SSE Riga Student Research Papers
2023 : 8 (260)

BETTING MARKET EFFICIENCY AND PROFITABLE STRATEGIES IN ESTONIAN TOP 2 FOOTBALL LEAGUES

Authors: Erik Daniel Gabor
Joonas Paju

ISSN 1691-4643
ISBN 978-9934-623-10-3

May 2023
Riga

Betting Market Efficiency and Profitable Strategies in Estonian Top 2 Football Leagues

Erik Daniel Gabor

and

Joonas Paju

Supervisor: Morten Hansen

JEL codes: C15, C53, C55

May 2023
Riga

COPYRIGHT DECLARATION AND LICENCE

Names of the authors in full: Erik Daniel Gabor, Joonas Paju

Title of the Thesis: Betting Market Efficiency and Profitable Strategies in Estonian Top 2 Football Leagues

We hereby certify that the above-named thesis is entirely the work of the persons named below, and that all materials, sources and data used in the thesis have been duly referenced. This thesis – in its entirety or in any part thereof – has never been submitted to any other degree commission or published.

In accordance with Section 1 of the Copyright Law of Latvia, the persons named below are the authors of this thesis.

Pursuant to Article 40 of the Copyright Law the authors hereby agree and give an explicit licence to SSE Riga to deposit one digital copy of this thesis in the digital catalogue and data base at SSE Riga Library for an unlimited time and without royalty. The licence permits SSE Riga to grant access to the duly deposited thesis to all users of the catalogue and data base without royalty and limitations to downloading, copying and printing of the digital thesis in whole or in part provided we are indicated as the authors of the thesis according to Clause 4 Section 1 Article 14 of Copyright Law. We assert our right to be identified as the authors of this thesis whenever it is reproduced in full or in part.

Signed digitally

_ /signed/___________________ _ /signed/___________________

Erik Daniel Gabor Joonas Paju

Date

13.04.2023

Table of Contents

Abstract

The study examines Estonian top two football leagues', Meistriliiga and Esiliiga, market efficiency by exploiting arbitrage opportunities, bookmaker calibration, behavioral biases such as favorite-longshot and home-away bias during 2013-2022, and a Poisson model during the 2022 season. Frequent and profitable arbitrage opportunities were discovered both in Meistriliiga and Esiliiga in each year of the research period, indicating market inefficiency. Bookmaker calibration showed that Meistriliiga is subject to favorite-longshot bias, thus favorites are underestimated, and longshots overestimated by bookmakers. Favorite-longshot bias simulation indicated that Esiliiga is economically inefficient as it was possible to make abnormal profits while betting on under 10% implied probability events. Home-away bias simulation showed that Esiliiga is subject to reverse home-away bias. The Poisson model combined with Kelly Criterion betting strategy proved to be profitable only for Esiliiga.

1. Introduction

Online sports betting market has grown significantly over time. The global sports betting market size was USD 76.8 billion in 2021, and is expected to grow year over year at an annual rate of 10.2% from 2022 to 2030. Compared to other sports, football betting generates the majority of the revenue, accounting for over 23.0% of the total online sports betting industry. (Grand View Research, 2021)

Estonia passed the lottery act in 1994 and gaming act in 1995, which allowed Estonian citizens to participate in state-owned lotteries and in offline gambling such as casinos. In 2009, Estonian government passed the gambling act, which allowed online gambling in Estonia for the first time in Estonia's history. In 2022, iGaming's (online gambling and betting) estimated gross gaming revenue was estimated to be EUR 170 million in Estonia (Delasport, 2022). According to Kantar Emor (2019), 50% of residents of Estonia aged 15-74 had encountered gambling for money in the past two years, and 70% during their lifetime. Compared to their previous research in 2014, participation in online gambling had risen 2.5 times indicating a rapid level of growth. At the same time, participation in offline gambling had dropped from 48% to 41%. 29% of Estonian residents aged 15-20 had participated in online gambling during 2017-2019. 14% of active gamblers in Estonia are recognized to belong to the risk group of being a problematic gambler, out of which, sports bettors comprise the majority. Estonia's biggest domestic online betting markets are Estonian football leagues Meistriliiga and Esiliiga (Delasport, 2022).

According to Thaler and Ziemba (1988), for measuring market efficiency, sports betting markets are better suited than the stock market. The main benefit of betting markets is that each asset (bet) has a well-defined time period over which its value becomes definite. As opposed to a stock, which has an infinite life and is valued based on the present value of future cash flows and the price another will pay for the security tomorrow. Moreover, Thaler and Ziemba argue that the properties of betting markets are more likely to be efficient than other financial markets.

The first papers including sports betting arbitrage were primarily intended to examine market efficiency (Pope & Peel, 1989), (Dixon & Pope (2004) due to the lack of online betting and the small number of bookmakers, resulting in no arbitrage opportunities. Further in time, (Constantinou & Norman, 2013), (Franck, Verbeek, and Nüesch, 2013) discovered that, due to an increase in the number of bookmakers and a larger divergence in betting odds, arbitrage opportunities were frequently found, yielding substantial profit given that the profit was risk-free. Market efficiency in the context of sports betting can be measured by the likelihood of generating abnormal profits. Pope and Peel (1989), Dixon and Pope (2004) considered the market to be efficient due to the lack of opportunities for abnormal profits, but additional research Constantinou & Norman (2013), Franck et al, (2013) revealed the market to be inefficient by finding arbitrage opportunities generating abnormal profits.

Contribution the authors want to make to the existing literature is the analysis of Estonian football sports betting markets, which are much lesser known than previously analyzed leagues (English Premier League, La Liga etc). Additionally, authors look for disparities in arbitrage opportunities, market efficiency, behavioral biases and application of betting strategies between a larger and smaller league from the same nation, which have not before been analyzed. Firstly, we will analyze the arbitrage opportunities of Estonian top two football leagues and find, which form of efficiency the Estonian betting market is in terms of arbitrage. Secondly, we will test if bookmakers tend to include behavioral biases such as favorite-longshot and home-away bias to their odds in Estonian top two football leagues. Lastly, we will implement a predictive Poisson model developed by Zebari et. al. (2021) combined with Kelly criterion and fixed unit betting strategy in order to see if it is possible for bettors to make abnormal returns in those two football leagues.
Thus, our research investigates the following three questions:

1. **How efficient in terms of availability of arbitrage opportunities are the top two Estonian football leagues' betting markets during 2013-2022?**
2. **Are Estonian top two football leagues subject to favorite-longshot and home-away bias during 2013-2022?**
3. **Can Kelly criterion and fixed unit betting strategies yield a positive return when betting on Estonian top two football leagues?**

The structure of the thesis is as follows. First, we provide context to our work by reviewing previous research in the same field. For the subsequent step, we describe the data and the method used to acquire it. Following that, we provide an in-depth description of the methodology, followed by a description of the results. In the results section we also explain how our findings fit in with existing studies. Finally, we discuss any potential limitations of the work, provide suggestions for further research, and report our conclusions.

2. Review of literature

2.1 Arbitrage as a predictor of market efficiency

Sauer (1998) proposed that for a betting market to be called efficient, it should include all relevant information available in order to eliminate punters exploiting market opportunities and generating profits. According to Kuypers (2000), for a betting market to be weakly efficient, abnormal returns cannot be made by either the punter or the bookmaker using only price information, with abnormal returns defined as the bookmaker's commission. If a bookmaker earned more or a punter gained more than the bookmaker's commission, the market would indicate weak efficiency. For a market to be semi-strongly efficient, betting odds must reflect all publicly available information, and no abnormal profits can be earned using the information.

Research on sports betting arbitrage has been conducted by several authors and has been well defined. Arbitrage in sports betting would imply betting on all potential outcomes of an event and securing a risk-free profit. These possibilities can occur if various bookmakers' odds across the market fluctuate enough that a punter could bet on each of the outcomes across various bookmakers and profit from it. In our thesis using football, we have three different outcomes: home win, draw, and away win.

Among the first to explore arbitrage in sports betting were Pope & Peel (1989), who found arbitrage opportunities offered by four bookmakers in the Association Football betting market in the UK during 1980-1982. Back then bets used to be taxed and only one after-tax arbitrage opportunity was found, concluding an absence of generating abnormal profits indicating the betting market to be efficient. Similarly, extending to Pope & Peel (1989), Dixon & Pope (2004) analyzed fixed-odds for the UK Association Football matches during 1993-1996 offered by three bookmakers. They found no arbitrage opportunities, resulting from odds divergence being lower during 1993-1996 than in previous years.

Forrest, Goddard, and Simmons (2005) found no arbitrage opportunities in English football presented by 5 bookmakers between 1998 and 2003 using a sample of 10,000 matches. They observed that if they bet on each of the outcomes on any of the individual bookmakers, the average bookmaker's margin would be in the 10% to 12% range, but utilizing the best available odds, the margin would be 6.6%. Bookmakers offer implied odds instead of true odds that are valued at greater than 100% for the total of outcomes of a sports event, thus, profiting by the amount of the margin if the bets on an event are distributed according to true probabilities of each outcome, which is called bookmakers' margin.

Similarly, in analyzing the odds of five major online European bookmakers from 2002 to 2004, Vlastakis, Dotsis, and Markellos (2009) discovered 10 arbitrage chances out of 10,374 football matches equaling to 0.096%, indicating the betting market to be inefficient.

Since then, the number of bookmakers has increased as a result of the internet's advent, increase in popularity of online sports betting, and more firms wanting to participate in the market due to the profitable business model, further research has revealed growth in arbitrage opportunities in sports betting markets. Constantinou & Norman (2013) identified frequent arbitrage opportunities by examining 14 European football leagues from 2005 to 2012 while using odds from 25-60 different bookmakers for a single match. They also discovered that the bookmakers' margins were dropping year after year as the number of bookmakers increased, resulting in increased competition in the market. Furthermore, they found proof that betting on lower division matches leads to increased arbitrage profitability due to the much greater divergence in odds between bookmakers in lower divisions. Franck et al. (2013) studied sports betting arbitrage by combining with five major European bookmakers and a betting exchange. Betting exchanges differ from regular betting markets such that punters can both purchase and sell bets (bet against a result to take place). Since there is no counterparty risk, betting exchanges typically charge cheaper commissions. Arbitrage opportunities were found in 19.2% of the five largest European football divisions' matches between 2004 and 2011. Concluding the above, we can derive our first hypothesis.

Hypothesis 1. The Estonian top two football leagues' betting markets show a weak form of efficiency in terms of arbitrage, and, Esiliiga (second league) shows a weaker form of efficiency than Meistriliiga (first league).

2.2 Biases as predictors of market efficiency

Betting markets, like securities markets, rely on public information and a great number of market players. Bookmakers' odds should reflect the probabilities of the outcomes of the events on the highest possible level, while adding a small margin to ensure their part of the profit, resulting in no opportunities for profitable betting. Still, it has been found that bookmakers have their own biases on the odds.

2.2.1 Favorite-longshot bias

According to Cain, et al. (2000), bookmakers' odds for football games played in the UK leagues tend to back favorites over long shots. Indicating that bettors on average prefer to wager on longshots rather than favorites. In other words betting on a team with a lower implied probability even when the odds favor the opposite side. Similarly Graham & Stott (2008) as they built an outcomes-based probit model to forecast football results in the UK and then compare them with the odds offered by William Hill (UK bookmaker), they found the bookmaker to show a bias toward favorites over underdogs. Moreover, they found that William Hill odds for games between powerful and weaker teams consistently differed. When the away team was stronger, the result-based probit model assigned a home team less of a chance (higher odds) than William Hill. Additionally, the authors' probit model projected a larger chance of home victory than William Hill in cases where the home team occurred to be stronger than the away team in terms of league ranking. Moreover, a study by Angelini & De Angelis (2019) examined the effectiveness of online betting markets using a forecast-based methodology over the course of 11 years for 11 top football leagues in Europe discovered that 8 of such leagues had been efficient, while 3 (Italian Serie A, Greek Super League, and Spanish La Liga) showed inefficiencies that suggest potential profits for punters. Thus, there are evidential bases towards favorite-longshot bias, which both punters and bookmakers can profit from. The results also indicated that when using maximum odds, the market looks to produce sizable positive returns for bettors in the Spanish Liga as they used the best odds offered by 41 different bookmakers. Existence of favorite-longshot bias in different European football leagues has also been reviewed by other researchers (Cain, Law & Peel, 2003; Deschamps & Gergaud, 2007; Forrest et al., 2005; Koning & Zijm, 2022; Oikonomidis, Bruce & Johnson, 2015; Reade, Singleton & Vaughan Williams, 2020).

However, what justifies bookmakers' propensity to utilize the favorite-longshot bias? One justification is to guard against fresh information being used by the public and more knowledgeable insiders (Lahvička, 2014). This makes sense if bettors discover information before the bookmakers, such as when a team's lineup turns out to be far weaker than anticipated. In such a case, it would be more beneficial for a bookmaker to slightly increase the odds of the favorite rather than the longshot in order to prevent bettors from making more money from an underdog wager. Conlisk (1993) explains this riskier method of betting on underdogs by pointing out that punters who correctly predict the underdog bets could also boast about it to their friends. Additionally, it is possible to surmise that the majority of gamblers exhibit risk-loving tendencies, which leads them to wager on the outcomes that are less likely to occur. (Constantinou & Fenton, 2013). Moreover, as was already said, when the majority of bettors in a market prefer longshots over favorites, bookmakers are encouraged to use this bias in order to benefit from those bettors.

