Analysing Efficiency of Predictive Sports Markets

Zdeněk Syrový

Supervisor: Ing. Gustav Šír
May 2020
I. Personal and study details

Student's name: Syrový Zdeněk
Personal ID number: 474404
Faculty / Institute: Faculty of Electrical Engineering
Department / Institute: Department of Cybernetics
Study program: Open Informatics
Branch of study: Computer and Information Science

II. Bachelor's thesis details

Bachelor's thesis title in English: Analysing Efficiency of Predictive Sports Markets
Bachelor's thesis title in Czech: Analýza efektivity prediktivních sportovních trhů

Guidelines:
While there is a variety of literature discussing efficiency of predictive markets in generic economic terms, the amount of scientific work with actually testable hypotheses is scarce. Most of the available research is based on creating statistical models designed to achieve profits in some market, invalidating the efficiency hypothesis as a consequence of potential success. The goal of this work is not an attempt to create a profitable model, but to analyse existing approaches by sound statistical means on as large corpora of market data as possible, to identify weaknesses and typical biases in existing market efficiency testing.

1) Review existing literature on prediction market efficiency, identify possibilities for empirical testing.
2) Identify possible weaknesses and biases in existing testable approaches.
3) Propose several methods for statistical hypothesis testing of the biases and market (in)efficiency.
4) Research existing online data sources for predictive sports markets (bookmakers, tipsters).
5) Scrape and parse suitably large datasets for statistical hypotheses testing over different markets.
6) Setup suitable testing framework, evaluate proposed hypotheses, and provide confidences to your findings.
7) Analyze and discuss your results.

Bibliography / sources:

Name and workplace of bachelor's thesis supervisor:
Ing. Gustav Šír, Intelligent Data Analysis, FEE

Name and workplace of second bachelor's thesis supervisor or consultant:

Date of bachelor's thesis assignment: 10.01.2020   Deadline for bachelor thesis submission: 25.05.2020
Assignment valid until: 30.09.2021
III. Assignment receipt

The student acknowledges that the bachelor's thesis is an individual work. The student must produce his thesis without the assistance of others, with the exception of provided consultations. Within the bachelor's thesis, the author must state the names of consultants and include a list of references.

______________________________________________
Date of assignment receipt

______________________________________________
Student’s signature
Author statement for undergraduate thesis:
I declare that the presented work was developed independently and that I have listed all sources of information used within it in accordance with the methodical instructions for observing the ethical principles in the preparation of university theses.

Prague, May 22, 2020

.................................
Signature
Acknowledgements

I would like to thank my supervisor Gustav Šourek for his never ending optimism and patience.
Abstract

In this work we aim to analyze efficiency of sports betting markets by statistical means. For that purpose, we rely on a large collection of historical records across variety of sports, markets, and bookmakers. Subsequently, we utilize diverse statistical testing methods and generic betting strategies to discover potential inefficiencies and biases in the markets. Additionally, we analyze a set of successful bettors to find out whether their success can be attributed to true skill or mere luck.

Keywords: sports betting, market efficiency, statistics, hypothesis testing

Supervisor: Ing. Gustav Šír
Contents

1 Introduction ........................................ 1
   1.1 Definitions ............................... 1
       1.1.1 Bookmaker .......................... 1
       1.1.2 Bettor ............................. 2
       1.1.3 Betting ............................. 2
       1.1.4 Odds ............................... 2
       1.1.5 Bookmaker margin .................. 3
       1.1.6 Types Of Odds ..................... 3
   1.2 Biases .................................. 5
       1.2.1 Favourite–Long-Shot Bias ....... 5
       1.2.2 Survival Bias ....................... 5
       1.2.3 Outcome Bias ....................... 5
   1.3 Related Work ............................. 5

2 Market Efficiency ........................... 7
   2.1 Betting Market ........................... 7
   2.2 Degree of Efficiency ...................... 7
       2.2.1 Weak Form Efficiency .......... 7
       2.2.2 Semi-strong Form ............... 8
       2.2.3 Strong Form ....................... 8
   2.3 Efficient Market in Betting ............ 8
   2.4 Refuting the Efficient Market Hypothesis ........................................... 8
   2.5 Generating Method ....................... 9
   2.6 Economic Tests ........................... 10
   2.7 Opening and Closing Odds ............. 10
   2.8 K-L-Divergence .......................... 11
   2.9 Regression Tests ....................... 11
   2.10 Bettors on the market ................ 13
   2.11 Group test ................................ 14
       2.11.1 Table ............................. 14
       2.11.2 Probabilities of Elements ... 15
       2.11.3 Asian handicap point five tables .............................................. 16
       2.11.4 System of inequalities .......... 17
       2.11.5 Finding a solution .......... 17
   2.12 Luck vs skill ........................... 17
       2.12.1 Importance of binomial distribution in betting ......................... 18

3 Data ............................................ 21
   3.1 Bookmaker odds .......................... 21
   3.2 Bettors history data .......................... 21
   3.3 Blogabet data ............................ 22
       3.3.1 Crawling the Blogabet .......... 22
   3.4 Rebelbetting data ............................ 23

4 Results ....................................... 25
   4.1 Bookmakers odds .......................... 25
       4.1.1 1x2 Odds .......................... 25
       4.1.2 Over-under Odds .................. 28
       4.1.3 Both-to-score Odds ............. 30
       4.1.4 Home-away Odds ................. 32
       4.1.5 Random Bettor ..................... 32
       4.1.6 Probability Method ............ 32
       4.1.7 Using all the odds available ... 34
       4.1.8 Explanation ........................ 34
   4.2 Blogabet tipsters data .................. 35
       4.2.1 Results of t-test ................. 35
       4.2.2 Results of Rebelbetting ......... 36

5 Conclusion ................................... 37
   5.1 Future Work ............................. 37

Bibliography .................................. 39

A Longer t-test Table .................... 41

B Other Sources of Tips Considered .... 43
Figures

2.1 Showing decrease of percent standard deviation ............ 18
3.1 Shows basic structure of the Blogabet tipster page. On top of the page there is some basic information about the bettor and in the middle there is a table with all the picks, which continues all the way down to the bottom of the page. ......... 23
3.2 Shows the bottom of the page with the button one need to programatically activate to load more data. ......................... 23
4.1 Cumulative sum of betting on football 1x2 Pinnacle Odds .... 26
4.2 Cumulative sum of betting on football 1x2 Chance.cz Odds ...... 26
4.3 Difference between Pinnacle random bettor and Tipsport random bettor on 1x2 football odds ....... 27
4.4 Cumulative sum of betting on Over-under Pinnacle Odds ...... 28
4.5 Cumulative sum of betting on over-under Pinnacle Odds ...... 29
4.6 Cumulative sum of betting on Over-under Pinnacle Odds ...... 29
4.7 Difference between pinnacle random bettor and Bet-at-Home random bettor on over-under baseball odds ....................... 29
4.8 Cumulative sum of betting on Both-to-score Bwin Odds ....... 30
4.9 Cumulative sum of betting on Both-to-score BetVictor Odds .... 31
4.10 Difference between Betvictor random bettor and Bwin random bettor on both-to-score football odds ......................... 31
4.11 Cumulative sum of betting on home-away 188BET Odds ....... 32
4.12 Cumulative sum of betting on home-away Pinnacle Odds ...... 33
4.13 Difference between 188BET random bettor and Pinnacle random bettor on home-away hockey odds 33

Tables

1.1 AH handicap CTU example ...... 4
1.2 Indicates to which result the box belongs ..................... 15
1.3 Indicates to which result the box belongs ..................... 15
1.4 Indicates to which result the box belongs for over-under limit of 2.5 16
1.5 Indicates to which result the box belongs for -1.5 home team handicap 16
3.1 Available odds for each sport ... 21
4.1 Summarizing the dataset ...... 35
4.2 T-test result of the most followed tipsters on blogabet .......... 35
A.1 T-test Result from the top 90 available tipsters .............. 42
Chapter 1

Introduction

Betting has been with us for centuries in the form of various prediction markets ranging from sports and economics to politics. As the predictive accuracy continuously increases, the markets are becoming more efficient year by year. This trend comes as no surprise since the amount of data we have today is incomparable to the amount we had 40 years back, let alone few centuries back. The data gives us a perfect opportunity to use various statistical methods and machine learning to predict the outcomes of individual events. However, building a successful model from data is not easy, since for a successful betting, one needs to directly beat other models on the highly competitive markets. In this work, we focus on how to analyse the models and markets using diverse statistical methods and strategies aimed at both the models coming from the bookmakers, and from the bettors.

1.1 Definitions

Let us start with some definitions of the sports-betting related concepts which we use in this thesis.

1.1.1 Bookmaker

A bookmaker is a person or an organisation that offers odds for particular outcomes of stochastic events, and allows other subjects on the market to bet, i.e. place wagers on the outcomes associated with the offered odds.

A bookmaker is a true player and wants to maximise his/her profit. Maximising profit can then be achieved by either setting a large margin or by estimating the probabilities, and the associated odds, very well.

The bookmakers can be generally divided into two categories. First are so-called “soft bookmakers”, who usually have a larger margin on the odds. Very often the predictions of soft bookmakers are less accurate and they adjust odds slowly. It is also common that skilled bettors are banned from such bookmakers as they would make too big of a profit. The other category are “sharp bookmakers”, who offer lower margin in exchange for a higher volume of bets, allowing for a more precise and responsive adjustment of the

\[1\] We explain this notion later in Section 1.1.5.
1. Introduction

odds w.r.t. the joint information possessed by the bettors and models on the market.

