,000 are more than enough data to make strong significant results. The concern is that many of the players have many transactions associated with them and thus data points attached to them. Depending on the player's success all the data point is either related to a winner or a loser and all the odds becomes success or failures. There are 157 matches that correspond to 157 winners and losers. Because the fate of one player has high influence on the extreme odds, a few outliners could interfere with all the data on extreme odds.

Basically there could be a problem with the amount of data material. Maybe 157 matched are not enough to secure robustness concerning odds the extreme. Although it seems like it when in total there are above 480,000 data points. The result show that there is positive yield from buying low odds, and that there is too few winners when the odds are high. It is not perceived to be a problem regarding the size of the yield on the extreme low odds. The problem arises when we look at the size of the yield when selling extreme high odds. Is it true that the market is so inefficient that it is possible to gain a 60% yield on such a strategy?

Since the results correspond to the assumption the people in general is risk averes, the findings are believed to be valid. It is the size of the yield especially on high extreme that is dubious.

10.2. Transaction cost

Until now transaction cost have been ignored, and the slightest deviation from the fundamental value could be use to generate a profit.

The exact calculation of the commission is a bit complicated. Overall you pay 5% of the net return from a wager. Heavily trading reduce the commission rate, see appendix 4 for the different discount rates.

The lowest possible commission rate is 3.0%. You only pay transaction cost based on the premium.

E.g. if you bet 100 at odds 3.0, you end up with 100 and the premium minus 3.0% (100+200*97.0=294)[6].

6 http://sports.betfair.com/da/aboutUs/Betfair.Charges/

The yield in this situation is 194/100=194%. Imagine the yield which generate from ten similar transactions thus 194/1000=0.194%. Without transaction cost: 200/100=200% and 200/100=0.20%.

The yield cannot be compared with and without transaction cost. It is not possible to simply substrate the commission rate with the yield (0,2-0,03≠0.194). It is mathematical not on the same scale. The percentage transactions cost have to be integrated in the calculations of the yield. That also means that before calculating the premium it is important to know if is from a buying or selling strategy.

I have selected the yield that differ most from zero and integrated a 3% transaction cost. The results are fund in table 10.2.

Table 10.2

Profitable trading strategies including transaction cost

	Yield
Buy when matched volume is around 15,000	5.50%
Sell it the trading volume is below 100 (Minimum)	1.84%
Buy low odds (1-1.1) highest yield	4.54%
Sell when the odds is high (above 10.5)	58.16%
Buy odds under 2.0	-0.28%
If the odds change in value sell, it dose not matter in which direction. This finding is likely the result of the low volume strategy. Remember that over 50% of the transactions happens when the trading volume is below 100, and where the yield from buying is negative. Hence in general it is best to be a seller.	1.33%

Almost all of the trading strategies are still profitable when transaction cost is included. The yield is actually quite robust. There are opportunities to profit even when you have to pay transaction cost of the premium. It is a surprise it is possible to profit when transaction cost are included. Some of the premium goes to a third party and betting is a null sum game. There are some participants who are following a bad trading strategy. It is also indicate that the market is not totally efficient, some would say far from it.

11. Conclusion

The research questions have been developed on the basis of different observations regarding human behaviour. The main driver has been behavioural finance theory. The essential elements used in thesis are anchoring, overconfident behaviour, the certainty effect, regret theory and representativeness. The behaviour finance theory has not been especially given, and its predictions are not really reflected in the results. There is a lack of connecting between the theory and the violations of the effective market hypothesis. In the end we do not give the theories much credit for the violations of the efficient hypotheses.

The betting exchange market offer several violations of the efficient market hypotheses, some more interesting than others. There are some main tendencies that are crisscrossing. Together they provide a deeper understanding of each other.

Overall two factors outline different trading phenomena's, odds at extreme values and trading volume. Odds at extreme values differ particularly form their fundamental value. Low odds provide a too high return and high odds a too low return. It is believed that risk aversion is a good explanation.

That interpretation seems to be valid and it corresponds well with theory, or at least the part of prospect theory that deals with gains. There have not been found a methodology that reveal how people behave when they are facing losses.

There was also found is a link between yield and trading volume. Low trading volume and that is below 100 provided a negative yield, while high trading volume provides a positive yield. The yield is highest at trading volumes around 15,000.

There is a mechanism that can explain why the return is negative at low trading volume, and it is related to how people trade. The acceptance of partly covered wagers could cause the odds to differ from its fundamental value. Partly matched odds lack forward and are still available when the situations change for the worse, and when the odds should have been higher.