Among the more recent studies, Angelini, De Angelis, and Singleton (2022) examined exchange odds prior to the match's start to determine whether the betting markets show a weak or semi-strong form and compared it afterwards to the news that the game's first goal had been scored. The results showed a preference for favorites over underdogs in both the live and the pre-match odds. This is the opposite of what was observed in fixed-odds bookmaker markets, where the reasons in favor of the favorite-longshot bias had been well documented.When Elaad, Reade & Singleton (2020) compared bookmakers' expectations to the outcomes of English football competitions since 2010, they discovered that, on average, bookmakers did not exhibit the favorite-longshot bias. Although these impacts were relatively tiny, individual bookmaker-specific exchanges were inefficient because they did not make use of the data in the odds of their rivals. Additionally, there is data that suggests bookmaker margin as well as overall profitability have decreased for bookmakers as a result of increasing competitiveness.

2.2.2 Home-away bias

The structure of home advantage throughout time in various football leagues is a multifaceted effect that does not appear to be explainable by a single factor. The social influence and audience backing of the home team are two of the most prominent factors (Dohmen, 2016; Goumas, 2014; Peeters & van Ours, 2021). The first principal way in which the crowd factor works is that crowds might motivate the home team to play better. Secondly, the referee may unintentionally favor the host team due to the noise made by the fans. Crowds have a tendency to protest loudly and aggressively at referees for making rulings that do not favor their favorite side (Sutter & Kocher, 2004). Travel exhaustion and altitude fluctuations for visiting teams are additional elements that support the competitive advantage of the home team. Rising above sea level, according to Van Damme & Baer (2019) is linked to an increase in home team benefit. The fact that less oxygen is available at higher altitudes may help to explain this. Furthermore, Pollard (2002) demonstrated that a team benefits from continuously playing in a familiar environment. In the first season played at the new stadium after the renovation, home advantage for teams was significantly lower than it was in the previous season's old stadium.

Although there appears to be an advantage for home teams in football it emerges that many bookmakers across different European football leagues have not given enough credit to this circumstance. Constantinou & Fenton (2013) examined the games of 14 European football leagues during a seven-year period in order to analyze the online betting industry for

European football. They proved that there existed a home-away bias, which was often regarded as being just as profound as the favorite-longshot bias. Betting exclusively on home victories throughout the course of the time frame resulted in better overall returns. However, it appears that this phenomenon is not profitable. In other words, wagers on away victories result in a significant cumulative loss, but wagers on home victories result in a relatively smaller cumulative loss. Elaad (2020) found that in English football leagues, the home team victories are over anticipated by bookmakers at the lower football divisions. Meaning that bookmakers tend to overvalue the home team advantage as the division gets lower. In contrast to these findings, Forrest & Simmons (2008) argued that for the top division of Scottish and Spanish football the bookmakers had understated the home advantage. Moreover, it appears that the number of fans backing each side in a match affects the odds, with those who wager on the more well-liked team receiving lower odds.

Hypothesis 2. The Estonian top 2 football leagues' betting markets are subject to a) favorite-longshot bias and b) home-away bias.

2.3 Predictive modeling

Maher (1982) proposed an independent Poisson order to estimate the outcome of a football game, making it one of the earliest football prediction models. There were less occurrences detected than anticipated where either no goals or a significant number of goals were scored, and the variations from such a model were minimal. Later a bivariate Poisson model was employed to significantly enhance the fit for the differences in scores because the Poisson model did not adequately account for them. Goddard & Asimakopoulos (2004) used an ordered probit regression model to predict the outcomes of the English football league. In comparison to other score prediction models, this model's simplicity is an advantage. The importance of the game, concerns with promotion and relegation, the clubs' participation in tournaments, and the distance between the clubs' home cities were all added for the first time as explanatory variables in this publication. In order to determine betting odds, Cortis, Hales, and Bezzina (2013) employed a Monte Carlo simulation approach with an integrated noise parameter. Using this strategy, they were able to earn a profit of almost 12% throughout the 2012 football tournament of the European Nations. Egidi et. al. (2018) developed a novel hierarchical Bayesian Poisson model that incorporates two separate sources of data - historical match outcomes and bookmakers' betting odds. Through the Skellam distribution, they calculated the scoring rates by converting the inverted bookmaker's betting odds into implied probabilities. The approach produced intended profits and demonstrated strong forecast accuracy in the top four European football leagues. Zebari et al. (2021) used a Poisson model to predict the outcomes of football matches during the 2016/2017 season in the Spanish first division. The model is based on creating attacking and defensive strengths of teams in the home and away fields. Due to only requiring historical match scores of football games played during the first part of a football season, their methodology has the advantage of being simple to apply. In their circumstances, the model was very accurate with correctly predicting the result of a football match in 8 out of 10 games. Modern ways of predictive modeling include machine learning (Hubáček, Šourek & Železný, 2019; Knoll, & Stübinger, 2020; Stübinger, Mangold, & Knoll, 2019). Although the machine learning technique has been effective in many situations, it is often far more complex than the methods mentioned above.

2.4 Kelly Criterion

Making informed predictions requires more than simply a predictive model. In order to reduce the risks associated with bets and avoid going ruin, there also needs to be a betting strategy. There have been numerous papers which apply Kelly criterion to various football leagues in order to seek abnormal returns (Andersen et. al., 2020; Boshnakov, Kharrat & McHale, 2017; Giani, 2022; Hassan & Londoño, 2017). A study by Matej et. al. (2021) has demonstrated that Kelly criterion can be viable strategies when betting on basketball, horse racing and football. The study displayed that by using right methods when betting then monetary gain could be achieved even in cases where the bookmaker's model forecasts are better. According to them, Kelly criterion achieved the best results when compared to other betting strategies (e.g. Markowitz's portfolio theory).

Hypothesis 3. Kelly criterion betting strategy yields a positive return when betting on Estonian top 2 football leagues, and the return is greater for Esiliiga.

3. Methodology

3.1 Data description

For the entirety of our analysis, we have extracted the required data from oddsportal.com, which is a site offering archived odds for various sporting leagues and events. Data was collected using various web scraping techniques. The process was time consuming on account of a lot of manual work as well as requiring a thorough examination in order to weed out even the smallest of errors.

We extracted the data of Estonia Meistriliiga and Esiliiga for 10 seasons, from the beginning of 2013 until the end of 2022. Reasoning behind it being that 2013 was the first fully recorded season for Esiliiga. For this period we have extracted 1708 game results for Meistriliiga and 1 646 for Esiliiga. For Meistriliiga the dataset varies from 2 to 27 bookmakers and for Esiliiga from 1 to 25 bookmakers. In the dataset the extracted odds (home win, draw, away win) were converted to average and maximum odds. These two types of odds from Meistriliiga were determined from 97 047 different bookmaker decimal odd observations and for Esiliiga from 63 354 observations. The historical odds used in the dataset are closing odds meaning they are the final value that a bookmaker has assigned before the start of a particular game. In addition to previously mentioned, the dataset consists of home team names, away team names, dates, home team scores, away team scores, final results and number of bookmakers.

3.2 Arbitrage model

The first part of our methodology consists of replicating the research conducted by Vlastakis et al. (2009) owing to the simplicity and straightforwardness of its approach. Bookmakers make money by adding a margin to all odds they offer. On an event with n outcomes, the expected margin for a bookmaker can be expressed as:

$$E(M) = 1 - \sum_{i=1}^{n} P_i * w_i * d_i \tag{1}$$

where the expected margin (M) depends on each outcome's probability (P_i), the weight of bets on each outcome (w_i), and the bookmaker's provided odds (d_i).

For calculating the actual margin that bookmakers apply to their odds, we would need the distribution of bets on each of the outcomes. This is not possible due to bookmakers not publishing the weights of bets placed. Further, we will calculate the bookmakers' implied margin, which implies that bets are evenly dispersed across all outcomes and that odds represent true probability of the outcomes. Implied margin can be expressed as:

$$E(M') = \left(\sum_{i=1}^{n} P_i'\right) - 1 = \left(\sum_{i=1}^{n} \frac{1}{d_i}\right) - 1 \tag{2}$$

where the implied margin is denoted as (M'), implied probability as (P_i'), and provided odds as (d_i). This equation shows that if bookmakers' provided odds are proportional to their true probability, they would not make any money in theory. The odds provided by bookmakers are smaller than their true probability to create a positive margin for themselves.

We consider a combination of different bookmaker odds to generate an arbitrage opportunity, being aware that betting on every outcome with a single bookmaker would result in a loss due to the margin of the bookmaker. We construct a combined bet, which includes the best odds available from bookmakers. In the event of an arbitrage opportunity, the margin of the total bet (M) is then negative

$$\widetilde{M} = \left[\left(\sum_{i=1}^{n} \frac{1}{\max d_{ij}}\right) - 1\right] < 0 \tag{3}$$

where (d_{ij}) represents the odd on match outcome (i) provided by bookmaker (j) and the highest odd on outcome (i) provided by all bookmakers on the market as (max d_{ij}). In case the equation above holds, we can exploit an arbitrage opportunity, which excessive returns will equal to the negative margin (M). The entire amount wagered must be weighed between each outcome type to form an arbitrage bet. Each bet will follow the weight ($_i$) of the total amount wagered

$$\widetilde{w}_i = \frac{P_i'}{\sum_{i=1}^{n} P_i'} \tag{4}$$

For demonstration, we have depicted two scenarios below where one includes an arbitrage opportunity, and the other does not. For scenario 1, bookmakers' maximum odds gives us an implied probability of 103%, which is obtained from summing up implied probabilities of each outcome of a football game, which is equal to the sum of the inverse odds and calculated in the following way: 1 / 2.5 + 1 / 3.5 + 1 / 2.9 = 1.03. This would mean that the probability of home win, draw and away win is predicted to be 103%, which implies that the bookmakers are overvaluing their probabilities and odds, which is also the bookmakers' margin. For an arbitrage opportunity to occur, the total implied probability of outcomes has to be under 100%, indicating undervaluing of odds. Weights of the bets are obtained in the following way: weight of home win equals to the proportion of home win's implied probability of the total implied probability in the following way: 1 / 2.5 / 1.03 = 0.39 or 39%. In these scenarios we have chosen the total bet size to be 1€, which implies that bet weights equal to bet amounts. The return is calculated by summing the bet amounts and dividing by the bookmakers implied probability: (0.39 + 0.28 + 0.33) / 1.03 = 0.97. For profit we subtract the total bet amount from the return: 0.97 - 1 = -0.03. In the 2nd scenario the sum of implied probabilities of each outcome equals to 0.87. If bets are distributed correctly as shown above previously, the punter will earn a risk-free profit of 15.1% from their total bet size.

	Home Win	Draw	Away Win	Implied Probability	Weight Home	Weight Draw	Weight Away	Return	Profit	Profit Margin
Scenario 1	2.5	3.5	2.9	1.03	0.39	0.28	0.33	0.97	-0.030	-3.0%
Scenario 2	2.9	4.2	3.5	0.87	0.40	0.27	0.33	1.15	0.151	15.1%

Table 1. Examples of arbitrage opportunity calculations. Created by the authors.

Next, we will gather all arbitrage opportunities to see whether there were any and conclude if the betting markets of top two Estonian football leagues are efficient or not. If we identify arbitrage opportunities, we will analyze them further covering the likelihood of the opportunities occurring, profitability, and comparing the results between two leagues.

3.3 Biases model

Favorite-longshot bias refers to bookmakers offering better odds for favorites than longshots, which implies that for favorites, the actual outcome is higher than bookmaker's implied probability, and for longshots, actual outcome is lower than bookmaker's implied probability. We want to see whether bookmakers' models undervalue favorites and overvalue longshots and for that, firstly, we will calibrate bookmakers' implied probabilities to a total implied probability of 100%, excluding bookmakers' added margin, by using a calibration method used by Brycki (2009). Bookmakers' average closing odds are used to calculate implied probabilities because they take into account the characteristics of the entire betting market as a whole rather than a specific bookmaker, which allows us to analyze variance in bookmakers' implied probabilities to actual outcome probabilities and determine whether favorite-longshot bias exists.

The following formula will be used to determine the implied probability of an outcome by bookmakers.

$$IP = \frac{\frac{1}{avg_i}}{\frac{1}{avg_h}+\frac{1}{avg_d}+\frac{1}{avg_a}} \tag{5}$$

where *IP*is the implied probability of an outcome *i, avgh, avgd, and avga* are the average closing odds provided by the bookmakers: home win, draw, and away win respectively.

Firstly, we create decile ranges and group bookmakers' implied probabilities of outcomes to the deciles during seasons 2013-2022. Secondly, the bookmakers' implied probabilities are compared with the actual outcomes' occurrences. Thirdly, the conclusion whether favorite-longshot bias exists is made. Fourthly, we compare the results and variance of Meistriliiga and Esiliiga. Lastly, we look for favorite-longshot bias during seasons 2021 and 2022, to see whether we find differences between a 10 season and 2 season time period. Implied probability predicted by the bookmaker would not systematically deviate from outcome probabilities in case of a weak form efficient betting market (Schervish, 1989).