1.1.2 Bettor

A bettor is a subject on the market that takes the offered odds by bookmaker and bets on them. Each bettor wants to again maximize his profit. The typical way for a bettor to gain profit is via beating the bookmakers in estimating the probabilities (finding “value bets”), or by exploiting the market by finding odds from different bookmakers that ensure positive profit – an approach called arbitrage betting (Jordan 2020). Recently, a new service emerged in the form of websites that advise the bettors in what bets to take. These websites often have automated systems of finding value bets from many different bookmakers. However, the websites target profit as well, for which they typically offer their services for some monthly subscription fee.

1.1.3 Betting

Betting is a game between the bookmakers and the bettors. Bookmaker estimates probabilities of sport event results and offers odds for the bettor. Bettor then places wagers to a subset of the offered odds by the bookmaker, while also predicting their results by own means. If the bettor guessed correctly, the bet is successful, and the bookmaker must pay off the winning to the bettor. The amount won is decided by the odds and the wager. If the bettor guessed wrongly, the bookmaker wins and keeps the wager.

1.1.4 Odds

Odds occur in many shapes and forms. The most common type in Europe are so-called decimal odds. These odds can be understood as inverse probabilities of each outcome $i$ (Equation 1.1) of an event (game), however, they typically do not sum to 1 because of the bookmaker’s margin (Sec. 1.1.5). The odds are fair when the sum of their inverse (probabilities) is exactly 1. If a bet is successful, then the amount won by the bettor is decided by multiplying the deposit and the odds. A net profit for a successful bet is then calculated by Equation 1.2 while, for an unsuccessful bet, the bettor loses the full deposit.

Also, the bookmaker constantly changes the odds according to other subjects on the market. This can prevent bettors from finding exploitable opportunities in the system. Recognising well-informed and skilled bettors is also crucial for the bookmaker. Adjusting to bettors can be seen in the difference between opening (first estimate) odds and closing odds (last odds just before the start of a match).

$$P_i = \frac{1}{odds_i} \quad (1.1)$$

$$profit = deposit \cdot odds - deposit = deposit \cdot (odds - 1) \quad (1.2)$$
1.1.5 Bookmaker margin

When a bookmaker estimates probabilities, he adds a certain margin to the odds, making the odds unfair for the bettor. There are many possibilities of how the bookmakers can divide this margin, e.g. uniformly, or according to a favourite-long-shot bias (Sec. 1.2.1). The margin of a bet with odds \( o \) can be calculated as

\[
\text{margin} = \left( \sum_{o \in \text{odds}} \frac{1}{o} \right) - 1 \quad (1.3)
\]

And a uniformly distributed margin can then be removed from the odds, using the following formula:

\[
P_{\text{fair}} = \frac{P_{\text{unfair}}}{\sum_i P_i} \quad (1.4)
\]

where \( P_i \) are odds converted via Equation 1.1, \( P_{\text{fair}} \) is a probability of outcome according to odds, and \( P_{\text{unfair}} \) is the original inverse odds, associated with the same outcome (Eq. 1.1). Basically, it is just simple normalization of the values to form a proper probability distribution.

1.1.6 Types Of Odds

A typical bookmaker offers odds for many matches from different sports, but even for an individual match, we can further divide the odds reflecting many different outcomes that can occur.

1x2 Odds

The most basic and common type of odds are “1x2”. The bookmaker offers three odds for a given match, representing the (1) home team win, the (x) draw and (2) the away team win, respectively. A bettor picks one of these odds, and according to the result he either wins or loses the bet.

Asian Handicap Odds

In asian handicapping, the bookmaker tries to overcome the relative difference between the teams in the match by giving the underdog an advantage before the match begins, or by giving the favourite a disadvantage. The advantage/disadvantage is called handicap and takes the form of a positive/negative number that is added to the final score of the side that the handicap belongs to. Bookmaker then offers two odds for each match to choose from – home team win or away team win.

In the most simple case, a handicap is an integer and there are only two possible outcomes of a game: home win and away win. In the case that the bettor guesses the winner correctly, he wins the bet, and if the guess is wrong, he naturally loses the bet, however, if the match was drawn, then the stakes are refunded. Another type of handicap is a “.5 handicap”, where the numbers
are half-way between two integers. In this instance, a draw is no longer a possibility as the score of a half of a point/goal is not achievable in any of the given sports. The last and most complex type is a quarter handicap. It is a combination of the two above, where half of the bet goes to an integer \( i \) handicap bet and the rest to the 0.5 handicap \( f \), with a constraint that:

\[
|i - f| = 0.5 
\]  

(1.5)

The notion from the bookmaker is either to give both handicaps that bet is divided to (an integer and a number higher or lower by a half, e.g. (0,0.5)) or to give the mean of two numbers (0.25).

Let us demonstrate this on an imaginary example. Let us have a hockey match between CTU and Charles University where the bookmaker’s model predicts that CTU team is slightly favourite and gives the CTU team a -0.25 handicap and offers odds 2 for both teams (note we use no margin here). Now imagine the bettor wants to stake two units of some currency. The possible gains/losses of this bet can be seen in Table 1.1.

<table>
<thead>
<tr>
<th></th>
<th>CU Wins</th>
<th>Draw</th>
<th>CTU Wins</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bet on CTU</td>
<td>-2</td>
<td>-1</td>
<td>2</td>
</tr>
<tr>
<td>Bet on CU</td>
<td>2</td>
<td>1</td>
<td>-2</td>
</tr>
</tbody>
</table>

**Table 1.1:** AH handicap CTU example

Importantly, note that the odds-implied probabilities in the context of an asian handicap are conditional probabilities, given that the match does not end as a draw, and not probabilities of the home/away team winning.

- **Home-Away Odds**

In this type of odds, bettors can choose from *two* possible outcomes of the game. Either the home team win or the away team win. A bettor wins if the outcome meets his choice and in case of a draw, all stakes are refunded. This market can also be seen as a special type of Asian handicap market with a handicap of 0.

- **Both-to-Score Odds**

Both to score (bts) is the first type of odds where bettor does not directly bet on the outcome of a match. The goal here is to predict whether both teams will score in the match. Also, bts is more restricted to the sports available, where a typical sport for this kind of market is football, as the occurrence of results with a maximum of one team scoring is common, and probabilities are not that extreme as would be in the case of, e.g., basketball.

- **Over-Under Odds**

The over-under is similar to both to score, but instead of betting whether both teams will score, the bettor stakes on whether the cumulative score of
the match will be over or under some offered limit.

## 1.2 Biases

While the markets are generally increasingly more precise and efficient, many studies have found systematic errors in the sports betting markets. These non-random errors are called *biases*. Additionally, many biases are also prevalent amongst the bettors, making them behave in a certain sub-optimal fashion. Finding biases in a given market decreases its reliability to serve as a match outcome predictor, and consequently offers opportunities for exploitation.

### 1.2.1 Favourite–Long-Shot Bias

The most connected bias with sports betting is the favourite–long-shot bias. It is a tendency to overvalue the chances of a less probable event happening. In betting, people often tend to bet on less probable events (long-shots) with a vision of fortune, but in reality, they often lose way more money than they would in backing up the favourites. Direr 2011 and Deschamps and Gergaud 2007 found this bias in odds provided by bookmakers and marked it as an inefficiency.

### 1.2.2 Survival Bias

An error of overlooking failures and concentrating only on the successful part of sample population is called the survival bias. This error can lead to a too optimistic view on a historical betting performance record. In the most extreme case, it can look like everyone is winning, but this is only an illusion because the rest simply did not survive, for which the statistics are not available or are not shown. This bias is more of a threat for analysing the history of bettors because, from a testing sample of successful bettors, it could look like the market is very inefficient. However, all of those records could be just lucky survivors, whereas the rest that was less lucky is simply missing.

### 1.2.3 Outcome Bias

The outcome of a bet is either a win or loss, and that can lead to the idea that if the bet was successful, then it had to be good. This is obviously false because betting is based on probability. Therefore, a mere luck could be behind a successful bet, and vice-versa, an unsuccessful bet could be caused by bad luck.

## 1.3 Related Work

Many authors tried to confirm or refute the efficient market hypothesis with different results. In (Cain, Law, and Peel 2000), the authors found
ineffectiveness in the form of favourite-longshot in UK football betting. In (Deschamps and Gergaud 2007), the authors also found evidence of favourite-longshot bias and draw bias in football using smaller odds. In (Kain and Logan 2014), a seemingly unrelated regression structure was used on NBA, NFL and college matches to find some inefficiencies. Semi-strong efficiency was then tested in (Bernardo, Ruberti, and Verona 2015) by observing movement of the odds after some significant changes to the teams prior to the match. Weak form inefficiency in low and high probabilities was found in (Titiu 2016) using European football matches. (Smith, Pato, and Williams 2006) studies effect of person to person betting on a classic bookmaker market. In (Paul and Weinbach 2002), log-likelihood ratio test was used to test over-under odds and showed that betting under 5+ leads to profit on data from 1979 to 2000. In contradiction to other studies, (L. M. Woodland and B. M. Woodland 2003) found reverse favourite-long-shot-bias (where favourites are overpriced) on the MLB data from 1990 to 1999. From another perspective, (Snowberg and Wolfers 2010) focused on explaining favourite-long-shot bias through whether bettors are risk-loving or probabilities are misinterpreted by bettors.
Chapter 2
Market Efficiency

2.1 Betting Market

A betting market is a place where the bookmakers and bettors meet. Bookmakers offer odds for many different sports and events from which bettors can choose. A bookmaker creates the market, while the bettor chooses what opportunities to invest in. As the bettor wants to make as good of a choice as possible, he typically compares the markets and opportunities offered over several bookmakers. Analyzing efficiency of the individual markets and bookmakers is a generally important statistical factor, since the higher the efficiency of a given market, the less of an opportunity for the bettor to make a profit.