When trading volume is high there is a positive return for the buyers. No good explanation seems to exist for understanding that tendency. The answer cannot be found in theory. Actually it has been observed throughout history that high trading volume is a sign of a general overconfidence and a belief that the prices will raise.

The belief is often misplaced and it is quite likely that the opposite is about to happen. High trading volume is an indicator of falling stock prices. A foreseen trader would sell under high trading volume, because the stock is going to decrease in value.

Drawing a parallel to the odds market. High trading volume should be bad for the buyer. High trading volume on the odds market indicates the contrary. Our results show that at high trading volumes, buyers gain a positive yield. That goes against the experience from the stock market.

Although it did not end with a genuine yield, investigation of trends before kick-off did provide considerable insight. The slope can be positive or negative, depending on the odds development before kick-off.

If the odd when the market opens is traded at a higher value compared to before kick-off, the slope is positive. In total there was no evidence that the slope is different from zero.

Afterwards all slopes where divided between winners and losers. There was observed significantly more positive slopes on the losers. A positive slope can be used to indicate the player who is going to lose the match.

Our investigation has been limited to only look at internal factors. That means that the investigation is based only on data that has to do with odds fluctuation over time. Hence external factors are discharged as explanatory variables.

The isolated focus on the odds has its advantages. You do not have to undergo the process of quantifying a lot of outside factors. Besides it is believed that technical analyses are sufficient to establish a satisfactory analysis. It is an approach that is capable of revealing violations on the efficient market hypothesis. The stock market is a good example of that fact.

Our main objective was to link behavioural finance theory and trading behaviour on the odds market.

There are many elements in the theory that give sound reasons why and where violations on the efficient market hypothesis should be found. Going from theory to practise have not been easy. It has been difficult to formulate research questions where the logic from the theory is manifested. The violations fund in this study can often be explained by some other logic or cannot be understood at all. Violations often seem to occur spontaneously. In other words the theories are too complicated to implement.

We have used patterns in the odds fluctuation to reveal any deviations from the fundamental value. We have only briefly discussed what it take to pursuit a profitable trading strategy, but is should not be difficult to imagine that it would be both risky and time consuming. Betting is a zero-sum game and there is the risk of losing all in each bet. Repeating efforts is necessary, and it takes a lot of time to identify the right situation to and react on it.

The Internet betting exchange market violates the hypothesis of a weak efficiency market. We have been able to uncover several situations where it is possible to profit. In some situations the market functions perfectly. I think it is important to notice that odds at kick-off are at there fundamental value. A profitable trading strategy has to be refined. I do not think the market is totally chaotic. There seems to be a limitation to the ineffectiveness. The returns are not extraordinary high expect in one incident. Selling odds above 10,5 does provide a superior return. But is it also a situation where the loss can be very high compared to the potential return. That is an important point. A very unequal payoff distribution is only favourable for one of the participants. Thinking about it maybe this is the reason for the premium. Even when it is known that the expected return is negative, the potential upside is still desirable. The potential upside and the high expected yield are considered as strong incentives for the participants to wager. The violation is not so irrational as it appears.