Another way of determining whether favorite-longshot bias exists is running a simulation, using the lower (<10%, 10-20%, 20-30%) and upper bound (70-80%, 80-90%, 90-100%) bookmaker's calibrated implied probability deciles where on each outcome, a stake of 1 unit will be bet on. For favorite-longshot bias to occur, betting on favorites (70-100% decile ranges) must result in a better return than betting on longshots (0-30%), which would imply that favorites are undervalued and longshots overvalued (Cain et. al 2000). Results between Meistriliiga and Esiliiga will be compared and also, we will compare the results of seasons 2021 and 2022 to see whether there are differences, compared to a 10-season period.

For determining whether home-away bias exists, we will run a simulation betting on each home win, draw and away win a 1 unit stake during 2013-2022 seasons and compare the results of the simulations between Meistriliiga and Esiliiga. Average closing odds of bookmakers will be used similarly to favorite-longshot bias. Home-away bias would occur when betting on home wins cumulatively returns more than betting on away wins (Constantinou and Fenton, 2013). Also, we will compare 2013-2022 seasons' results to seasons 2021-2022, to see whether there are any differences.

3.4 Prediction model

We have decided to implement Zebari et. al. (2021) Poisson model for a couple of reasons. First, our goal of the thesis is not to develop a new model, rather try and implement an already made model for Estonian top two football leagues in order to see if it is possible to achieve abnormal returns as well as compare the efficiencies of the two leagues. Secondly, as there is no accurate historical in-game statistics (e.g. possession, shots on target, pass accuracy) available for those two football leagues, the aforementioned model is appropriate since it can forecast the winner, draw, and loser of football matches using just the teams' previous game results. We extend Zebari et al. (2021) work by using a larger test set, comparing the model's accuracy across two distinct leagues, and combining the model's predictions with average bookmaker odds to determine the model's profitability in the top two Estonian football leagues.

A season in both Esiliiga and Meistriliiga consists of 36 rounds or game weeks where each team plays one football match. There are a total of 10 teams in both of these leagues. First step when implementing the above-mentioned model would be to define the training and testing sets for the model. We have chosen to use the first half of 2022 season (first 18 game weeks) as an initial training set which will be used to predict the outcomes of football games for the following game week (game week 19). When predicting each next game week, the previous game week's data will be added to the training set. For both leagues, we intend to forecast the results of 50 football matches spanned between game weeks 19 and 28.

Training set is used to create each football team's attacking and defending strengths on both home and away fields. For home team's attacking strength, the following formula will be used:

$$Home\ att_a = \frac{Average\ goals_a}{Home\ avg\ goals} \tag{6}$$

where $Home\ att_a$ is the attacking strength of home team a, $Average\ goals_a$ is the average number of goals scored by that same home team during the season so far (excluding future matches), and $Home\ avg\ goals$ is the average number of goals that all the home teams have scored so far during the season (excluding future matches).

For away team's defensive strength, the following formula will be used:

$$Away\ def_b = \frac{Average\ goals\ against_b}{Away\ avg\ goals\ against} \tag{7}$$

where $Away\ def_b$ indicates the defensive strength of away team b, $Average\ goals\ against_b$ is the average number of goals that same away team has conceded during the season so far (excluding future matches), $Away\ avg\ goals\ against$ is the average number of goals away teams in the league have conceded during the season so far (excluding future matches).

After having calculated the attacking and defensive strengths of all teams in the training set, the next step would be to calculate expected home and away goals for each football match. The following formula will be used to find out the Home team's expected goals for a single match:

$$Goal\ expectancy\ home_a = Home\ att_a * Away\ def_b * Home\ avg\ goals \tag{8}$$

where $Goal\ expectancy\ home_a$ is the number of goals the home team a is expected to score during the match against away team b, $Home\ att_a$ is the attacking strength of home team a, $Away\ def_b$ is the defensive strength of away team b, $Home\ avg\ goals$ is the average number of goals that all the home teams have scored so far during the season (excluding future matches).

For away team's expected goals, the following formula will be used:

$$Goal\ expectancy\ away_b = Away\ att_b * Home\ def_a * Away\ avg\ goals \tag{9}$$

where $Goal\ expectancy\ away_b$ is the number of goals the away team b is expected to score during the match against home team a, $Away\ att_b$ is the attacking strength of away team b, $Home\ def_a$ is the defensive strength of home team a, $Away\ avg\ goals$ is the average number of goals that all the away teams have scored so far during the season (excluding future matches).

In order to get probabilities of each possible outcome of a football match, we will first need to use the following Poisson formula to get probabilities of each team scoring up to six goals in the match:

$$P(k\ events\ in\ an\ interval = \ (\lambda^k e^{-\lambda})/k! \tag{10}$$

where P is the probability of a team scoring k amount of goals in a match, k is the number of predicted goals. The formula will be used 7 times for both home and away team, each time the k value changes from 0 to 6. For example, when we put $k = 3$ in the formula, we get the probability of a team scoring 3 goals. e is Euler's number, λ is the goal expectancy for either home or away team which we calculated in the previous step, $k!$ is the factorial of predicted goals. Choosing to get probabilities of up to 6 goals in a match is in accordance with Zebari et. al. (2021) methodology. Reasoning behind it being that teams very rarely score more than 6 goals in a football match.

The final step of the model is to create a goal distribution matrix for each football match. The matrix will contain the probabilities of home and away team scoring up to six goals in the match. "To get every possible score, multiply the probability of every possible score by each team by the probability of every score possible by the other team" (Zebari et. al., 2021). After having created the matrix we will be able to calculate probabilities of each possible outcome of a football match (home win, draw, away win). In order to calculate the probability of a match ending in a draw we need to add up the values of diagonal cells. To get the probability of a match ending in a home win we add up the values under the diagonal cells. Lastly, to get the probability of a match ending in an away win we add up the values which are over the diagonal cells.

3.5 Fixed unit betting vs Kelly criterion

In this section of the thesis, we will examine two betting strategies using the prediction model from the previous part in the intent to produce abnormal returns. From the prediction model we will get necessary probabilities to predict a football match outcome (home win, draw, away win). The first betting strategy will be a simple fixed unit strategy where we bet 1 unit on every football match that our model has predicted. Profit for a single match will be calculated with the following formula:

$$Profit_i = f_i * (b_i - 1) \tag{11}$$

where $Profit_i$ is the profit from an outcome i of a football match, f_i is the size of a bet on a predicted outcome i of a football match which in our case always equals 1 unit, b_i is the average decimal bookmaker odd for a predicted outcome i of a football match.

Our second betting strategy is called Kelly criterion. This strategy differs from the previous simple fixed unit strategy for a couple of reasons. First, the bet size changes with each game and it is calculated with the following formula:

$$f_i = p_i - \left[\frac{(1-p_i)}{(b_i-1)}\right] \tag{12}$$

where f_i is the size of a bet on a predicted outcome i of a football match, p_i is the probability assigned by our model on an outcome i of a football match, and b_i is the average decimal bookmaker odd for a predicted outcome i of a football match (Brycki, J., 2009).

Secondly, with Kelly criterion we do not bet on every football match that our model has predicted, but use the following formula to determine when to make a bet:

$$b_i > \frac{(1-p_i)}{p_i} \tag{13}$$

where b_i is the average decimal bookmaker odd for a predicted outcome i of a football match, and p_i is the probability assigned by our model on an outcome i of a football match. The Kelly betting strategy assumes that if this inequality persists, the bettor will have an edge over the bookmaker and therefore will bet on match outcome i in accordance with the bet size determined by the previous equation (Brycki, J., 2009).

4. Results

4.1 Data overview and the initial comparison between the leagues

The summary statistics shown below (Table 2) already gives a good understanding of the differences between Estonian top two leagues - Meistriliiga and Esiliiga. We can see that all the possible outcome odd type averages are lower for Esiliiga. This could be explained by the indent of bookmakers being more cautious towards the country's secondary league. Also, both the maximum amount and the mean of bookmakers per game is lower for the secondary league, which might indicate that some bookmakers are even scared to enter the Estonian secondary league market. What is in common for the two leagues, is that the home teams tend to score more goals than the away teams. In both cases this is compensated by the bookmakers giving out smaller odds for the home team. This might indicate that the bookmakers have already weighed in the home-away bias into their created odds due to home field advantage. As the mean odds are already quite high for both of the leagues (especially Meistriliiga), this might show us that the arbitrage possibilities may be apparent. Also, as the number of bookmakers per game tends to be higher for the Meistriliiga it may indicate more probable arbitrage opportunities for that particular league. For arbitrage purposes it is also vital to know the bookmaker margins for both of the leagues (Appendix 1). We can see that the average bookmaker margins per 10 year period for the Meistriliiga is at 10.1%, whilst for Esiliiga at 10.9%. For both leagues the highest point of margins were at the 2013 season and as the years went forward the margins started to slightly decrease. This might indicate that as the time goes on, bookmakers get more competitive with each other. This is in accordance with the findings of Grant et. al. (2018) which indicates that bookmaker margins have been continuously declining in time.

There are additional indications of home field advantage when examining the scoring densities for both home and away teams in Esiliiga and Meistriliiga (Appendix 2, Appendix 3). The home team's scoring density is slanted toward more goals in both leagues. The greatest point of density for the home team on the Esiliiga graph is closer to 2 goals per game, while the away club's density is skewed toward 1 goal per game. Although, the graph shows that the home team is more likely than the away team to score more than one goal each game, the highest points of density for the home and away teams in the Meistriliiga are both closer to one goal per game.

	Obs.	Mean	Std. Dev.	Min	Max
Meistriliiga					
Home odds	32349	4.833	5.994704	1.01	126
Draw odds	32349	5.774	2.847175	3.033	18.809
Away odds	32349	6.367	7.880988	1.014	151
Bookmakers	1708	18.94	3.186526	2	27
Home score	1708	1.739	1.739149	0	10
Away score	1708	1.546	1.578903	0	12
Esiliiga					
Home odds	21118	3.116	3.156846	1.01	76
Draw odds	21118	5.038	1.795452	3.183	26
Away odds	21118	4.455	4.759038	1.018	80
Bookmakers	1646	12.83	6.214423	1	25
Home score	1646	2.019	1.763421	0	9
Away score	1646	1.637	1.53905	0	9

Table 2. Summary Statistics Comparison Between the Two Leagues. Odd lines show the summary statistics for each possible game result in 32349 and 21118 observations respectively for Meistriliiga and Esiliiga. Bookmakers, home and away score lines show the summary statistics for 1708 and 1646 Meistriliiga and Esiliiga games respectively. Created by the authors.

4.2 Arbitrage

The arbitrage strategy which we used during the studied period involves finding out the highest possible odds for each outcome from the whole selection of available bookmaker odds. From those maximum odds we calculated the implied probabilities for those odds as well as the total implied probabilities (bookmaker probabilities). When bookmaker probability is lower than 1 then it means that an arbitrage opportunity exists and punters can take advantage of that. From Table 3 we can see that for Meistriliiga there were a total of 263 arbitrage opportunities out of 1708 games, meaning that arbitrage opportunities arose in 15.4% of the games. For Esiliiga the arbitrage opportunities were apparent in 21.39% of the games played. Those percentages are similar when we compare them to the findings from previous academic papers. For example, Franck et al. (2013) found arbitrage opportunities in 19.2% of the five largest European football divisions' matches between 2004 and 2011. Although it is vital to keep in mind that in bigger football leagues tend to be more bookmakers per game.

It is important to note that as we use only the closing bookmaker odds, the actual amount of arbitrage possibilities can be much higher. The results are even more surprising when looking at the first three-year results for the Esiliiga, because the combined number of arbitrage opportunities from these years was only 16. For Meistriliiga the lowest arbitrage opportunity years were 2019 and 2020 with 11 and 19 opportunities respectively. Best years for implementing arbitrage strategies would have been between 2015 and 2017 for Meistriliiga and for Esiliiga between 2015 and 2018. When looking at the table from the Meistriliiga standpoint there does not seem to be any trends which could indicate that as the years go by the arbitrage opportunities rise or fall. For Esiliiga however, it seems that as the time went on more opportunities became apparent. This can be explained by the average number of bookmakers per game. As we can see for the first years in the period for Esiliiga there was about twice as less number of different bookmakers per game than in later years. It is logical that as the number of bookmakers increases, so does the number of arbitrage opportunities, due to having a larger variety of odds to choose from.

					Meistriliiga						
Season	**Total**	**2022**	**2021**	**2020**	**2019**	**2018**	**2017**	**2016**	**2015**	**2014**	**2013**
Avg. nr. of bookmakers	**18.94**	24.1	22.6	19.6	18.6	19.5	18.3	19.2	17.8	15.3	15.5
No. of games in the sample	**1708**	182	137	148	184	179	183	183	180	182	183
No. of arbitrage opportunities	**263**	30	26	19	11	20	38	27	34	26	32
% of arbitrage opportunities	**15.4%**	16.5%	19.0%	12.8%	6.0%	11.2%	20.8%	14.8%	18.9%	14.3%	17.5%
					Esiliiga						
Avg. nr. of bookmakers	**12.83**	21	20.1	16.2	14.7	15.9	13.9	9.8	7.96	3.84	4.84
No. of games in the sample	**1646**	179	133	156	180	172	176	183	181	166	172
No. of arbitrage opportunities	**352**	57	48	45	52	72	37	25	11	2	3
% of arbitrage opportunities	**21.4%**	31.8%	36.1%	28.8%	28.9%	41.9%	21.0%	13.7%	6.1%	1.2%	1.7%

Table 3. Comparison of descriptive statistics of arbitrage opportunities in the Estonian top two leagues. Created by the authors.