2.2 Degree of Efficiency

The market is efficient if it reflects all information available in the available prices. In (Malkiel and Fama 1970), the authors divided market efficiency into three categories of (i) weak form, (ii) semi-strong form, and (iii) strong form in terms of a stock market. This terminology, however, is relevant even for sport prediction markets as there are many similarities.

2.2.1 Weak Form Efficiency

The market is weak-form efficient if it reflects all the past information available. That means past results cannot be used to get an advantage over the market. In sports betting, past information is usually used to predict the future match outcomes through the machine learning approaches. Once the predictive models reach their full potential, seeing a weak-form efficient market would be theoretically possible.

2.2.2 Semi-strong Form

In the semi-strong efficiency, the market reflects all new public information as well as past information, making it immune to analysis as everything gained
through analysis should already be counted for in prices. However, the market could still be exploited by using private information.

### 2.2.3 Strong Form

Private information is then accounted for in the strong form efficiency. Strong form basically means that prices already reflect all pieces of information available and making any profit is thus not possible.

### 2.3 Efficient Market in Betting

The efficient market hypothesis (EMH) states that the market reflects all the information. In betting terminology, this means that bookmaker estimates the probabilities so well, that making a consistent profit in the long run is impossible with skill alone. Efficiency is impossible to measure directly from the definition because the sheer amount of information is overwhelming, and it is defined only vaguely. Thus instead of proving that EMH is correct, it is better to define methods that could help us confidently refute the efficient market hypothesis.

### 2.4 Refuting the Efficient Market Hypothesis

One of the most intuitive ways to refute the efficient market hypothesis is to build a predictive model that is better than the model of a bookmaker. These days, with all the various machine learning methods and approaches, one could say that building such a model is manageable. However, bookmakers are aware of that possibility and a battle that a long time ago was about two humans predicting an outcome is now about two machine learning models. In this battle, the bookmaker seems to be favoured as the number of data the bookmakers have available is typically larger than that of a typical bettor. Nevertheless, the question here is whether the sheer amount is that important, and if, e.g. one and a half-year-old, statistics are relevant for the outcome of future games. Another question is what numbers of past games to use to predict the future result. Then come some other typical machine learning problems such as over-fitting of training data and computational complexity. Moreover, when a bettor can overcome all of those problems, the life expectancy of a successful model is also not long, as every subject on the market tries to improve their model daily, especially when a bettor builds a good model and the bookmaker reacts to minimise the loses.

In this work, we refrain from building a custom predictive model, and rather aim to analyze the efficiency of the existing models and markets by solid statistical means.
2.5 Generating Method

The most straightforward method to test the EMH is using randomly generated bets and seeing what fraction of the bets was “successful” in making a profit. There are many approaches to this method. The most straightforward one is to simply choose a random opportunity from all the available bets. Then there is a possibility to add more degrees of freedom, e.g., by generating wager sizes or choosing a random sub-sample from the available bets.

In (Buchdahl n.d.(b)), a Monte-Carlo simulation in betting was tried, even though not for testing of the EMH. This method gives us an intuition of how hard it is to be in profit without any information. If a high number of bettors can consistently beat the market without any information, then it is not very efficient indeed. Nevertheless, the opposite case where the bettors without any information cannot beat the market, does not necessarily mean that the market is efficient.

A big disadvantage of this method is the bookmaker’s margin. The larger it is the less information can random bettor provide, which is quite counter-intuitive, as normally we would expect that bookmakers with larger margin would be less confident in their predicting skill. Another addition to the method is thus removing the bookmaker’s margin and working with probabilities only.

There are more ways on how to remove the margin. The easiest one is using Equation 1.4. According to (Buchdahl 2016), another possibility of how the margin can be distributed is proportionally to the odds. The higher the odds, the higher the margin, which can now be removed using the following formula:

\[
\text{fair odds} = \frac{\text{outcomes} \cdot \text{odds}}{\text{outcomes} - \text{margin} \cdot \text{odds}}
\]  

(2.1)

With the different ways on how to remove margin, we now have 4 degrees of freedom of:

1. which matches to consider
2. which outcomes to take
3. how much to stake
4. consider the margin

These are the very questions every bettor asks before staking. Now we can run a simulation and see how well an average random bettor does. We note again that this method can help us to refute EMH but not to confirm it.

Another version of this method is following the “wisdom of the crowd”. This process uses collective opinion as a betting strategy, staking on the most voted outcome. Joseph Buchdahl used this method in his book (Buchdahl 2016), and demonstrated some good results.
2. Market Efficiency

2.6 Economic Tests

Following from the random strategy, we can move further to other simple strategies in trying to make profit. The simple strategies are defined by using only the odds available from the matches. One such a strategy is the max-odds, i.e. wagering on the highest odds for each match. The reverse strategy is wagering on the lowest odds available. Other simple strategies include constant betting on home team win, away team win and the draw (if available), respectively.

In the Section 1.2 about biases, we introduced the favourite-long shot bias as a form of possible inefficiency. This bias could then be easily discovered using these simple strategies. For that we can test profit of the strategies over multiple markets and evaluate whether they do about the same across the different markets.

2.7 Opening and Closing Odds

Opening odds are the first estimates of probabilities made by the bookmaker. However, every decent bookmaker makes changes in odds according to bettors distribution and other factors, which results in changing odds in time. Closing odds are odds right before a match begins, and in the literature are considered as the best representation of a market’s predictive ability, because closing odds should reflect more information than the opening ones. This means, by the very definition (Sec. 2.2), more efficient.

To test this, let us first define an advantage derived from the odds movement as:

\[
\text{advantage} = \frac{\text{used odds}}{\text{closing odds}}
\]  

(2.2)

In testing, we first calculate an average advantage over all available bets using opening odds, instead of the used odds. Then we can estimate standard deviation in the expected advantage. Subsequently, we want to relate the results to sample size. In Section 2.13.1 about binomial distribution, we will show that the standard deviation is proportional to the inverse of the square root of the number of bets. From here, we calculate the bettor’s average expected advantage. A distance between mean and expected advantage in multiples of standard deviation can now give us an idea of how lucky/unlucky must a bettor be to get a particular result. An intuition for this method is that if the bettor is good then he brings a new piece of information to the market and bookmaker changes his estimate to reflect this new information, consequently. The main advantage of this method is that it requires a smaller sample size to give some statistically significant results.
2.8 K-L-Divergence

Another method was suggested in Cover and Thomas [2012], where the authors defined market efficiency in terms of Kullback-Leibler divergence as:

\[
\text{efficiency} = 1 - \frac{\min(d_s(P_r||P_b), \log(n))}{\log(n)}
\]  

(2.3)

where

\[
d_s(p_r||p_b) = \frac{\sum_{i=1}^n D(P_{ri}||P_{bi})}{n}
\]  

(2.4)

\[
D(p_r||p_b) = \sum_{\text{outcomes}} p_r(\text{outcome}) \log \left( \frac{p_r(\text{outcome})}{p_b(\text{outcome})} \right)
\]  

(2.5)

where \(P_r\) is the probability of an outcome (meaning 1 or 0 in case of most bets), \(P_b\) is the bookmaker-estimated probability, and \(n\) is a number of bets in a sample. The value of [2.3] goes from zero to one as the market goes from random to efficient. The important part here is that the method accounts for the number of possible outcome, otherwise it would inherently favour less outcome odds (and correspond to a mere crossentropy).

2.9 Regression Tests

Very typical methods form EMH testing in the literature are regression tests. These are used to see if a bettor brings any new information to market. For this purpose, we use the most simple linear probability model which has the following form

\[
y = a + b \cdot mp + \epsilon
\]  

(2.6)

where \(mp\) is the market prediction, \(y\) is the outcome, \((a, b)\) are parameters estimated by the linear regression, and \(\epsilon\) is a normally distributed error. The parameter \(a\) in Equation [2.6] can be considered a systematic bias and could be tested in the null hypothesis. However, the result from this test would give us less information than some other possibilities described further below.