12. References

- Ackert, L. F. and Deaves, R. 2010. Behavioural Finance: Psychology, Decision-Marking, and Markets. International Edition. South-Western Cengage Learning.
- Agresti, A. And Franklin, A. C., 2007. Statistics: The art and science of learning from data. Person Prentice Hall.
- Allais, M. 1953. Le Comportement de l'Homme Rationnel devant le Risque, Critique des Postulats et Axiomes de l'Ecole Americaine. Econometrica, 21, 503-546.
- Alpert, M., and H. Raiffa 1982. A Progress Report on the Training of Probability Assessors, in Judgment Under Uncertainty: Heuristics and Biases, pp. 294–305, Cambridge. Cambridge University Press.
- Amos T. and Daniel K., 1974. Judgment under uncertainty: Heuristics and biases. Science, 185 (4157).
- Bar-Hillel M. 1973. On the Subjective Probability of Compound Events. Organizational Behaviour and Human Performance, 9 (396-406).
- Barberis, N. C. and Thaler. H. 2003. A survey of behavioural finance. Handbook of the Economics of Finance: Volume 1B, Financial Markets and Asset Pricing. Elsevier North Holland, Chapter 18, pp. 1053–1128.
- Betfair. Charges. [online]http://sports.betfair.com/da/aboutUs/Betfair.Charges/ [Accessed 21 Sep 2011].
- Betfair. History. [online] Available at: http://corporate.betfair.com/about-us/history/history-archive.aspx [Accessed 15 May 2011].
- Edwards, W. 1968. Conservatism in Human Information Processing. In Kleinmuntz, B. Formal Representation of Human Judgment. Wiley.
- Fama, E F. 1998. Market efficiency, long-term returns, and behavioural finance, Journal of Financial Economics 49 283—306.
- Fischhoff, B., P. Slovic, and S. Lichtenstein. 1977. "Knowing with certainty: The appropriateness of extreme confidence," Journal of Experimental Psychology: Human Perception and Performance, 3, 552-564.
- Fuglesang, L. Hageorn-Rasmussen, P. and Olsen, P.B., 2007. Teknikker I samfundsvidenskaberne. Roskilde Universitetsforlag.
- Jegadeesh N. and Titman S. 2001. Profitability of Momentum Strategies: An Evaluation of Alternative Explanations. The journal of finacne Vol. LVI N0. 2 April
- Kahneman, D. and Tversky, A. 1974. Judgment under Uncertainty: Heuristics and Biases. Science, New Series, 4157: 1124-1131
- Kahneman, D. and Tversky, A. 1979. Prospect theory: an analysis of decision under risk, Econometrica, 47:263−291.
- Kahneman, D. and Tversky, A. 1981. The framing of decisions and the psychology of choice. Science, 211, January.
- Koppel, R.and Abell H., 1994. The Innergame of TRADING. United States of America: BB.

- Lord C. G., Ross L., and Lepper M. R. 1979. Biased Assimilation and Attitude Polarization: The Effects of Prior Theories on Subsequently Considered Evidence. Journal of Personality and Social psychology, Vol. 37, No. 11, 2098-2109
- Monitier, J. 2003. Behavioural Finance: Insights into irrational minds and markets. John wiley & sons, LTD.
- Pedersen, C. V. And Plenborg, T. 2005. Regnskabsanalyse for beslutningstagere. Thomson
- Svenson, O., 1981. Are we all less risky and more skilful than our fellow driver?. Acta Psychological 47.
- Tvede, L. 1999. The Psychology of Finance. 2nd ed. England: John Wily & Sons Ltd.
- WERNER F. M. De BONDT and RICHARD THALER 1985* Does the Stock Market Overreact? THE JOURNAL OF FINANCE * VOL. XL, NO. 3 * JULY
- Whewell, William. 1837. History of the Inductive Sciences, from the Earliest to the Present Times. 3 vols, London. 2nd ed 1847. 3rd ed 1857. 1st German ed 1840-41.
- Woods A. S. and West M. A., 2010. The psychology of work and organizations. Thomas Rennie.
- Yudkowsky S. E. An Intuitive Explanation of Bayes' Theorem [online] http://yudkowsky.net/rational/bayes [Accessed 27 May 2011].

13. Appendix 1

The likelihood that the urn is containing 2/3 red balls and 1/3 white balls given 2 specific draws. That would be the likelihood of the conditioned probability *P (content|draw)*. This problem is solved by the of use Bayes' law and common probability theory. It is assumed that the ball is put back after each draw.

Definitions:

P (the urn contains 2/3 red balls and 1/3 white balls |draw)=P (content|draw)

P (draw|the urn contains 2/3 red balls and 1/3 white balls)=P (draw|content)

P (draw|the urn contains 1/3 red balls and 2/3 white balls)=P (draw|¬content)

Bayes' Law (Yudkowsky S. E online):

P (content|draw)= P (content & draw)/(P (draw & content)+ P (draw & ¬content))

First Incident: 4 red and 1 white:

P (draw|content)=5!/(4! • 1!) • (2/3)^4 • 1/3=80/243

P (draw|¬content)= 5!/(4! • 1!) • (1/3)^4 • 2/3=10/243

P (content|draw)= 80/243/(80/243+10/243)=8/9

Second Incident: 12 red and 8 white:

P (draw|content)=(20!)/(12! • 8!) • (2/3)^12 •(1/3)^8=125970 • 2^12/3^20

P (draw|¬content)= (20!)/(12! • 8!) • (1/3)^12 • (2/3)^8=125970 • 2^8/3^20

*P (content|draw)= (125970 • 2^12/3^20)/(125970 • 2^12/3^20+125970 • 2^8/3^20)=**16/17***

<u>16/17 > 8/9</u>

Conclusion it is more likely that the urn contains 2/3 red and 1/3 white balls based on the drew of 12 red and 8 white, compare to 4 red and 1 white balls.