In order to find the profits which could be acquired from implementing arbitrage strategy in those two leagues, we decided to choose only games where the above-mentioned bookmaker probability was less than 1. The results show that by perfectly weighing a total of 1 euro bet on all three possible outcomes of every football game in the period, the profit which could have been made from Meistriliiga was 7.19 euro (Appendix 4). In other words, when a person would have executed perfect wagers on every single game where the arbitrage opportunity arose, then that person would have 7.19 timed his initial stake. For Esiliiga however, the profit would have been 18.5 times the initial stake. Also, from this table we can see that the average arbitrage profit per game for Meistriliiga would have been 0.027 euros per 1 euro bet and 0.053 euros per 1 euro bet for Esiliiga. The maximum arbitrage profit from a single game in Meistriliiga during the period would have resulted in 0.423 euros of profit and 0.44 euros of profit for Esiliiga. Vlastakis et al. (2009) study noted even higher arbitrage profitability by individual games, stating that they discovered arbitrage chances that produced profits between 12% and 200% between 2002 and 2004. Nonetheless, they only discovered 10 arbitrage opportunities out of 10,374 football games. To summarize, we accept Hypothesis 1 since Esiliiga has larger arbitrage profits and a higher percentage of arbitrage opportunities than Meistriliiga.

4.3 Biases

4.3.1 Favorite-Longshot Bias

This section examines the bookmakers' calibration results during ten seasons from 2013 to 2022 using average closing odds. Appendix 5.1-5.4 sets includes implied probability decile mid-points, number of observations for each decile range, mean implied probability used by bookmakers' models, and actual outcomes as a percentage. Visual representation of the table is provided in Appendix 6.1-6.4.

For Meistriliiga, looking at the lower bound (5%, 15% and 25% decile mid points), the results of bookmakers' calibration show that bookmakers indeed have overestimated longshots by implying a larger probability than the actual outcomes occurred (Appendix 5.1). We can see that for the 5% implied probability decile mid-point, which corresponds to 0-10% decile, the mean implied probability for bookmakers' models was 6.64% on average, but winning outcomes occurred only in 3.97% from 906 instances. Similarly, 15% and 25% implied probability decile mid points that correspond to 10-20% and 20-30% respectively, mean implied probability is also higher than the actual instances although for 10-20% decile, the difference was 45bp and for 20-30% decile, difference was 94bp, which is relatively small. The upper bound of probabilities implied by bookmakers (75%, 85%, 95% decile mid points corresponding to 70-80%, 80-90% and 90-100% respectively) shows that for 70-80% implied probability, bookmakers have overestimated outcome probabilities, but have underestimated 80-90% and 90-100% decile ranges as their outcome probabilities were higher than implied by bookmakers. For 70-80% decile, the difference is 427bp and for 90-100% decile, difference is 582bp. The biggest divergence comes from 60-70% decile range where outcomes occurred in 71.06% of times compared to bookmakers' implied probability of 65.09%, resulting a 597bp difference. For Meistriliiga, during 2013-2022, evidence suggests that favorite-longshot bias exists as longshots are overestimated and favorites tend to be underestimated by the bookmakers on average, which is in line with previous literature by (Cain, et al., 2000; Cain, Law & Peel, 2003; Forrest et al., 2005; Deschamps & Gergaud, 2007; Graham & Stott, 2008; Oikonomidis, Bruce & Johnson, 2015; Reade, Singleton & Vaughan Williams, 2020; Koning & Zijm, 2022).

For Esiliiga, considering the same time period and variables as for Meistriliiga, we see quite the opposite results (Appendix 5.3). During 2013-2022, Esiliiga's lower bound (0-10% and 10-20%) decile outcome probabilities were underestimated by bookmakers as opposed to Meistriliiga. During 350 instances for bookmakers' implied probability of 7.30%, the outcome probability was 8.86%. Similarly, 10-20% decile range had bookmakers' implied probability of 15.93% while outcome probability was 16.48%. For the whole upper bound, we can see that 70-100% decile ranges were strongly overestimated by the bookmakers. 70-80% decile range had an overestimation of 447bp, while 80-90% decile range had 149bp, and 90-100% decile range had an overestimation of 2408bp (Appendix 6.3). We cannot fully account for the 90-100% decile range as it only had 6 instances, making it vulnerable for large fluctuations. The reason behind the decile range only having 6 instances can be described by the league not having many teams that are way better than the worst teams in the league. Also, the league's teams could be more inconsistent resulting the bookmakers not wanting to assign a team such a high probability of winning. Moreover, neither did bookmakers' implied probabilities on longshots fall to 0-10% decile range as frequent than in Meistriliiga. For Esiliiga, 7.1% of the instances had an implied probability between 0-10% while for Meistriliiga, it was 17.7%. From evidence, we can conclude that bookmakers were rather biased to reverse favorite-longshot, underestimating longshots and overestimating favorites, for Esiliiga during 2013-2022.

To see whether the phenomena still exists and is relevant, we used a two-year period consisting of the previous seasons 2021 and 2022 for comparison. For Meistriliiga, we can see that each favorite decile range (70-100%) is underestimated by the bookmakers and each longshot decile range (0-30%) is overestimated during 2021-2022, confirming the favorite-longshot bias (Appendix 5.2 & 6.2). This could be due to Meistriliiga gaining more popularity by punters, thus having larger betting volumes on games and bookmakers adding their bias, which is more common in more popular football leagues similarly to previous literature by (Cain et al., 2000; Koning & Zijm, 2022) who found favorite-longshot bias to exist in English Premier League and La Liga. Another justification on adding the bias to a higher betting volume league is due to an average punter's tendency of betting on longshots as the upside potential is larger (Constantinou & Fenton, 2013).

In the case of Esiliiga, using seasons 2021 and 2022, a definite conclusion cannot be made as for longshots 0-10% and 20-30% decile range is overestimated but 10-20% decile range is underestimated by the bookmakers as seen on (Appendix 5.4 & 6.4). Similarly, for favorites, two decile ranges (10-80% and 90-100%) are underestimated and 80-90% is overestimated by the bookmakers. Decile range 90-100% only consists of four instances, which cannot be used for a significant estimation.

The results of favorite-longshot betting simulation shows whether previously found bias in Meistriliiga can generate abnormal profits in which case, the betting market would be economically inefficient. Additionally, Constantinou & Fenton (2013) used the betting simulation technique on analyzing the favorite-longshot bias. During seasons 2013-2022 for Meistriliiga, abnormal profits could not be generated as betting on both favorite and longshot decile ranges made loss on average (Appendix 7.1) due to bookmakers' added margin, which was 10.1% on average (Appendix 1). Applying Constantinou & Fenton (2013) methodology that favorite-longshot bias exists if betting on favorites generates a greater return, than betting on longshots, we get another confirmation of favorite-longshot bias existing in Meistriliiga. Using the same strategy on Meistriliiga for seasons 2021 and 2022 we obtain the same results as for seasons 2013-2022 except betting implied probability of 90-100% generated a profit. Unfortunately, we cannot conclude economical inefficiency from it due to number of instances only being 12 and 100% of them winning, making it impossible to generate a loss.

In opposite, favorite-longshot simulation of Esiliiga during 2013-2022 generated a 0.15 unit profit on 1 unit bet on average resulting a profit margin of 15.0% while betting on implied probability 0-10%, which previous literature has not yet captured and reflects one of the purposes of this thesis of lesser known markets being more inefficient (Appendix 7.3) Betting on favorites cumulatively generates more loss than betting on longshots indicating an overestimation of favorites and underestimation of longshots by bookmakers, concluding reverse favorite-longshot bias. On the other hand, this phenomenon disappears while looking at seasons 2021 and 2022 (Appendix 7.4). Favorite-longshot bias starts to appear as betting on favorites generates a better return than betting on longshots. Possible reason to explain this could be that Esiliiga has gained more popularity in previous years, betting volumes have increased, and bookmakers have started adding the bias to their model. Another explanation could be that bookmakers have improved their models as they have more data to train it on, predicting outcomes more precisely.

In conclusion, we accept hypothesis 2a, that the Estonian Meistriliiga is subject to favorite-longshot bias during 2013-2022. On the other hand, we reject hypothesis 2a for Estonian Esiliiga during 2013-2022. As the bookmakers' calibration revealed bookmakers' implied probabilities deviate from actual outcome probabilities for both Estonian Meistriliiga and Esiliiga, we conclude a form of statistical inefficiency. During 2013-2022 Estonian Esiliiga showed economical inefficiency as it was possible to generate abnormal profits betting on 0-10% implied probability outcomes.

4.3.2 Home-Away Bias

The results of home-away bias simulation for Meistriliiga during 2013-2022 showed that the bias would occur during the seasons 2013, 2015, 2016, 2021 and 2022 as betting on home win resulted a higher return than betting on away win. For example, in 2013, betting on home win returned 6.56 units while betting on away win returned a loss of 35.58 units. In 2015, betting on home win returned a loss of 4.72 units and betting on away win a loss of 24.34 units, 2016 results were a loss of 6.77 units and a loss of 79.60 units respectively. Cumulatively over the 10-season period, we cannot conclude the presence of home-away bias as home win returned a loss of 290.60 units and betting on away win a loss of 291.53 units, which are interestingly almost the same (Appendix 8.1). During seasons 2021 and 2022, a marginal home-away bias existed as betting on home win returned a loss of 49.1 units and betting on away win a loss of 57.67 units (Appendix 8.2). This is consistent with Constantinou & Fenton (2013) and Forrest & Simmons (2008) who found home-away bias in top division of Spanish, English, Italian and Scottish football leagues.

Esiliiga shows a marginal presence of reverse home-away bias considering 1645 games played during 2013-2022 with cumulative loss on home win of 22.19 units and a loss on away win of 10.46 units (Appendix 8.3). During seasons 2021 and 2022, the results also concluded a reverse home-away bias as betting on home win cumulatively made a loss of 20.62 units and betting on away win resulted a small loss of 0.74 units, which was also found by Elaad (2020) analyzing English lower football divisions as is Esiliiga compared to Meistriliiga in this thesis (Appendix 8.4).

In conclusion, we reject hypothesis 2b for both Estonian Meistriliiga and Esiliiga being subject to home-away bias during 2013-2022. We suggest Estonian Esiliiga to be subject to reverse home-away bias, as betting on away win cumulatively resulted a higher return.

4.4 Prediction model and betting strategies

We used Poisson model in order to assign probabilities to each possible match result (home win, draw, away win) for football matches in Estonian top two football leagues -

Meistriliiga and Esiliiga. For both of these leagues we predicted 50 match outcomes between game weeks (rounds) 19-28 in the 2022 season. We started with creating the initial testing set for game weeks 1-18 in order to predict the outcomes of football matches for the 19^{th} game week. When predicting each next game week, the previous game week's data was added to the training set.

We will first go through an explanatory phase where we describe how the calculations led us to the predicted outcomes of game week 19 and later, we will summarize the results of the entire model from game weeks 19 to 28. We will use the Meistriliiga as an example for these calculations.

First step when building the model was to create home and away team tables which can be seen down below (Table 4 & Table 5). From these tables we can see that the games played between different teams do not exactly match. In a perfect world each team would have played 9 times at the home stadium and 9 times away, but in realistic world some games get cancelled or postponed. We recognize that home teams for Meistriliiga tend to score more goals at home and also concede less goals at home than away on average. The same sight appears when we look at the Esiliiga (Appendix 9 & Appendix 10).

Meistriliiga GW 1-18					
Home team	**Games played**	**Goals**	**Average goals**	**Goals against**	**Average goals against**
Flora	9	27	3.00	4	0.44
Kalju	9	17	1.89	5	0.56
Kuressaare	9	13	1.44	10	1.11
Legion	9	12	1.33	15	1.67
Levadia	9	25	2.78	4	0.44
Narva	8	10	1.25	11	1.38
Paide	9	29	3.22	11	1.22
Parnu JK Vaprus	8	9	1.13	19	2.38
Tallinna Kalev	8	6	0.75	16	2.00
Tammeka	9	9	1.00	17	1.89
Total	**87**	**157**	**1.80**	**112**	**1.29**

Table 4. Meistriliiga home score of each team for game weeks 1-18 in 2022. Created by the authors.

Meistriliiga GW 1-18					
Away team	**Games played**	**Goals**	**Average goals**	**Goals against**	**Average goals against**
Flora	8	21	2.63	27	3.38
Kalju	8	14	1.75	17	2.13
Kuressaare	9	8	0.89	13	1.44
Legion	8	7	0.88	12	1.50
Levadia	9	17	1.89	25	2.78
Narva	8	6	0.75	10	1.25
Paide	9	14	1.56	29	3.22
Parnu JK Vaprus	9	8	0.89	9	1.00
Tallinna Kalev	10	9	0.90	6	0.60
Tammeka	9	8	0.89	9	1.00
Total	**87**	**112**	**1.29**	**157**	**1.80**

Table 5. Meistriliiga away score of each team for game weeks 1-18 in 2022. Created by the authors.