Kuper [2012] suggested usage of a linear probability model in determining whether the fans collectively bring something new to the market. For the purpose of testing the null hypothesis, it is defined as:

\[
H_0 : y = a + b \cdot est + c \cdot mp + \epsilon, b = 0
\]  

(2.7)

\[
H_1 : y = a + b \cdot est + c \cdot mp + \epsilon, b \neq 0
\]  

(2.8)

where \(y\) is a dummy variable – set to 1 when the market-implied favorite won and zero otherwise, \(est\) is an estimated probability (by some model or bettor) of the favorite winning(max-odds outcome), and \(mp\) is the market implicated probability of winner. The parameters \(a, b, c\) are then the regression coefficients, and \(\epsilon\) is the error that we assumed to be normally distributed. The null hypothesis says that market had sufficient information and model
have not brought anything new. To confidently test this hypotheses, a t-test can be used in the following form:

\[ t_{\text{score}} = \frac{(b - 0)}{SE_b} \]  

(2.9)

where \( SE_b \) is the standard error of the slope coefficient described formally as:

\[ SE_b = \frac{\sqrt{\frac{1}{n-2} \cdot \sum_{i=0}^{n}(y_i - \hat{y}_i)^2}}{\sqrt{\sum_{i=0}^{n}(\text{est}_i - \text{est}_i)^2}} \]  

(2.10)

Equation 2.9 follows student t-distribution with \( n - 2 \) degrees of freedom, and from the \( t_{\text{score}} \) we can then compare to a \( p \)-value of our confidence level. For the \( \hat{y} \) we then have the following formula:

\[ P(y = 1 | \text{est}, mp) = \hat{y} \]  

(2.11)

where \( \hat{y} \) can be interpreted as the probability of a favorite winning according to this regression model. However as the coefficients are not restricted, values of \( \hat{y} \) can reach values generally higher than 1, which is a dangerous property of this method.

In Equation 2.11 we showed how the linear regression can be interpreted as a probability, however with the flaw of the possibility to have values higher than 1. This drawback can be eliminated by transforming our assumed linear dependency by an appropriate function.

\[ P(y = 1) = \phi(a + b \cdot \text{est} + c \cdot mp + \epsilon) \]  

(2.12)

An example of such a function can be the logistic sigmoid:

\[ f(x) = \frac{1}{1 + e^{-x}} \]  

(2.13)

Applied to our case we then get:

\[ y = \frac{1}{1 + e^{-(a + b \cdot \text{est} + c \cdot mp + \epsilon)}} \]  

(2.14)

where the parameters \( a, b, c \) can then be estimated using standard logistic regression and \( \epsilon \) is once again a normally distributed error.

For us to be able to test the null hypothesis, we need to use logistic regression to estimate parameters in case that \( b \) is zero. The wanted \( p \)-value for our null hypothesis can then be calculated through a likelihood ratio test as

\[ LR = 2 \cdot (L(H_0) - L(H_1)) \]  

(2.15)

where \( L(H_0) \) is the log likelihood of the model with \( b = 0 \).

LR follows chi-squared distribution with one degree of freedom (difference in the number of the estimated parameters) and if the result is found to be statistically significant, then the market should be inefficient as the model with \( b \neq 0 \) fits the data better. Logit model values, unlike in linear probability
2.10 Bettors on the market

model, can be directly interpreted as true probabilities in all cases, because the values are bounded between 0 and 1, since sigmoid function has limit 1 in \( \infty \) and 0 in \( -\infty \) and is continuous.

An alternative to the logit model is a probit model that uses the cumulative distribution function of the normal distribution, described as \( \phi \) in Equation 2.12 and parameters are estimated using maximum likelihood estimate (same as in logistic regression). After that, the procedure is the same as in the case of the logit model and logistic regression.

The model that we used above is simple and generic, but by adding more parameters we could search for more concrete biases. For example in (Golec and Tamarkin 1991) the authors used a model of:

\[
y = a + b \cdot mp + c \cdot f + d \cdot h + \epsilon
\]  

(2.16)

that uses dummy variables \( h \) to account for the home team advantage and a dummy variable \( f \) that signals the favourite in a match, where again \( mp \) is the market probability estimate, \( a, b, c \) and \( d \) are the estimated parameters, and \( \epsilon \) is a normally distributed error.

2.10 Bettors on the market

While in the previous chapter we introduced possibilities to test the bookmakers, here we propose to test the bettors. For that we take different bettors and their respective wagering histories to test if they are skilled enough to be able to find value bets and gain an edge over the bookmaker, or if their success can be attributed to a mere chance. For this purpose we can again use the t-test, but let us start with some definitions first.

A return over investment (ROI) is defined as:

\[
ROI = (\text{profit} - \text{investment})/\text{investment} + 1
\]  

(2.17)

where investment is the state of bankroll before staking and profit is the amount won. ROI is one of the many possible indicators describing how well a bettor does in terms of making profit. The standard deviation of the bettor’s history can then be estimated as:

\[
\sigma = \sqrt{\text{ROI}(\text{odds}_{\text{average}} - \text{ROI})}
\]  

(2.18)

and then, following (Buchdahl n.d.(a)), the statistics can be calculated as:

\[
t = \frac{\sqrt{n}(ROI - 1)}{\sigma}
\]  

(2.19)

The statistics \( t \) follows student-t distribution with the degrees of freedom equal to sample size -1. Equation 2.19 is then a special form of a classic t-test of the form

\[
t = \frac{\sqrt{n}(x - x_0)}{\sigma}
\]  

(2.20)
2. Market Efficiency

Where $x$ is the mean of population, $x_0$ is the expected mean, and $\sigma$ is the estimate of the standard deviation.

The null hypothesis is whether ROI is equal to 1. In other words, if we cannot reject null hypothesis on our significance level, then the bettor’s ROI is at maximum of 1 and the bettor is not in a profit. After the t-score is calculated using Equation 2.19, its value can be used to calculate a $p$-value as $1 - \text{cdf}(t)$. If the $p$-value is lower than our significance level, then we can reject the null hypothesis and say that it is not a coincidence that the bettor is winning.

### 2.11 Group test

Another quite intuitive method proposed in (Tiitu et al. 2016) is dividing the odds into $n$ categories with similar odds and trying to stake one unit bet per each category while observing return $R_i$. For the null hypothesis, we have:

$$H_0 : R_i \leq 0 \quad (2.21)$$

For testing, we use once again the t-test to determine the significance level of each group. However, this time we only compare to a single inverse distribution function value to determine our confidence level. This procedure could potentially again reveal the favourite-long shot bias – the most common bias proven in all kinds of works and papers.

### 2.12 Probabilities method

Lastly, we propose a new sophisticated method to test the efficiency jointly over several markets. In the method, we exploit the fact that we have more types of odds available on each match. Particularly, we use the both-to-score, over-under, and 1x2 odds to find if the probabilities implied by these odds make sense altogether, i.e. make up for an actual distribution\[^1\].

#### 2.12.1 Table

For each match, we create an $n \times n$ table where $n$ is our estimate of maximum points achievable in the given sport by one side. For example in football, we consider $n = 9$ as a reasonable maximum of goals scored by one side. Each element in the table then represents one possible outcome: rows represent the number of goals scored by an away team and columns the number of goals for a home team.

\[^1\]In case the probabilities do not make sense, there is a chance that they can be exploited to make profit, however in this work we focus merely on finding these situations.
2.12.2 Probabilities of Elements

After our table is created, we shall start forming our soon to be the system of equations in the following way: we take all odds implied probabilities for which it applies that they can be directly divided into individual outcomes represented by the elements in the table. This division splits our table into varying number of parts, depending on the type of odds. We then create an equation for each part of the table. All the elements that form part of the table sum up to the probability of the respective outcome happening, as derived from the odds given by the bookmaker. An example of how such a divided table can look like for the 1x2 market can be seen in Table 2.1.

### 1x2 table

To form the equations for 1x2 odds, we firstly remove margin from the odds and then get implied probabilities. Now on the diagonal of the table we created before, we have all the considered draw outcomes, and we know that they must add to the probability of a draw we derived from the odds. The next equation is that the upper triangle elements sum up to the home team win probability. Lastly, all the lower triangle elements sum up to the away team winning probability.

<table>
<thead>
<tr>
<th></th>
<th>Home</th>
<th>Draw</th>
<th>Home</th>
<th>Home</th>
<th>Home</th>
</tr>
</thead>
<tbody>
<tr>
<td>Away</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Away</td>
<td>Draw</td>
<td>Home</td>
<td>Home</td>
<td>Home</td>
<td>Home</td>
</tr>
<tr>
<td>Away</td>
<td>Away</td>
<td>Draw</td>
<td>Home</td>
<td>Home</td>
<td>Home</td>
</tr>
<tr>
<td>Away</td>
<td>Away</td>
<td>Away</td>
<td>Away</td>
<td>Draw</td>
<td>Home</td>
</tr>
<tr>
<td>Away</td>
<td>Away</td>
<td>Away</td>
<td>Away</td>
<td>Away</td>
<td>Draw</td>
</tr>
</tbody>
</table>

**Table 2.1:** Indicates to which result the box belongs

### Both to score table

The first equation here comes from the sum of elements of the first row and column, and the other equation comes from all the boxes that do not form the first equation.