14. Appendix 2

Definitions

Initial wealth: W

EU: Expected utility

U: Utility function, in this situation it is unknown but it is unimportant.

As a result of the subjects preference:

$$EU=U(W+3000)*1 > EU=U(W+4000)*0,8$$

and,

$$EU=U(W+3000)*0,25 < EU=U(W+4000)*0,2$$

The expressions are rearranged to clarify that this is a violation of the Exp. Utility framework.

$U(W+3000) / U(W+4000) > 0,8/1 > 0,8$ (first choice)

$U(W+3000) / U(W+4000) < 0,20/0,25 < 0,8$ (second choice)

Clearly both can be true.

15. Appendix 3

Results when we examine if odds at extreme situations are either too low or too high.

Odds Equal	Win (counts)	Lose (Counts)	P-value	Yield
1,01	5307	7	6.20E-65	1%
1,02	5252	14	8.23E-120	2%
1,03	5371	20	7.70E-191	3%
1,04	5293	32	4.87E-190	3%
1,05	5069	66	1.29E-103	4%
1,06	4678	108	3.21E-55	4%
1,07	4843	104	2.14E-100	5%
1,08	4648	175	6.93E-44	4%
1,09	4331	218	1.63E-27	4%
1,1	4719	397	3.77E-04	1%
1,11	4501	545	4.16E-02	-1%
1,12	5328	679	1.49E-01	-1%
1,13	4471	842	5.69E-18	-5%
1,14	4647	855	2.76E-11	-4%
1,15	5090	830	3.05E-02	-1%
1,16	4916	914	7.67E-05	-2%
1,17	4388	710	2.14E-01	1%
1,18	4302	735	1.83E-01	1%
1,19	3914	662	3.95E-03	2%
1,2	4521	918	6.77E-01	0%
6	249	1433	0.03159618	-11%
6,5	0	0	#DIV/0!	#DIV/0!
7	72	511	0.155953569	-14%
7,5	0	0	#DIV/0!	#DIV/0!
8	174	1114	0.289478528	8%
8,5	0	0	#DIV/0!	#DIV/0!
9	21	194	0.507624894	-12%
9,5	0	0	#DIV/0!	#DIV/0!
10	113	1079	0.539975177	-5%
10,5	15	185	0.278506656	-21%
11	78	1149	9.15171E-05	-30%
11,5	0	80	#DIV/0!	-100%
12	40	966	2.84176E-12	-52%
12,5	6	228	3.2374E-07	-68%
13	1	74	7.86304E-06	-83%
13,5	37	878	2.94583E-07	-45%
14	1	218	2.04934E-34	-94%
14,5	0	76	#DIV/0!	-100%
15	14	889	6.33838E-33	-77%

15,5	3	244	1.05029E-12	-81%
16	0	166	#DIV/0!	-100%
17	0	72	#DIV/0!	-100%
18	1	299	6.19335E-41	-94%
18,5	0	43	#DIV/0!	-100%
19	0	49	#DIV/0!	-100%
19,5	0	14	#DIV/0!	-100%
20	0	76	#DIV/0!	-100%
50	4	630	1.56007E-05	-68%
100	0	375	#DIV/0!	-100%
200	0	130	#DIV/0!	-100%
500	0	39	#DIV/0!	-100%
1000	0	40	#DIV/0!	-100%

16. Appendix 4

Discount rates on Betfair. Trading volume triggers Betfair points and they are the basis for calculation the discount on the commission. The activity of betting per week regulates the amount of Befair points. You earn one point for every 10 pence you have paid in commission.

EARNED BETFAIR POINTS AND THEIR ASSOCIATED DISCONT

From	To	Discount on the commission
0	999	0%
1000	2,499	2%
2500	3,999	4%
4000	5,499	6%
5500	6,999	8%
7000	8,999	10%
9000	10,999	12%
11000	12,999	14%
13000	14,999	16%
15000	16,999	18%
17000	18,999	20%
19000	20,999	22%
21000	22,999	24%
23000	25,999	26%
26000	28,999	28%
29000	31,999	30%
32000	35,999	32%
36000	39,999	34%
40000	43,999	36%
44000	48,999	38%
49000	54,999	40%
55000	60,999	42%
61000	66,999	44%
67000	72,999	46%
73000	79,999	48%
80000	87,999	50%
88000	95,999	52%
96000	107,999	54%
108000	125,999	56%
126000	149,999	58%
150000	-	60%