With the average goals scored and conceded from the tables above we can calculate teams' attacking and defensive strengths at home and away fields which is necessary to later predict expected goals in each game. For example, we want to calculate the attacking strength of Kalju at home and Kuressaare's defensive strength at away:

Home att. (Kalju) = 1.89 / 1.80 = 1.05

Away def. (Kuressaare) = 1.44 / 1.8 = 0.8

Attacking strength which is larger than 1 can be perceived as strong because it means that the team scores more goals than the teams in the league on average. Opposingly defensive strength which is larger than 1 can be viewed as weak because it means that the team concedes more goals than the teams in the league on average. Other teams' attacking and defensive strengths for both Esiliiga and Meistriliiga game weeks 1-18 can be found at the end of the thesis (Appendix 11 & Appendix 12).

Next step is to calculate expected goals for each football match in 19th game week of the 2022 season. For example, number of expected goals that Kalju will score against Kuressaare:

Exp. goals (Kalju) = 1.05 * 0.8 * 1.8 = 1.51

Meistriliiga GW 19			
Home team	Goal expectancy	Away team	Goal expectancy
Legion	2.49	Flora	3.40
Kalju	1.51	Kuressaare	0.38
Narva	2.23	Paide	1.66
Levadia	0.92	Tallinna Kalev	0.31
Parnu JK Vaprus	0.62	Tammeka	1.64

Table 6. Meistriliiga goal expectancies of home vs away teams for game week 19 in 2022. Created by the authors

From the Meistriliiga table above we can already get a good view on how these football games should end in terms of goals (same table for Esiliiga could be found in the Appendix 13). Although, it gets more difficult to predict the match outcome when goal expectancies of teams are similar or closer to zero. To get the probabilities of each probable amount of goals scored by home and away teams we have used Poisson formula. After, using the Poisson formula 7 times for both teams in a match we can create the goal distribution matrix.

No. of goals	Kuressaare goals	0	1	2	3	4	5	6
Kalju goals	Probability	0.681	0.261	0.050	0.006	0.001	0.000	0.000
0	0.220	0.150	0.058	0.011	0.001	0.000	0.000	0.000
1	0.333	0.227	0.087	0.017	0.002	0.000	0.000	0.000
2	0.252	0.172	0.066	0.013	0.001	0.000	0.000	0.000
3	0.127	0.087	0.033	0.006	0.001	0.000	0.000	0.000
4	0.048	0.033	0.013	0.002	0.000	0.000	0.000	0.000
5	0.015	0.010	0.004	0.001	0.000	0.000	0.000	0.000
6	0.004	0.002	0.001	0.000	0.000	0.000	0.000	0.000

Table 7. Meistriliiga goal distribution matrix for a matchup between Kalju and Kuressaare from game week 19 in 2022. Created by the authors.

From the matrix above we can see that the most probable score of the game would be 1:0 for Kalju followed by 2:0, 0:0, 3:0 and so on. In light blue colors the probabilities of a team scoring certain amount of goals can be seen. In order to get probability of a match ending in draw all the diagonal or yellow cells should be added up. To get the home win probability all the orange cells should be added up and in order to get probability of away team win all the green cells should be added up. Same matrix for a match from Esiliiga game week 19 can be found in the Appendix 14.

Meistriliiga GW 19						
		Probability				
Home team	Away team	Win	Draw	Loss	Match score	Correct prediction
Legion	Flora	0.29	0.17	0.55	0:1	Yes
Kalju	Kuressaare	0.66	0.25	0.09	3:1	Yes
Narva	Paide	0.51	0.20	0.29	1:1	No
Levadia	Tallinna Kalev	0.50	0.38	0.12	3:1	Yes
Parnu JK Vaprus	Tammeka	0.14	0.24	0.62	1:1	No

Table 8. Meistriliiga Poisson model outcome probabilities for home vs away teams from game week 19 in 2022. Created by the authors.

From the table above we can see that our example match between Kalju and Kuressaare ended up in a 3:1 home win. The model was able to correctly predict the outcomes of three out of five matches in the 19th game week in the Meistriliiga. Both of the wrongly predicted matches ended in a draw. When we compare those results with Esiliiga's game week 19 predictions from Appendix 15, we can see that the model guessed only two out of five games correctly. One of those wrongly predicted games was predicted as a draw which might indicate that our model has difficulties in correctly estimating draw outcomes.

After, having finished with the game week 19 predictions we added historical match data regarding this game week to the already existing training set in order to predict the following round of football matches. We continued this loop until we predicted the outcomes of 50 football matches in both Meistriliiga and Esiliiga. The results of the whole period for the Meistriliiga can be seen in Table 9 below. First, we can see that the model's accuracy in predicting correct match outcomes was 60% on average. This is equal to the accuracy of fixed bet strategy where we simulated betting 1 unit on each of those 50 matches. Moreover, based on our testing period we cannot conclude that the model would have gotten better in time. We also observe that for Meistriliiga the Kelly criterion strategy failed, due to selecting wrong games which to bet on. We used Kelly criterion formula to determine the bet size taking 1 unit as a maximum or as our bankroll. This resulted in much smaller bet sizes when compared to fixed unit strategy, betting 9.98 units as a total for 21 games. In terms of profit we can see that even though by wagering a total of 50 units with fixed unit strategy we made a smaller loss than with Kelly criterion strategy. The average profit margin with the fixed betting strategy for the whole period was -10% and with Kelly -50%. There were a total of four profitable game weeks for fixed unit and one profitable game week for Kelly criterion strategy. From the cumulative profit graph in Appendix 16 we are able to observe that for the fixed unit strategy our account was in profit between game weeks 21 and 23 whereas with Kelly our account balance was never in a positive state. For both strategies the cumulative loss was highest at the end of game week 28.

Meistriliiga											
2022	**GW19**	**GW20**	**GW21**	**GW22**	**GW23**	**GW24**	**GW25**	**GW26**	**GW27**	**GW28**	**Total**
No. of games	5	5	7	4	5	6	5	5	5	3	**50**
No. of Kelly bets	1	0	3	1	2	1	4	4	4	1	**21**
Fixed bet size (per GW)	5	5	7	4	5	6	5	5	5	3	**50**
Kelly bet size (per GW)	1.96	0.00	1.02	0.39	0.12	0.52	1.96	1.69	1.95	0.37	**9.98**
Accuracy (fixed bet)	60%	80%	71%	75%	60%	67%	20%	60%	40%	67%	**60%**
Accuracy (Kelly)	0%	-	33%	0%	50%	0%	25%	50%	25%	0%	**29%**
Profit (fixed bet)	-1.42	0.58	1.69	0.15	-0.70	-1.44	-2.50	0.39	-0.78	-0.76	**-4.79**
Profit (Kelly criterion)	-0.38	0.00	-0.18	-0.39	0.05	-0.52	-1.44	0.00	-1.79	-0.37	**-5.02**
Profit margin (fixed bet)	-28%	12%	24%	4%	-14%	-24%	-50%	8%	-16%	-25%	**-10%**
Profit margin (Kelly criterion)	-19%	-	-18%	-100%	37%	-100%	-74%	0%	-92%	-100%	**-50%**

Table 9. Meistriliiga betting strategies and profitability for the whole test period of game week 19-28 in 2022. Created by the authors.

When examining the results for Esiliiga from Table 10, we see that the 52% prediction accuracy of the model for the whole period is slightly worse when compared to Meistriliiga. We also cannot conclude whether the model got better with time. In total, number of Kelly bets is higher as is the accuracy of the model when compared to the Meistriliiga results from Table 6. There were a total of five profitable game weeks for both fixed unit and Kelly criterion betting strategies for Esiliiga. This time we detect that the Kelly strategy is actually more profitable, being able to result in a profit of 0.47 units compared to a loss of 0.17 units with fixed unit strategy. Thus, we can say that the betting market in Esiliiga appears to be less efficient than the betting market in Meistriliiga based on the profitability of the prediction model. From the Esiliiga cumulative profit graph in Appendix 17 we see that the lowest state of our balance when using Kelly strategy was at the end of game week 20 with a cumulative loss of 2.73 units thus far. For fixed unit strategy the lowest point was game week 23 with a cumulative loss of over 7 units. The highest states of our balance for both strategies was at the end of game week 28.

Esiliiga											
2022	**GW19**	**GW20**	**GW21**	**GW22**	**GW23**	**GW24**	**GW25**	**GW26**	**GW27**	**GW28**	**Total**
No. of games	5	6	5	5	6	5	5	5	5	3	**50**
No. of Kelly bets	3	4	4	5	4	3	3	4	4	1	**35**
Fixed bet size (per GW)	5	6	5	5	6	5	5	5	5	3	**50**
Kelly bet size (per GW)	1.43	2.30	1.82	2.13	1.52	1.31	1.37	1.46	2.07	0.45	**15.86**
Accuracy (fixed bet)	40%	33%	80%	20%	33%	60%	60%	60%	60%	100%	**52%**
Accuracy (Kelly)	0%	0%	100%	20%	25%	33%	33%	50%	50%	100%	**37%**
Profit (fixed bet)	-2.80	-3.34	5.42	-3.65	-2.72	1.36	3.87	-0.60	1.10	1.19	**-0.17**
Profit (Kelly criterion)	-1.43	-2.30	3.02	-1.95	-0.75	0.33	3.10	-0.44	0.53	0.36	**0.47**
Profit margin (fixed bet)	-56%	-56%	108%	-73%	-45%	27%	77%	-12%	22%	40%	**0%**
Profit margin (Kelly criterion)	-100%	-100%	100%	-91%	-49%	25%	226%	-30%	26%	80%	**3%**

Table 10. Esiliiga betting strategies and profitability for the whole test period of game week 19-28 in 2022. Created by the authors.

As Kelly criterion betting strategy yielded a positive return for Esiliiga and not for Meistriliiga we need to reject the first part of our Hypothesis 3. Although, we can confirm the second part of the Hypothesis 3. as the return is greater for Esiliiga.

Zebari et. al. (2021) who's prediction model we used received an 80% match outcome prediction accuracy when using their model compared to our 60% for Meistriliiga and 52% for Esiliiga. Although, it is important to mention that Zebari et. al. (2021) used the model to predict only 10 match outcomes compared to our 50. Moreover, we extended their methodology in terms of combining it with betting strategies in order to see the possible monetary value this model could add when betting in a real-world scenario. Our Kelly criterion betting strategy results are in accordance with Matej et. al. (2021) study which demonstrated that Kelly criterion strategy can be profitable in cases where the bookmakers' model forecasts are better. There are also other academic papers which found the Kelly criterion to be profitable strategy when betting on football (Andersen et. al., 2020; Boshnakov, Kharrat & McHale, 2017; Brycki, J. 2009; Hassan & Londoño, 2017). In addition, Matej et. al. (2021) pointed out that fractional Kelly would in many cases be more appropriate strategy than Kelly, due to the reason that Kelly assigns too large bet sizes compared to one's bankroll. We can also confirm it, with looking at Appendix 17 we see that the lowest state of our balance when using Kelly strategy is minus 2.73 units for Esiliiga. Indicating that we would have needed a bankroll larger than 2.73 units not to go broke.

5. Conclusion

The study evaluates the effectiveness of the betting markets for the top two football leagues in Estonia by assessing arbitrage opportunities as well as behavioral biases which are built into bookmakers' odds. A prediction model and two distinct betting techniques are also tested in the study to determine whether it is feasible to generate abnormal returns.

Arbitrage opportunities were apparent in 15.4% of the games for Meistriliiga and 21.39% for Esiliiga in a 2013-2022 period, showing weak-form market efficiency for both leagues. It is important to note that as we only used the closing bookmaker odds, thus the actual amount of arbitrage possibilities can be much higher. In terms of profitability, when a person would have executed perfect wagers on every single game where the arbitrage opportunity arose, then that person would have 7.19 timed his initial stake for Meistriliiga games. For Esiliiga however, the profit would have been 18.5 times the initial stake.

We found that Meistriliiga was subject to favorite-longshot bias both during 2013-2022 and 2021-2022, while Esiliiga was subject to reverse favorite-longshot bias during 2013-2022. For Esiliiga, betting on events with 0-10% implied probability by bookmakers returned an average profit of 0.15 units per 1-unit stake with a profit margin of 15%, which implies that Esiliiga's betting market is economically inefficient. Bookmakers' implied probabilities deviated from outcome probabilities for both Meistriliiga and Esiliiga, indicating a weak form of statistical inefficiency. For Meistriliiga, we could not conclude home-away bias during 2013-2022 due to divergence in cumulative returns being too marginal. For Esiliiga, we concluded reverse home-away bias as cumulative returns of the simulation for home win resulted in a loss of 22.19 units and for away win, a loss of 0.74 units.

Our findings show that the Poisson model's accuracy in predicting correct match outcomes was 60% on average for Meistriliiga and 52% for Esiliiga. For fixed odds strategy we placed a 1 unit bet on each of the 50 games for both football leagues. This resulted in a -4.79 unit total loss for Meistriliiga and -0.17 unit total loss for Esiliiga. However, for our second approach, the Kelly criterion method, we bet on 21 games for the Esiliiga and 35 games for the Meistriliiga. Due to choosing the incorrect games to wager on, this approach was unsuccessful for Meistriliiga, resulting in a total loss of -5.02 units. Whilst, for Esiliiga Kelly criterion strategy was profitable, resulting in a total profit of 0.47 units.

Further studies on the effectiveness of the betting market can build on the findings of our study. The prediction model might be improved in future study by integrating more internal historical match information (e.g. possession, shots on target, pass accuracy) as well as testing it for a longer time period. To increase returns, one should also modify the Kelly criterion betting strategy's bet sizes according to the bankroll. Additionally, one could include opening odds in order to achieve as many arbitrage opportunities as possible.