<table>
<thead>
<tr>
<th></th>
<th>Home</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Away</td>
<td>Neither scored</td>
<td>only Home</td>
<td>only Home</td>
<td>only Home</td>
<td>only Home</td>
<td>only Home</td>
<td>only Home</td>
<td>only Home</td>
</tr>
<tr>
<td></td>
<td>only Away</td>
<td>only Home</td>
<td>only Home</td>
<td>only Home</td>
<td>only Home</td>
<td>only Home</td>
<td>only Home</td>
<td>only Home</td>
</tr>
<tr>
<td></td>
<td>only Away</td>
<td>only Home</td>
<td>only Home</td>
<td>only Home</td>
<td>only Home</td>
<td>only Home</td>
<td>only Home</td>
<td>only Home</td>
</tr>
<tr>
<td></td>
<td>only Away</td>
<td>only Home</td>
<td>only Home</td>
<td>only Home</td>
<td>only Home</td>
<td>only Home</td>
<td>only Home</td>
<td>only Home</td>
</tr>
<tr>
<td></td>
<td>only Away</td>
<td>only Home</td>
<td>only Home</td>
<td>only Home</td>
<td>only Home</td>
<td>only Home</td>
<td>only Home</td>
<td>only Home</td>
</tr>
<tr>
<td></td>
<td>only Away</td>
<td>only Home</td>
<td>only Home</td>
<td>only Home</td>
<td>only Home</td>
<td>only Home</td>
<td>only Home</td>
<td>only Home</td>
</tr>
</tbody>
</table>

**Table 2.2:** Indicates to which result the box belongs
2. Market Efficiency

### Over-under table

Both 1x2 odds and both-to-score odds form systems of equations with the same left side, i.e. where only the right sides change. In the over-under system type, it is different as it also consists of another numeric parameter which tells us the over/under limit on the number of cumulative points that should be scored in a match. The “under part” of the table then contains all the elements in which the sum of the elements at indices offseted by one is smaller than the over/under limit given with the odds. The “over part” is them formed by all the remaining elements. An example for an over-under limit of 2.5 can be seen in Table 2.3.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Home</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Under</td>
<td>Over</td>
<td>Over</td>
<td>Over</td>
<td>Over</td>
<td>Under</td>
<td>Over</td>
<td>Over</td>
<td>Over</td>
<td>Over</td>
<td>Under</td>
<td>Over</td>
<td>Over</td>
<td>Over</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Over</td>
<td>Over</td>
<td>Over</td>
<td>Over</td>
<td>Over</td>
<td>Under</td>
<td>Over</td>
<td>Over</td>
<td>Over</td>
<td>Over</td>
<td>Under</td>
<td>Over</td>
<td>Over</td>
<td>Over</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2.3: Indicates to which result the box belongs for over-under limit of 2.5

### 2.12.3 Asian handicap point five tables

In general, asian handicap odds cannot be used for this method as they introduce the stake refund possibility. However, the “point five” handicaps are an exception as there is always a winning side after adding/removing a half of a point. Consequently, all the indices that represent outcomes in which the home team won, after the handicap is considered, form the home-winning equation. In other words, if we transform the considered handicap to affect home team and round it, then to home equation box $A_{ij}$, contributes if and only if $i + \text{handicap} \geq j$, where $i$ is number of points home team achieved and $j$ is the number of points away team achieved. Furthermore, as usual, the rest goes to form the away team equation.

<table>
<thead>
<tr>
<th></th>
<th>Home</th>
</tr>
</thead>
<tbody>
<tr>
<td>Away</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Away</td>
</tr>
<tr>
<td></td>
<td>Away</td>
</tr>
<tr>
<td></td>
<td>Away</td>
</tr>
<tr>
<td></td>
<td>Away</td>
</tr>
<tr>
<td></td>
<td>Away</td>
</tr>
</tbody>
</table>

Table 2.4: Indicates to which result the box belongs for -1.5 home team handicap
2.12.4 System of inequalities

Once we transformed the odds to the system of equations, we represent it by a matrix $A^p_n$, where $p$ is the number of equations and $n$ is number of variables (probabilities). However, for our problem to make sense, we need to add additional constraints. Each of our variable must lie between 0 and 1, since it is a probability. In the end, we need to find a solution to the system of equations in a unit hyper-cube of a dimension equal to the number of probabilities, i.e.

$$Ax = b, \ A \in \{0,1\}^{p \times v}, \ b \in [0,1]^v$$

$$x_i < 1 \ i = 1..v$$

$$x_i > 0 \ i = 1..v$$

(2.22)

2.12.5 Finding a solution

To find a solution we can use the fact that our system of inequalities is not large (about 15 equations and 100 constraints for all variables), which allows us to define and solve our problem in a more general terminology of linear programming without worrying about computational complexity. Linear programming is an optimization method of minimizing a linear function constrained with linear inequalities. Our problem can be transformed to a linear program in the following form:

$$\min c^T x$$

subject to

$$Ax = b$$

$$0 < x_i < 1 \ i = 1..n$$

(2.23)

where $c$ is an arbitrary non-zero vector\footnote{having the probabilities sum up to one is one of our equations.}. However, for our purposes, we only need to know whether the program has a feasible solution or not. If the program turns out to have some solution, then variables in $x$ give us the potential probabilities, and the bookmaker’s odds do make sense, as there is a particular probability distribution over the match outcomes. However, if the program has no solution, then bookmaker probabilities do not make sense as there is no legal assignment of probabilities to our table. The non-existent solution should then imply some error on the side of bookmaker that might be exploitable, depending on the margin.

2.13 Luck vs skill

Even a bettor that estimates the probability values wrong can still make profit if he is lucky enough. However, with a larger number of bets this situation
becomes less and less likely. The same reasoning works on the other side as well, as if some bettors can make a profit it could be just a bad luck for the bookmaker, but with a higher number of bettors this explanation gets more and more unlikely. However, betting sample sizes should be considerably large, as we would like to eliminate luck out of the equation.

### 2.13.1 Importance of binomial distribution in betting

The binomial distribution is a good choice for us to give an example of how the sample size matters. It is described by two parameters \( n \) and \( p \). In our case, the distribution represents \( n \) successful bets with probability \( p \) as

\[
\text{Mean} = n \cdot p
\]  
\[
\sigma = \sqrt{n \cdot p(1 - p)}
\]  
\[
\sigma_p = \frac{\sqrt{p(1 - p)}}{\sqrt{n}}
\]

Equation 2.26 shows how much the probabilities differ from \( p \) in percent. Also, it can be seen that \( \sigma_p \) is decreasing with the square root of the number of trials, which directly shows the relevance of sample size to make a statistically significant conclusion.

![Figure 2.1: Showing decrease of percent standard deviation](image)

Figure 2.1 then shows how much in percent can an \( n \)-trial experiment differ from the expected value for a simple coin toss. For example, \( n = 100 \)
makes up for a standard deviation of 5%, which means that 66% of values lie between 40% and 60% win rate, which is with a little exaggeration all the way from “I give up” to “I am one of the best”.

Joseph Buchdahl uses Binomial distribution in (Buchdahl n.d.(a)) to give a confidence interval of how lucky/unlucky a bettor was to achieve different results from those he claims. Note that this method is only suitable for 50:50 bets, which are common in, e.g., the Asian handicap markets.

\[ \text{Binomial distribution is close to a normal distribution for large } n \text{ according to central limit theorem.} \]
Chapter 3

Data

For our experiments we used multiple types of datasets. We used two distinct sources of data for our analysis of the bookmakers and the bettors, and one dataset from the Rebel-Betting system\(^1\) to also see how well the paid services do.

### 3.1 Bookmaker odds

Our largest dataset were odds from many different bookmakers for millions of different matches collected from 6 sports: football, hockey, basketball, baseball, handball and volleyball. The Table 3.1 below shows which types of odds are available for each sport, respectively.

<table>
<thead>
<tr>
<th></th>
<th>Football</th>
<th>Hockey</th>
<th>Basketball</th>
<th>Baseball</th>
<th>Volleyball</th>
<th>Handball</th>
</tr>
</thead>
<tbody>
<tr>
<td>1x2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Home-away</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Both-to-score</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Over-under</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Asian handicap</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 3.1: Available odds for each sport

### 3.2 Bettors history data

For our analysis, we needed to have a betting history of multiple bettors with sufficient length to make a solid conclusion. As we did not find any dataset directly, we decided to make our own. On the internet, there are many different websites, where the bettors offer tips for advantageous bets to others and as part of the offer they give away their history for the user to confirm the quality of a bettor. These sites are perfect for us to gather all the relevant data for the test. There are dozens of these sites on the internet, however some websites have a possibility for the bettor to delete

\(^1\)https://www.rebelbetting.com/
unsuccessful bets and if we happened to get data from such a site, our results would be biased and basically useless. After we researched many different sources (List B) and considered the other options, we decided that the best site for our purposes is Blogabet\(^2\), one of the most important and viewed site with history of bettors. Blogabet met our condition of inability for the bettors to delete particular bets and at the same time provided a large enough collection to build a reasonable dataset just from the single site.

### 3.3 Blogabet data

Blogabet is a social betting platform where the bettors publish and share their betting histories with the community. While many of the bettors do this freely, some tipsters are also behind a pay wall.

#### 3.3.1 Crawling the Blogabet

The most problematic part about the site is the fact that it is generated by JavaScript, which means that a simple HTTP get request cannot be used because the resulting HTML returns little to no data. Thus we resorted to use the python framework Scrapy\(^3\) together with a JavaScript rendering engine called Splash\(^4\) running in a separate Docker\(^5\) container. However, this method sadly did not give us desired results, so we had to resort to the least effective, but most robust, method using headless chrome browser as a rendering tool together with a Selenium WebDriver\(^6\).

When we had sufficient tools to render the data, we could finally scrape them from the website. The first logical step was to obtain a list of bettor pages on the website to simplify navigating around the site. Luckily for us the list was available, and we simply crawled it with with Selenium, as mentioned above. After we acquired the list we could simply iterate through it and visit the respective bettors pages one by one. The structure of the bettor page on the site was not complicated. Basic structure of top of the page shows Figure 3.1 On the bottom of each page (Figure 3.2), there is a button that we had to scroll to and press for more bets to load, which was also the main reason why we resorted to JavaScript rendering through a headless web-driver. During the process of crawling, we also encountered problems with timeouts, for which a delay between each request had to be properly tuned.