Our study has its limits, despite the fact that our technique may be utilized to examine any other football sports betting market. Due to the intricacy of the approach and the accessibility of the data, transaction costs are not taken into account while looking for arbitrage opportunities. Due to the short testing period and general absence of continuously changing variables, the prediction model's results could be inconclusive.

Given that we are aware of our limitations, we firmly feel that the present growth in sports betting and internet gambling makes our study highly relevant. With the examination of the two Estonian football league betting markets, which to our knowledge have not been studied before, our study enhances the amount of known literature. Also, both bettors and bookies may find value in the findings from this research. Bettors can develop a betting strategy to take advantage of irregularities to increase their winnings by being aware that the market is vulnerable to inefficiencies and that bookmakers include behavioral biases in the odds they are offering. On the other hand, bookmakers can change their odds to profit more through biased bettors.

6. Reference list

Andersen, R., Hassel, V., Hvattum, L. M., & Stålhane, M. (2020). IN-GAME BETTING AND THE KELLY CRITERION. *Math. Appl*, *9*, 67-81.

Angelini, G., & De Angelis, L. (2019). Efficiency of online football betting markets. *International Journal of Forecasting*, *35*(2), 712-721.

Angelini, G., De Angelis, L., & Singleton, C. (2022). Informational efficiency and behaviour within in-play prediction markets. *International Journal of Forecasting*, *38*(1), 282-299.

Boshnakov, G., Kharrat, T., & McHale, I. G. (2017). A bivariate Weibull count model for forecasting association football scores. *International Journal of Forecasting*, *33*(2), 458-466.

Brycki, J. (2009). Statistical and Economic Tests of Efficiency in the English Premier League Soccer Betting Market.

Cain, M., Law, D., & Peel, D. (2000). The Favourite-Longshot Bias and Market Efficiency in UK Football betting. Scottish Journal of Political Economy, 47(1), 25–36. doi:10.1111/1467-9485.00151

Cain, M., Law, D., & Peel, D. (2003). The favourite-longshot bias, bookmaker margins and insider trading in a variety of betting markets. *Bulletin of Economic Research*, *55*(3), 263-273.

Conlisk, J. (1993). The utility of gambling. *Journal of risk and uncertainty*, *6*(3), 255-275.

Constantinou, A. C. & Fenton, N. E. (2013). Profiting from Arbitrage and Odds biases of the European Football Gambling Market. *The Journal of Gambling Business and Economics*, 7(2), 41-70.

Cortis, D., Hales, S., & Bezzina, F. (2013). Profiting on inefficiencies in betting derivative markets: The case of UEFA Euro 2012. *The Journal of Gambling Business and Economics*, *7*(1), 39-51.

Delasport. (2022). Estonia iGaming Market Report: A safe bet. Retrieved from: https://www.delasport.com/estonia-igaming-market-report/

Deschamps, B., & Gergaud, O. (2007). Efficiency in betting markets: evidence from English football. *The Journal of Prediction Markets*, *1*(1), 61-73.

Dixon, M. J., & Pope, P. F. (2004). The value of statistical forecasts in the UK association football betting market. International Journal of Forecasting, 20(4), 697–711. doi:10.1016/j.ijforecast.2003.12.007

Dohmen, T., & Sauermann, J. (2016). Referee bias. *Journal of Economic Surveys*, *30*(4), 679-695.

Egidi, L., Pauli, F., & Torelli, N. (2018). Combining historical data and bookmakers' odds in modelling football scores. *Statistical Modelling*, *18*(5-6), 436-459.

Elaad, G. (2020). Home-field advantage and biased prediction markets in English soccer. *Applied Economics Letters*, *27*(14), 1170-1174.

Elaad, G., Reade, J. J., & Singleton, C. (2020). Information, prices and efficiency in an online betting market. *Finance Research Letters*, *35*, 101291.

Forrest, D., Goddard, J., & Simmons, R. (2005). Odds-setters as forecasters: The case of English football. International Journal of Forecasting, 21(3), 551–564. doi:10.1016/j.ijforecast.2005.03.003

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

Franck, E., Verbeek, E., & Nüesch, S. (2013). Inter-market Arbitrage in Betting. Economica, 80(318), 300–325. doi:10.1111/ecca.12009

Giani, B. (2022). Optimal betting with the Kelly criterion: applications to sports betting and stock market.

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

Goumas, C. (2014). Home advantage in Australian soccer. *Journal of Science and Medicine in Sport*, *17*(1), 119-123.

Graham, I., & Stott, H. (2008). Predicting Bookmaker Odds and Efficiency for UK Football. *Applied Economics*, 40, 99-109.

Grand View Research (2021). Sports Betting Market Size, Share & Trends Analysis By Platform, By Type, By Sports Type (Football, Basketball, Baseball, Horse Racing, Cricket, Hockey, Others), By Region, And Segment Forecasts, 2022 - 2030

Grant, A., Oikonomidis, A., Bruce, A. C., & Johnson, J. E. (2018). New entry, strategic diversity and efficiency in soccer betting markets: the creation and suppression of arbitrage opportunities. *The European Journal of Finance*, *24*(18), 1799-1816.

Hassan, A. R., & Londoño, M. G. (2017). Profiting from the English premier league: predictive elicitation, the Kelly criterion, and black swans. *International Journal of Sport Finance*, *12*(4), 386-395.

Hubáček, O., Šourek, G., & Železný, F. (2019). Exploiting sports-betting market using machine learning. *International Journal of Forecasting*, *35*(2), 783-796.

Kantar Emor. (2019) Eesti elanike kokkupuuted hasartmängudega.

Knoll, J., & Stübinger, J. (2020). Machine-learning-based statistical arbitrage football betting. *KI-Künstliche Intelligenz*, *34*(1), 69-80.

Koning, R. H., & Zijm, R. (2022). Betting market efficiency and prediction in binary choice models. *Annals of Operations Research*, 1-14.

Kuypers, T. (2000). Information and efficiency: an empirical study of a fixed odds betting market. Applied Economics, 32(11), 1353–1363. doi:10.1080/00036840050151449

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

Maher, M. J. (1982). Modelling association football scores. *Statistica Neerlandica*, *36*(3), 109-118.

Matej, U., Gustav, Š., Ondřej, H., & Filip, Ž. (2021). Optimal sports betting strategies in practice: an experimental review. IMA Journal of Management Mathematics, 32(4), 465-489.

Oikonomidis, A., Bruce, A.C. & Johnson, J.E.V. (2015) Does Transparency Imply Efficiency? The Case of the European Soccer Betting Markets

Peeters, T., & van Ours, J. C. (2021). Seasonal home advantage in English professional football; 1974–2018. *De Economist*, *169*(1), 107-126.

Pollard, R. (2002). Evidence of a reduced home advantage when a team moves to a new stadium. *Journal of Sports Sciences*, *20*(12), 969-973.

Pope, P. F., & Peel, D. A. (1989). Information, Prices and Efficiency in a Fixed-Odds Betting Market. *Economica*, *56*(223), 323–341. https://doi.org/10.2307/2554281

Reade, J., Singleton, C., & Vaughan Williams, L. (2020). Betting markets for English Premier League results and scorelines: evaluating a forecasting model. *Economic Issues*, *25*(1), 87-106.

Schervish, M. J. (1989): “A General Method for Comparing Probability Assessors,” The Annals of Statistics, 17, 1856–1879.

Stübinger, J., Mangold, B., & Knoll, J. (2019). Machine learning in football betting: Prediction of match results based on player characteristics. *Applied Sciences*, *10*(1), 46.

Sutter, M., & Kocher, M. G. (2004). Favoritism of agents–the case of referees' home bias. *Journal of Economic Psychology*, *25*(4), 461-469.

Thaler, R. H., & Ziemba, W. T. (1988). Anomalies: Parimutuel Betting Markets: Racetracks and Lotteries. Journal of Economic Perspectives, 2(2), 161–174. doi:10.1257/jep.2.2.161

Van Damme, N., & Baert, S. (2019). Home advantage in European international soccer: Which dimension of distance matters?. *Economics*, *13*(1).

Vlastakis, N., Dotsis, G., & Markellos, R. N. (2009). How efficient is the European football betting market? Evidence from arbitrage and trading strategies. Journal of Forecasting, 28(5), 426–444. doi:10.1002/for.1085

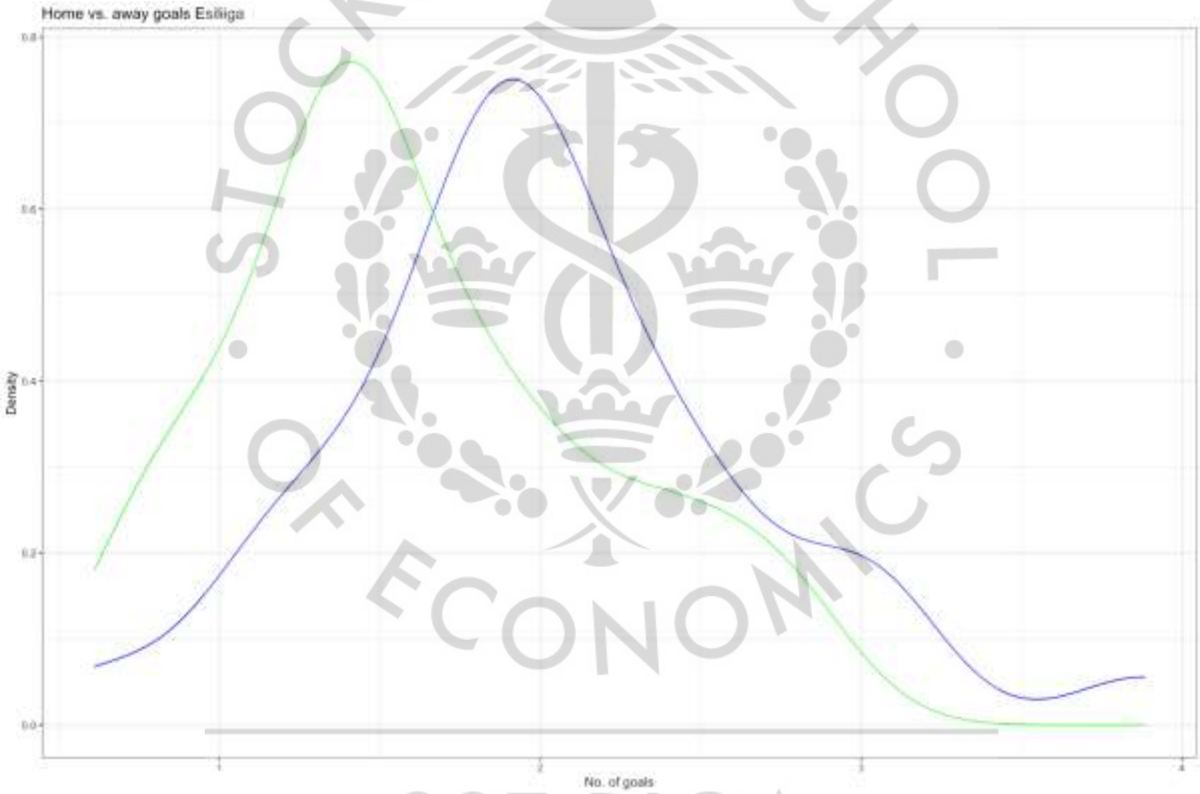
Zebari, G. M., Zeebaree, S. R., Sadeeq, M. A., & Zebari, R. R. (2021). Predicting Football Outcomes by Using Poisson Model: Applied to Spanish Primera División. Journal of Applied Science and Technology Trends, 2(04), 105-112.

7. Appendices

Appendix 1. Average bookmaker margin per season in Estonian top two football leagues.

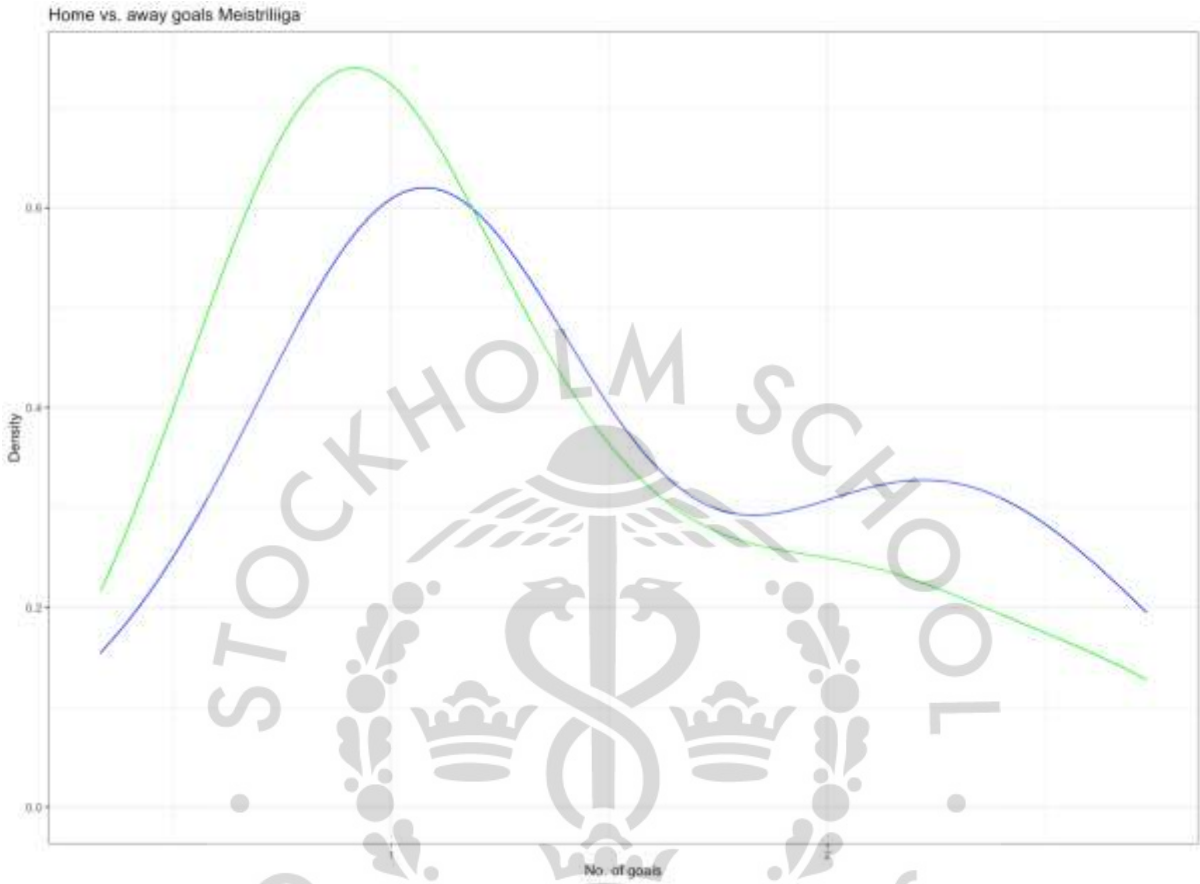
Average Bookmaker Margin per Season											
Season	**Total**	**2022**	**2021**	**2020**	**2019**	**2018**	**2017**	**2016**	**2015**	**2014**	**2013**
Meistriliiga	**10.1%**	9.0%	8.9%	9.8%	9.8%	9.4%	10.3%	10.4%	10.5%	10.4%	10.7%
Esiliiga	**10.9%**	9.2%	9.2%	10.3%	10.7%	10.2%	11.7%	11.7%	11.8%	12.0%	12.0%

Appendix 2. Home vs. away goals in Esiliiga.



Blue line indicates density of home team goals and green line away team goals per game. Created by the authors.

Appendix 3. Home vs. away goals in Meistriliiga.



Blue line indicates density of home team goals and green line away team goals per game. Created by the authors.

Appendix 4. Arbitrage profit comparison for Estonian top two leagues.

Season	Total	2022	2021	2020	2019	2018	2017	2016	2015	2014	2013
Meistriliiga											
Max. arbitrage profit per game (€)	**0.423**	0.310	0.120	0.089	0.094	0.047	0.250	0.140	0.101	0.111	0.423
Avg. arbitrage profit per game (€)	**0.027**	0.036	0.029	0.030	0.028	0.017	0.026	0.029	0.018	0.028	0.033
Cumulative arbitrage profit (€)	**7.19**	1.09	0.762	0.545	0.302	0.331	0.968	0.787	0.612	0.726	1.07
Esiliiga											
Max. arbitrage profit per game (€)	**0.440**	0.382	0.255	0.294	0.196	0.440	0.273	0.250	0.126	0.020	0.155
Avg. arbitrage profit per game (€)	**0.053**	0.064	0.048	0.058	0.047	0.049	0.055	0.050	0.043	0.012	0.061
Arbitrage profit (€)	**18.5**	3.67	2.32	2.59	2.46	3.51	2.05	1.24	0.477	0.025	0.182

Results are shown as a profit per 1-euro wagers. Only games with arbitrage opportunities are chosen and the bet sizes have been perfectly weighed on all three outcomes. Created by the authors.

Appendix 5.1. Bookmakers mean implied probability vs outcome probability in Meistriliiga during 2013-2022.

Meistriliiga 2013-2022

Implied Probability Mid Point	Observations	Mean Implied Probability	Outcome Probability
5%	906	6.64%	3.97%
15%	1136	14.89%	14.44%
25%	1119	24.98%	24.04%
35%	387	34.99%	34.63%
45%	303	44.47%	47.52%
55%	220	55.08%	55.00%
65%	273	65.09%	71.06%
75%	362	75.37%	74.59%
85%	358	84.56%	88.83%
95%	60	90.85%	96.67%

Appendix 5.2. Bookmakers mean implied probability vs outcome probability in Meistriliiga during 2021-2022.

Meistriliiga 2021-2022

Implied Probability Decile Mid Point	Observations	Mean Implied Probability	Outcome Probability
5%	153	6.43%	4.58%
15%	208	15.25%	14.90%
25%	223	24.39%	22.42%
35%	69	35.18%	34.78%
45%	43	44.64%	53.49%
55%	62	55.02%	54.84%
65%	59	65.24%	66.10%
75%	46	75.16%	78.26%
85%	67	84.35%	86.57%
95%	12	90.77%	100.00%

Appendix 5.3. Bookmakers mean implied probability vs outcome probability in Esiliiga during 2013-2022.

Esiliiga 2013-2022

Implied Probability Mid Point	Observations	Mean Implied Probability	Outcome Probability
5%	350	7.30%	8.86%
15%	1165	15.93%	16.48%
25%	1377	23.48%	20.99%
35%	475	34.88%	38.53%
45%	427	44.56%	47.31%
55%	401	54.77%	55.36%
65%	348	64.73%	66.38%
75%	251	74.99%	70.52%
85%	138	84.82%	83.33%
95%	6	90.75%	66.67%

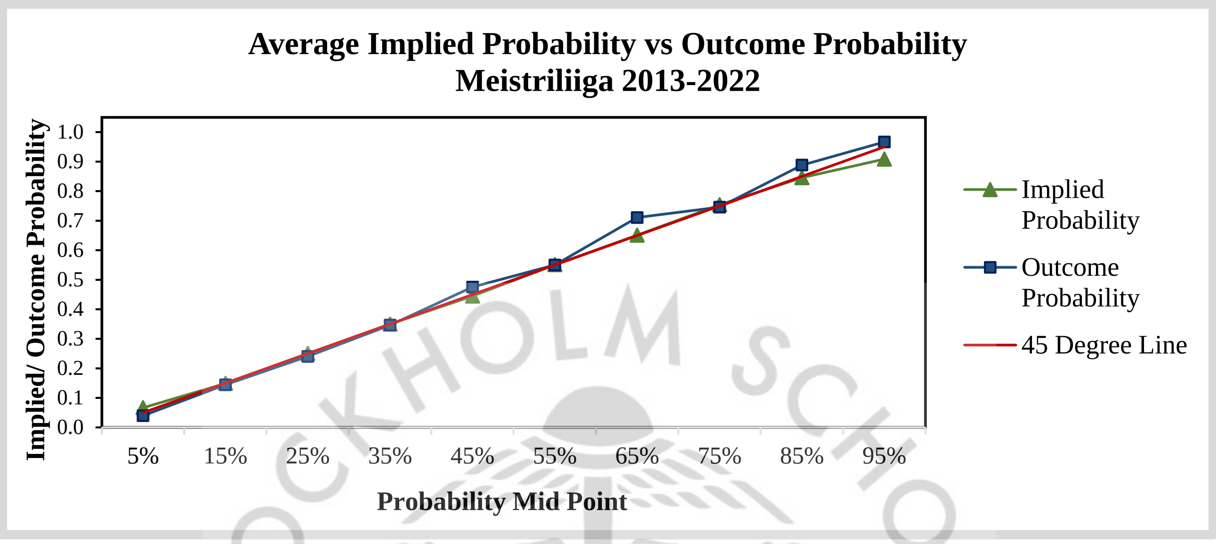
Appendix 5.4. Bookmakers mean implied probability vs outcome probability in Esiliiga during 2021-2022.

Esiliiga 2021-22

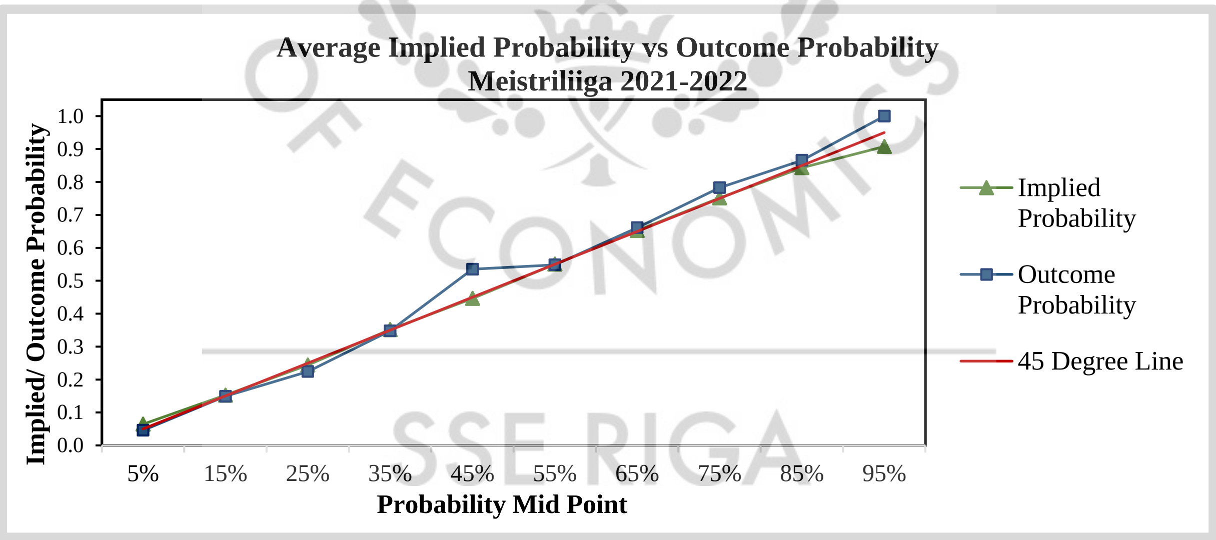
Implied Probability Decile Mid Point	Observations	Mean Implied Probability	Outcome Probability
5%	130	7.04%	5.38%
15%	225	15.48%	16.89%
25%	225	23.63%	18.22%
35%	54	34.38%	44.44%
45%	53	44.56%	47.17%
55%	78	54.42%	60.26%
65%	55	64.98%	63.64%
75%	63	75.63%	77.78%
85%	52	85.03%	82.69%
95%	4	90.43%	100.00%

These tables show the results of bookmaker calibration to check whether bookmakers' implied probabilities are true to actual outcome probability. 5% implied probability decile mid-point corresponds to implied probability of 0-10%. Created by the authors.

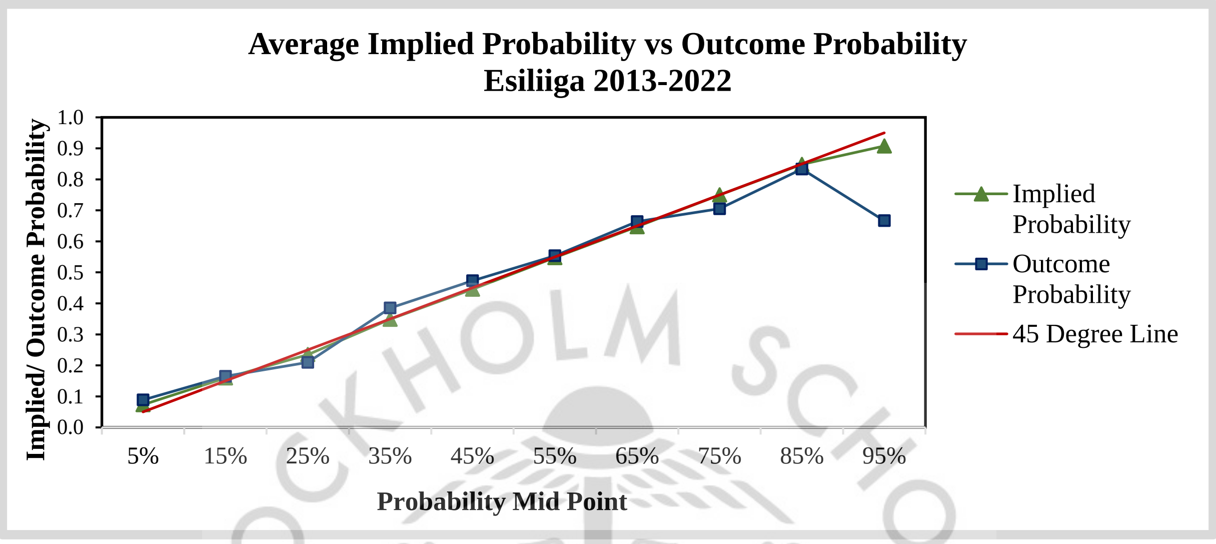
Appendix 6.1. Average implied probability vs outcome probability in Meistriliiga during 2013-2022.