Once the data were crawled we designed a simple relational database to store the data, consisting of two tables. In the first table we stored HTML file from the site for future reference and in the second we stored each match crawled with all the necessary attributes such as odds, result of the bet etc., with a foreign key pointing to the respective bettor and HTML.

\(^2\)https://blogabet.com/
\(^3\)https://scrapy.org/
\(^4\)https://splash.readthedocs.io/en/stable
\(^5\)https://www.docker.com/
\(^6\)https://www.selenium.dev/projects/
3.4 Rebelbetting data

For a duration of one month, we were able to gain free access to a site that specialises in detecting arbitrage and value opportunities, mentioned in Section 1.1.2 as one of the possible ways for a bettor to gain profit. The monthly cost of the service as such is considerable, usually about 100 Euro, so we were excited to see how well this system does.
Chapter 4

Results

4.1 Bookmakers odds

Firstly we take a look at results of all tests that we applied on the odds we had from different bookmakers. For our analysis, the data from the following bookmakers were used:

- bet-at-home
- bwin
- Unibet
- BetVictor
- Tipsport.cz
- Betsafe
- Pinnacle
- Chance.cz
- GoldBet
- 188BET

The set consists of large global bookmakers such as Pinnacle or bwin and also local bookmakers such as Tipsport or Chance. We then tried each test on both opening and closing odds, to get an idea on how well the bookmakers adjust to the new information present in each market.

4.1.1 1x2 Odds

Let us start with the most basic and common odds on the match outcome itself.
4. Results

Simple strategies

Firstly, we tried applying the simple strategies of unit betting. The cumulative profit development, as displayed e.g. in Figure 4.2, was fairly typical across the bookmakers. In Section 1.2.1 about biases, we mentioned the favourite-long shot bias and, indeed, betting on the underdog is less profitable than betting on the favorite of a match.

**Figure 4.1:** Cumulative sum of betting on football 1x2 Pinnacle Odds

![Pinnacle closing odds](a) ![Pinnacle opening odds](b)

**Figure 4.2:** Cumulative sum of betting on football 1x2 Chance.cz Odds

![Chance.cz closing odds](a) ![Chance.cz opening odds](b)

Random bettor

Staking a unit bet on random outcome of the game can give us two important pieces of information. Firstly, the median of the profit is about as large as the margin (it could be also calculated using equation 1.3) of the bookmaker, which agrees with the mechanics of the market odds. Secondly, we can see a systematic change of profit between the closing and opening odds. In Figure 4.3, we see box-plot of two random bettor realizations. In the case of the bookmaker Pinnacle, we notice a margin of about 0.05, where the closing odds profits deviate less from the median and overall we can say, that
closing odds profit values are lower than the opening ones, which agrees with the hypothesis of bettors bringing new information to the market. On the other hand with the Tipsport bookmaker, we see more than 2 times as large margin. A larger margin is not that surprising as Tipsport is considered a soft bookmaker on the global market.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure4.3}
\caption{Difference between Pinnacle random bettor and Tipsport random bettor on 1x2 football odds}
\end{figure}

\section*{K-L Divergence}

When we divide the markets by sport, we can observe that values of K-L divergence for one sport are very balanced across the bookmakers. The value itself seems to be dependant on probability of a draw in a given sport. The highest K-L values were observed in basketball. The ordered list of bookmakers with the corresponding football 1x2 -KL value is listed below:

1. Bet-at-home 0.1022
2. Bwin 0.1015
3. UniBet 0.1004
4. Betsafe 0.0980
5. BetVictor 0.0975
6. 188BET 0.0953
7. Pinnacle 0.0918
8. Tipsport.cz 0.083
9. Chance.cz 0.079
10. GoldBet 0.075
4. Results

- **Over-all**

Using the 1x2 odds, we were not able to make any profit out of the simple strategies we used. Although there seems to be favourite-long-shot bias in the odds given by the bookmakers, as profit from betting max-odds is higher than betting the min-odds strategy. Random strategy showed us that movement of odds between the opening and closing line can be different across the bookmakers (especially soft and sharp bookmakers), and K-L testing showed that probabilities from different bookmakers are in some sense similar.

- **4.1.2 Over-under Odds**

- **Simple strategies**

While simple strategies applied to the 1x2 odds did not bring any surprising facts, the opposite can be said about the over-under odds. Figure 4.4 shows some interesting traits of the basketball over-under Pinnacle market. First of all betting under on all available bets surprisingly seems to result into a positive return. Also, the return is higher on the closing odds, which even more surprising as these odds are generally considered more efficient.

In contrast in Figure 4.5 displaying the Bet-at-home baseball market, we see a different situation as betting all under and all over gives us about same loss. Another interesting property is an occasional huge change between opening and closing odds as showed in Figure 4.6. One possible explanation of this behavior is that the bettors love to bet on over odds with the vision of an interesting match to watch.

![Pinnacle baseball closing odds](a) Pinnacle baseball closing odds
![Pinnacle baseball opening odds](b) Pinnacle baseball opening odds

**Figure 4.4:** Cumulative sum of betting on Over-under Pinnacle Odds

- **Random Bettor**

Figure 4.7 shows us once again difference between bookmaker margins. Movement of the odds is not that visible here as in applying the simple strategies. Pinnacle is known for confidence in their predictive model and offers lowest
4.1. Bookmakers odds

(a) : Bet-at-Home baseball closing odds  (b) : Bet-at-Home baseball opening odds

Figure 4.5: Cumulative sum of betting on over-under Pinnacle Odds

(a) : Pinnacle basketball closing odds  (b) : Pinnacle basketball opening odds

Figure 4.6: Cumulative sum of betting on Over-under Pinnacle Odds

margin (about 0.025 in this case) out of all available bookmakers and mistakes in their model can possibly lead to profit as in Figure 4.4

(a) : Pinnacle baseball random bettor  (b) : Bet-at-Home baseball random bettor

Figure 4.7: Difference between pinnacle random bettor and Bet-at-Home random bettor on over-under baseball odds
4. Results

K-L Divergence

In terms of the K-L values, the situation is basically the same as in the case of 1x2 odds. Each sport has basically some K-L value that does not differ that much between bookmakers. The highest K-L value then occurs for football. On the other side, lowest K-L values belong to basketball and baseball. This seem to imply that lower the points/goals achieved in a game, the higher the K-L value as it easier to set the over/under limit. An ordered list of bookmakers with the corresponding hockey K-L value is below:

1. UniBet 0.1459
2. Bwin 0.1277
3. Betsafe 0.1077
4. Bet-at-home 0.0996
5. BetVictor 0.0525
6. Pinnacle 0.0455
7. 188BET 0.0161

4.1.3 Both-to-score Odds

Simple strategies

In both-to-score market it was consistently the best option to take the maximum of the available odds. Often, the differences from other options are quite small as in Figure 4.8, but for some bookmakers the difference was more significant as in Figure 4.9.

Figure 4.8: Cumulative sum of betting on Both-to-score Bwin Odds

for Tipsport.cz, Chance.cz and GoldBet we did not have enough data
4.1. Bookmakers odds

(a) : BetVictor closing odds
(b) : BetVictor opening odds

Figure 4.9: Cumulative sum of betting on Both-to-score BetVictor Odds

Random Bettor

Margin difference shown in Figure 4.10 is smaller than in previous section. Besides that, Random bettor does not show us anything interesting.

(a) : Betvictor random bettor
(b) : Bwin random bettor

Figure 4.10: Difference between Betvictor random bettor and Bwin random bettor on both-to-score football odds

K-L Divergence

The situation here is the same as in the 1x2 odds and over-under odds. This time however we have only two sport options of football and hockey. The K-L value is really low for football, the lowest over all the markets and sports. This is simply because the implied probabilities are near 50% and that is the worst case scenario for this method, as any distance quickly adds up to high number. So in this case the low K-L value does not imply inefficient market, because the probabilities are naturally even. K-L value for hockey is on the other side higher and the reason for that can be in the imbalanced probabilities for an outcome to occur. Ordered list of bookmakers and corresponding football
both-to-score K-L value is below:

1. 188BET 0.0123
2. UniBet 0.0104
3. Betsafe 0.0099
4. Bet-at-home 0.0093
5. Bwin 0.0091
6. BetVictor 0.0076

### 4.1.4 Home-away Odds

#### Simple Strategies

A typical development of simple strategies for the home-away odds is shown in Figure 4.12 and in Figure 4.11. We see that neither strategy dominates across all the bookmakers. Also, the changes of odds seems to play lesser role on this market.

![Figure 4.11: Cumulative sum of betting on home-away 188BET Odds](image)

(a) : 188BET home-away hockey closing odds
(b) : 188BET home-away hockey opening odds

### 4.1.5 Random Bettor

Figure 4.13 Shows once again low Pinnacle margin in comparison with more typical one.

### 4.1.6 Probability Method

We applied this method on football, mainly because the number of goals scored in a match does not differ that much thus the table can be small. The significant limitation here is data requirements. All odds considered must be taken from one bookmaker and at the same time, since bookmakers tend
4.1. Bookmakers odds

(a) : Pinnacle home-away hockey closing odds

(b) : Pinnacle home-away hockey opening odds

Figure 4.12: Cumulative sum of betting on home-away Pinnacle Odds

(a) : 188BET random bettor

(b) : Pinnacle random bettor

Figure 4.13: Difference between 188BET random bettor and Pinnacle random bettor on home-away hockey odds

to change the odds according to the market. If we took data from different times, then underlying box probabilities would have most likely changed by then, and the linear program would be infeasible as we wanted, however not because of bookmaker mistake as we anticipated, but because of mentioned change of probabilities. For our testing we choose table 10x10 as the most extensive number of scored goals in a match was nine in our data set.

### 1x2 and Both-to-Score Odds

We started simply with the odds that are unique in a given moment as it would make no sense for a bookmaker to offer more 1x2 or both to score odds in one time. Combining the odds, we got five equations every single time. For solving the linear program, we used python scipy framework[^2] that uses the interior point as a default method. In this simple test, bookmakers showed

[^2]: https://www.scipy.org/
4. Results

no inconsistencies in determining probabilities, and to no surprise, we were not able to find any exploitable matches.

**Adding Over-Under Odds**

Subsequently, we added over-under odds to our testing set. A single pair of these odds add only two equations to our system, however, typically a bookmaker offers more over-under odds at the same time and we can form two equations from each. If the match had only one over-under odds, then the linear program had no problem finding a solution for our matches. However, as we added more over-under limit equations, the inconsistencies began to emerge. More often than not linear programs started to be infeasible as in our system of equations rows became linearly dependent. However this alone would not be such a problem for linear program as it could simply remove redundant constrains and move on, nevertheless it often happened that the linear combination of rows that nullified left side of equation did not nullified right side of equations, which implies the constraints to be incompatible for which the solution could not be found.

**4.1.7 Using all the odds available**

If we used all (1x2, both-to-score, over-under and point five handicap) possible probabilities we could for our program, then almost exclusively the program found a solution if there was only one probability added for each of four groups. If point five handicap or over-under contributed by more than that, the program became infeasible, because of the reasons mentioned in Section 4.1.6.

**4.1.8 Explanation**

This method theoretically seemed as a good option to find inefficiencies in the market, but after using it on real data we found big drawbacks. Let us give two examples. In the first example, a bookmaker offers these 1x2 odds: 2.45 for home team win, 3.0 for away team win and 2.95 for a draw. Additionally, he offers odds 2.41 and 1.52 for -0.5 home team handicap. Betting on home team in 1x2 is same as betting on home team -0.5 handicap. If we remove margin from odds 2.45 and 2.41 we get implied probabilities 37.7 % and 38.6 %. Implied probabilities are different for the same outcome just as we expected, because the odds are different. In the second example let us have the same types of odds, however with the following values: 2.45, 3.0, 3.1 for 1x2 and 2.45 and 1.53 for -0.5 handicap. If we once again remove margin from the odds we get probabilities 38.4 % respectively 38.5 %. Here we get different probabilities. The margins are basically same (1.059 and 1.06) so this is not cause of the problem. The problem here is in assumed uniformly distributed margin. In the 1x2 odds it is distributed between 3 possibilities so one outcome has lower part of the margin and implied probability is higher. However choosing different approach for removing margin might help in some
cases, but make situation even worse in different cases. Another approach could be not to remove margin at all. This fails when bookmaker use different size of margin in his odds and that is usually the case.

### 4.2 Blogabet tipsters data

The data set consists of the most followed tips-giving bettors on the website and it is really important to realize that it is therefore strongly influenced by survival bias. The successful bettors stay on the site and keep giving more and more tips, but the less successful ones simply disappear. As we described in Chapter 2.13, a profit is not enough for us to be able to say that bettor is skilled. The bettor could be just 1 out of thousands that bets randomly and still make seemingly consistent profit. Before we get into testing, let us summarize our data set in Table 4.1.

<table>
<thead>
<tr>
<th>Number of tipsters</th>
<th>average odds</th>
<th>number of bets</th>
<th>overall profit</th>
</tr>
</thead>
<tbody>
<tr>
<td>380</td>
<td>2.4</td>
<td>445915</td>
<td>158496.5</td>
</tr>
</tbody>
</table>

**Table 4.1:** Summarizing the dataset

#### 4.2.1 Results of t-test

Intuitively, it would make sense that making the same profit using shorter odds should require more skill than by using longer odds, which are odds for low probability outcomes. A method that accounts for this we defined in Section 2.10. We tested our hypothesis on two levels of significance, 1 % and 5 %, respectively. Table 4.2 below shows results of the test for the 10 most followed bettor on Blogabet.

<table>
<thead>
<tr>
<th>Name</th>
<th>followers</th>
<th>sample size</th>
<th>average odds</th>
<th>t-score</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>dedi22</td>
<td>11900</td>
<td>2026</td>
<td>3.6</td>
<td>11.41</td>
<td>0</td>
</tr>
<tr>
<td>Aussie126</td>
<td>7854</td>
<td>1215</td>
<td>2.1</td>
<td>9.4</td>
<td>0</td>
</tr>
<tr>
<td>cheser</td>
<td>7004</td>
<td>1161</td>
<td>1.9</td>
<td>6.82</td>
<td>10^{-12}</td>
</tr>
<tr>
<td>Safin</td>
<td>5948</td>
<td>20</td>
<td>1.9</td>
<td>0.33</td>
<td>0.36</td>
</tr>
<tr>
<td>Iranian Sports</td>
<td>2596</td>
<td>294</td>
<td>2</td>
<td>11.9</td>
<td>0</td>
</tr>
<tr>
<td>Semas_Biggest_Fan</td>
<td>2581</td>
<td>284</td>
<td>2.5</td>
<td>3.5</td>
<td>10^{-4}</td>
</tr>
<tr>
<td>NINIBET</td>
<td>2303</td>
<td>275</td>
<td>2.1</td>
<td>10.2</td>
<td>0</td>
</tr>
<tr>
<td>BabyJumpHook</td>
<td>2231</td>
<td>866</td>
<td>1.9</td>
<td>8.4</td>
<td>0</td>
</tr>
<tr>
<td>doko</td>
<td>2203</td>
<td>8754</td>
<td>2</td>
<td>9.0</td>
<td>0</td>
</tr>
<tr>
<td>JimBeam</td>
<td>2085</td>
<td>465</td>
<td>1.93</td>
<td>4.1</td>
<td>10^{-5}</td>
</tr>
</tbody>
</table>

**Table 4.2:** T-test result of the most followed tipsters on blogabet

Nine out of ten times we reject our null hypothesis on both 1 % and 5 % significance levels. We can clearly see that the most followed tipsters earned \(^3\)from the ones which we were able to crawl data from.
their followers by more than just a lucky coincidence. The same table as Table 4.2 but without the number of followers, for all of the available tipsters in our data set can be seen in Table A.1. From the p-values in the table it can be seen that in more than half of the cases we reject the null hypothesis, thus more than half of the tipsters seems to be actually skilled, but we must remind, that data are affected by selection bias.

4.3 Results of Rebelbetting

We virtually staked on everything the site had to offer, and during the scope of one month we were able to gain a profit of about 180 virtual units. If these unit were euros, it could seem like profit about 80 euro a month just by thoughtlessly staking everything. However, one must take into consideration that bets are from about 15 different bookmakers and keeping the bankroll and account with each of them is not an easy job and would actually require a decent skill and considerable time.
Chapter 5

Conclusion

The goal of our work was to use available statistical methods to find inefficiencies and biases in sports betting markets. For that purpose we used bookmaker odds and staking histories of successful bettors to decide whether they are really skilled or just lucky.

For testing of the efficiency we used methods proposed in previous work and also came up with our own method for searching joint inefficiencies across the markets. We also considered the odds movement through the opening and closing lines, and generally considered details and data volumes beyond previous works. The results we found then mostly agree with the previous research.

Particularly, we found evidence of a favourite-long-shot bias in 1x2 odds across different bookmakers and sports using simple staking strategies. We demonstrated big differences in margin sizes across the different bookmakers and, surprisingly, we were even able to demonstrate (virtual) profitability by using a simple under-only strategy at the Pinnacle over-under market.

While the predictive markets are ever-changing and the biases that we found in our research might already be fixed in the bookmaker’s odds by now, new inefficiencies might have been introduced in the process, possibly again creating new opportunities for exploitation in this never-ending evolution towards a truly efficient market.

5.1 Future Work

In our work we were not able to utilize all the proposed methods due to lack of time. A logical direction for future work is thus a simple followup to provide a complete set of tests over all the possible combinations of markets, sports and bookmakers. We also encountered problems with our own proposed method, which could be further improved so as to calculate directly with the provided odds, removing the issues with the presumed margin distribution.


Tiitu, Tero et al. (2016). “Abnormal returns in an efficient market? Statistical and economic weak form efficiency of online sports betting in European soccer”. In:


40
## Appendix A

### Longer t-test Table

<table>
<thead>
<tr>
<th>Name</th>
<th>Sample Size</th>
<th>Average Odds</th>
<th>T-score</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>4eventeVB365</td>
<td>120</td>
<td>1.991</td>
<td>2.941</td>
<td>0.002</td>
</tr>
<tr>
<td>@betinvolley</td>
<td>610</td>
<td>2.131</td>
<td>4.963</td>
<td>0.0</td>
</tr>
<tr>
<td>ACE APUESTAS</td>
<td>569</td>
<td>2.213</td>
<td>3.8</td>
<td>0.0</td>
</tr>
<tr>
<td>ATAT</td>
<td>2170</td>
<td>2.45</td>
<td>8.874</td>
<td>0.0</td>
</tr>
<tr>
<td>Abay_Anzali</td>
<td>2230</td>
<td>1.904</td>
<td>0.736</td>
<td>0.231</td>
</tr>
<tr>
<td>AceOn2ndServe</td>
<td>1570</td>
<td>2.349</td>
<td>5.868</td>
<td>0.0</td>
</tr>
<tr>
<td>African Arabic</td>
<td>425</td>
<td>2.286</td>
<td>2.823</td>
<td>0.002</td>
</tr>
<tr>
<td>Albanian Sports</td>
<td>215</td>
<td>1.867</td>
<td>5.287</td>
<td>0.0</td>
</tr>
<tr>
<td>Algorithms</td>
<td>114</td>
<td>2.926</td>
<td>2.327</td>
<td>0.011</td>
</tr>
<tr>
<td>Alias Core</td>
<td>1071</td>
<td>1.967</td>
<td>0.786</td>
<td>0.216</td>
</tr>
<tr>
<td>Alix</td>
<td>424</td>
<td>1.921</td>
<td>3.152</td>
<td>0.001</td>
</tr>
<tr>
<td>Amassin</td>
<td>627</td>
<td>2.019</td>
<td>1.359</td>
<td>0.087</td>
</tr>
<tr>
<td>Amateur-bets</td>
<td>1709</td>
<td>1.846</td>
<td>4.736</td>
<td>0.0</td>
</tr>
<tr>
<td>Andi Rettig</td>
<td>1744</td>
<td>2.412</td>
<td>0.156</td>
<td>0.438</td>
</tr>
<tr>
<td>Arenas Picks</td>
<td>288</td>
<td>1.952</td>
<td>-0.675</td>
<td>0.25</td>
</tr>
<tr>
<td>Argentin Picks</td>
<td>1180</td>
<td>2.148</td>
<td>2.891</td>
<td>0.002</td>
</tr>
<tr>
<td>Atenasball</td>
<td>45</td>
<td>2.459</td>
<td>4.929</td>
<td>0.0</td>
</tr>
<tr>
<td>Aussie1126</td>
<td>1215</td>
<td>2.1</td>
<td>9.433</td>
<td>0.0</td>
</tr>
<tr>
<td>BBbets</td>
<td>1201</td>
<td>2.331</td>
<td>2.252</td>
<td>0.012</td>
</tr>
<tr>
<td>BabyJumpHook</td>
<td>866</td>
<td>1.908</td>
<td>8.454</td>
<td>0.0</td>
</tr>
<tr>
<td>Bario</td>
<td>456</td>
<td>1.857</td>
<td>5.958</td>
<td>0.0</td>
</tr>
<tr>
<td>Basketball Vision</td>
<td>508</td>
<td>2.086</td>
<td>2.219</td>
<td>0.013</td>
</tr>
<tr>
<td>BeDoBet</td>
<td>362</td>
<td>2.668</td>
<td>0.841</td>
<td>0.2</td>
</tr>
<tr>
<td>Begemot</td>
<td>665</td>
<td>2.171</td>
<td>6.315</td>
<td>0.0</td>
</tr>
<tr>
<td>Betsinlocation</td>
<td>54</td>
<td>1.883</td>
<td>-0.241</td>
<td>0.405</td>
</tr>
<tr>
<td>Betting Coach Tips</td>
<td>900</td>
<td>1.913</td>
<td>2.256</td>
<td>0.012</td>
</tr>
<tr>
<td>Betting Department</td>
<td>437</td>
<td>2.3</td>
<td>2.397</td>
<td>0.008</td>
</tr>
<tr>
<td>Betting Project</td>
<td>164</td>
<td>1.972</td>
<td>1.903</td>
<td>0.029</td>
</tr>
<tr>
<td>Billy Walters</td>
<td>1260</td>
<td>2.37</td>
<td>5.997</td>
<td>0.0</td>
</tr>
<tr>
<td>BoomBoomRoom</td>
<td>444</td>
<td>2.207</td>
<td>7.185</td>
<td>0.0</td>
</tr>
<tr>
<td>Brazilian Bets</td>
<td>2513</td>
<td>2.0</td>
<td>6.622</td>
<td>0.0</td>
</tr>
<tr>
<td>Bruno No.1</td>
<td>358</td>
<td>1.886</td>
<td>2.382</td>
<td>0.009</td>
</tr>
<tr>
<td>Butsa</td>
<td>165</td>
<td>2.004</td>
<td>3.162</td>
<td>0.001</td>
</tr>
<tr>
<td>Carlosn245</td>
<td>1929</td>
<td>2.166</td>
<td>6.8</td>
<td>0.0</td>
</tr>
<tr>
<td>Cassandra_BET</td>
<td>2557</td>
<td>2.374</td>
<td>5.942</td>
<td>0.0</td>
</tr>
<tr>
<td>Claudia</td>
<td>385</td>
<td>2.505</td>
<td>3.836</td>
<td>0.0</td>
</tr>
<tr>
<td>Clint</td>
<td>555</td>
<td>2.016</td>
<td>2.738</td>
<td>0.003</td>
</tr>
<tr>
<td>Tipster</td>
<td>T-statistic</td>
<td>P-value</td>
<td>Significance</td>
<td></td>
</tr>
<tr>
<td>-----------------------</td>
<td>-------------</td>
<td>---------</td>
<td>--------------</td>
<td></td>
</tr>
<tr>
<td>CozZmo</td>
<td>1.988</td>
<td>0.001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Curruncho3579</td>
<td>1.831</td>
<td>0.024</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Daniel’s Tipstube</td>
<td>1.829</td>
<td>0.031</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Del_Pi</td>
<td>3.275</td>
<td>0.001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Denarius</td>
<td>3.017</td>
<td>0.001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DobrAzmiava</td>
<td>1.628</td>
<td>0.053</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Doctor4k</td>
<td>0.894</td>
<td>0.186</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dropp System</td>
<td>3.662</td>
<td>0.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DutchBettingTips</td>
<td>0.216</td>
<td>0.424</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EL POLAKO</td>
<td>5.282</td>
<td>0.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Effic</td>
<td>9.618</td>
<td>0.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eishockey-Expert</td>
<td>-2.06</td>
<td>0.02</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EliteBets</td>
<td>3.326</td>
<td>0.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EuroHockeyTips</td>
<td>2.055</td>
<td>0.02</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ExperimenTalus</td>
<td>3.08</td>
<td>0.001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eye of the storm</td>
<td>2.133</td>
<td>0.017</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FR-Courses by J.Vallet</td>
<td>0.97</td>
<td>0.166</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fannullonibet</td>
<td>2.771</td>
<td>0.003</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fernando_Luiz</td>
<td>2.9</td>
<td>0.002</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Florentin_!</td>
<td>3.267</td>
<td>0.001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Forrest Gump</td>
<td>3.164</td>
<td>0.001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Francis Begbie</td>
<td>6.922</td>
<td>0.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FrancisBegbie</td>
<td>11.935</td>
<td>0.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Franck7</td>
<td>7.139</td>
<td>0.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Frappe</td>
<td>-0.156</td>
<td>0.438</td>
<td></td>
<td></td>
</tr>
<tr>
<td>From Denmark with love</td>
<td>2.26</td>
<td>0.012</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fuleco Sports</td>
<td>1.714</td>
<td>0.048</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Futbol</td>
<td>6.32</td>
<td>0.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gambler_07</td>
<td>4.071</td>
<td>0.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Geh</td>
<td>8.09</td>
<td>0.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gekest</td>
<td>5.52</td>
<td>0.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GeniusTips</td>
<td>13.042</td>
<td>0.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GeorgeBell22</td>
<td>6.172</td>
<td>0.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gerald Woods</td>
<td>4.153</td>
<td>0.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>German Cards</td>
<td>2.69</td>
<td>0.004</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GermanRegio</td>
<td>3.244</td>
<td>0.001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Giorgio Tips</td>
<td>6.207</td>
<td>0.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Giorgoulakos</td>
<td>1.376</td>
<td>0.085</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Global Resistance</td>
<td>4.277</td>
<td>0.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Green Bet365</td>
<td>3.565</td>
<td>0.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>H@NDI©@PPE©</td>
<td>3.553</td>
<td>0.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Handball Profit</td>
<td>0.233</td>
<td>0.412</td>
<td></td>
<td></td>
</tr>
<tr>
<td>High ROI Society</td>
<td>3.243</td>
<td>0.001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hugo6666</td>
<td>0.619</td>
<td>0.273</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hungarian football tips</td>
<td>3.729</td>
<td>0.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>I OAP I Soccer</td>
<td>0.952</td>
<td>0.171</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IALYSOS_17</td>
<td>5.948</td>
<td>0.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ignazio</td>
<td>3.489</td>
<td>0.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inadel</td>
<td>1.855</td>
<td>0.032</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Incognito.</td>
<td>5.551</td>
<td>0.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IncognitoWolf</td>
<td>0.567</td>
<td>0.287</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table A.1: T-test Result from the top 90 available tipsters
Appendix B

Other Sources of Tips Considered

- https://tipstrr.com/
- https://smartbettingclub.com/
- https://www.bettingexpert.com/
- https://pyckio.com/
- https://myfairbet.com/
- https://www.actionnetwork.com/
- https://nflpickwatch.com/
- https://www.oddsportal.com/
- https://www.bettinggods.com/
- https://tradematesports.com/
- https://www.betgps.com/
- http://bettingtips.expert/
- https://sporita.com/
- https://www.bethelp.com/
- https://www.sportytrader.com/
- https://bettoriq.com/
- https://bookies.com/