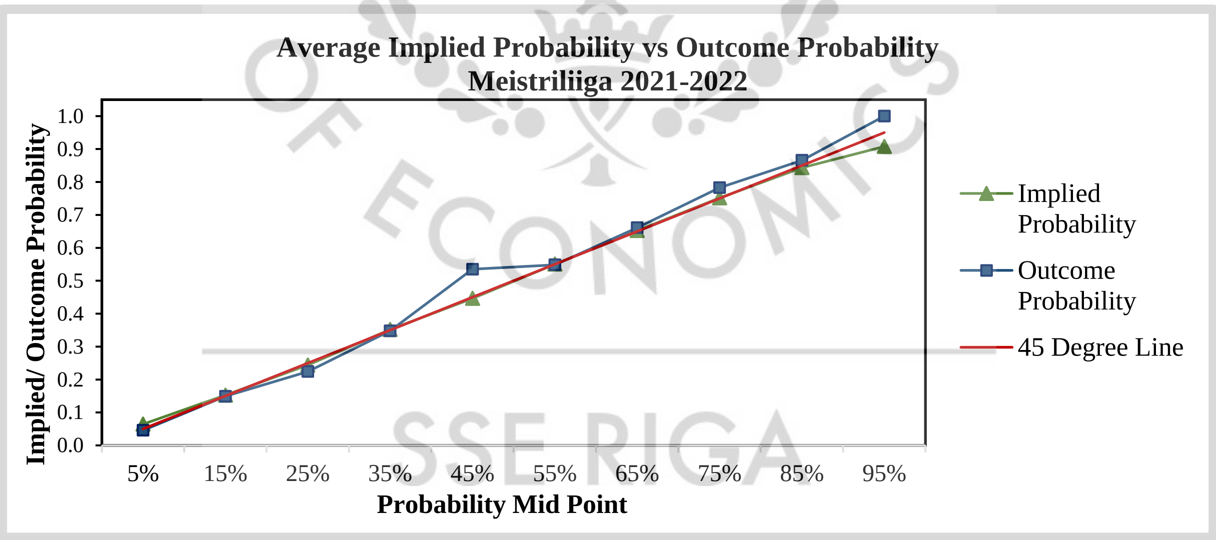
Appendix 6.2. Average implied probability vs outcome probability in Meistriliiga during 2021-2022.



Appendix 6.3. Average implied probability vs outcome probability in Esiliiga during 2013-2022.

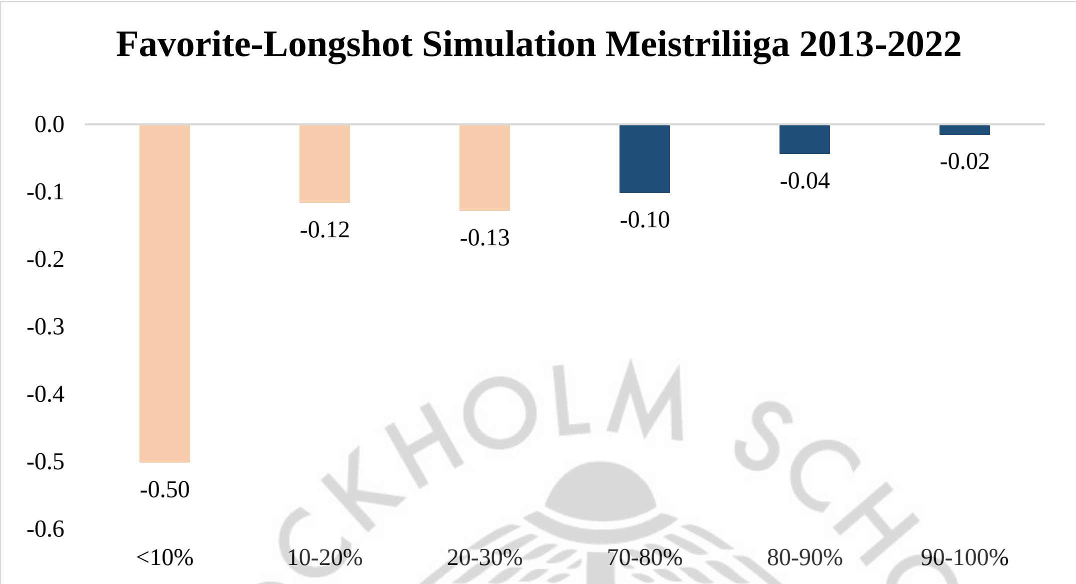


Appendix 6.4. Average implied probability vs outcome probability in Esiliiga during 2021-2022.

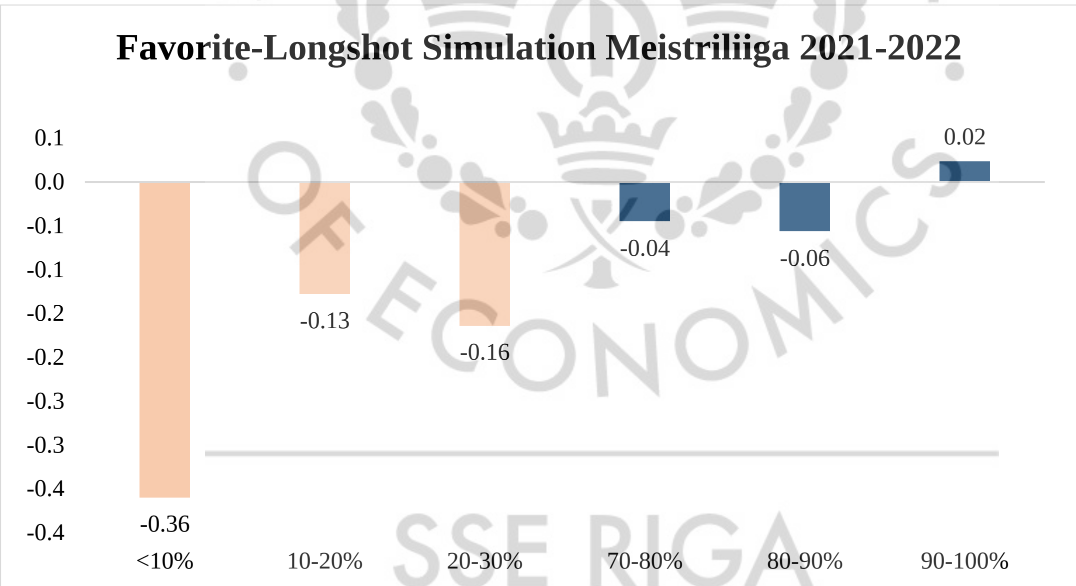


These graphs are visualized versions of Appendix depicting bookmakers' calibration compared to outcome probability. Created by the authors.

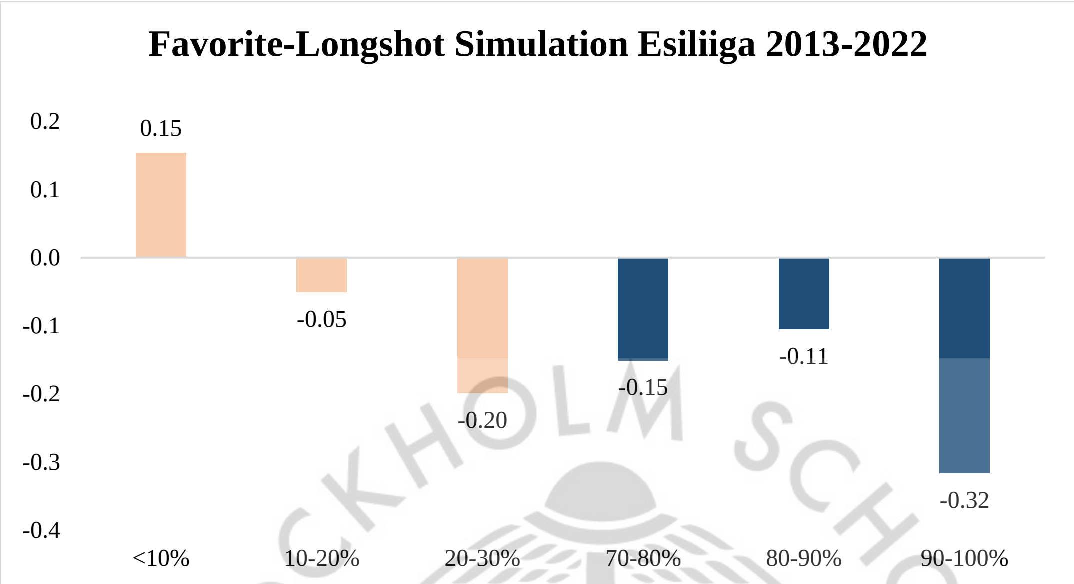
Appendix 7.1. Favorite-longshot bias simulation in Meistriliiga during 2013-2022.



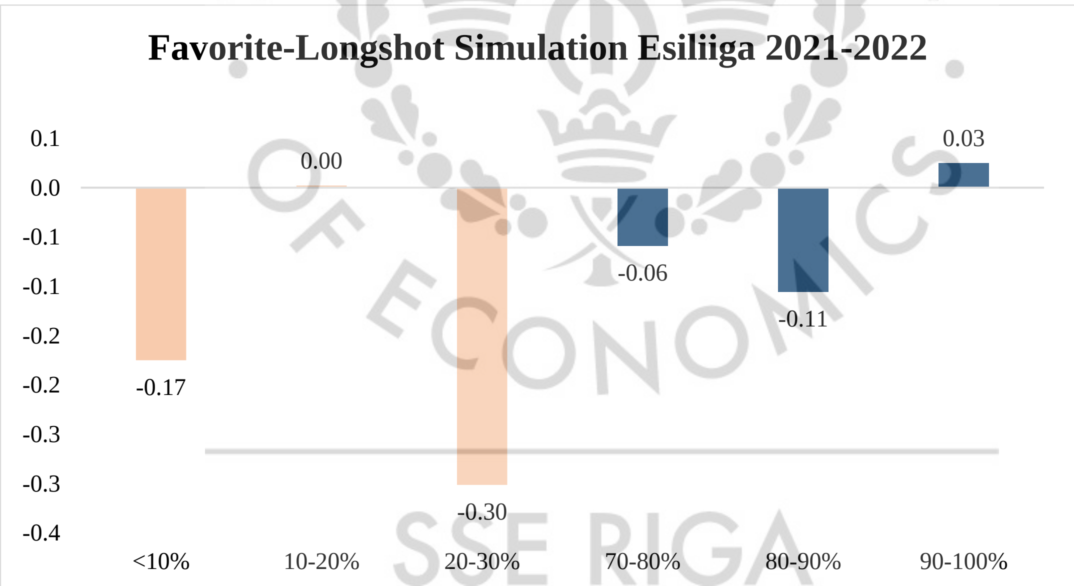
Appendix 7.2. Favorite-longshot bias simulation in Meistriliiga during 2021-2022.



Appendix 7.3. Favorite-longshot bias simulation in Esiliiga during 2013-2022.

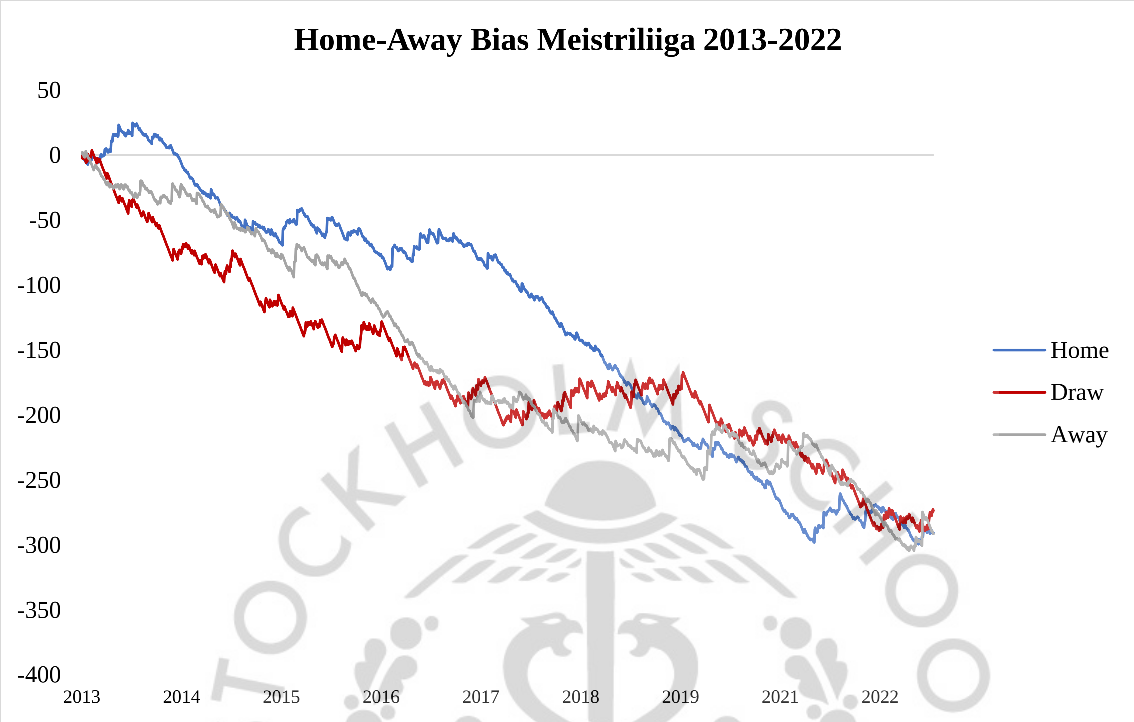


Appendix 7.4. Favorite-longshot bias simulation in Esiliiga during 2021-2022.

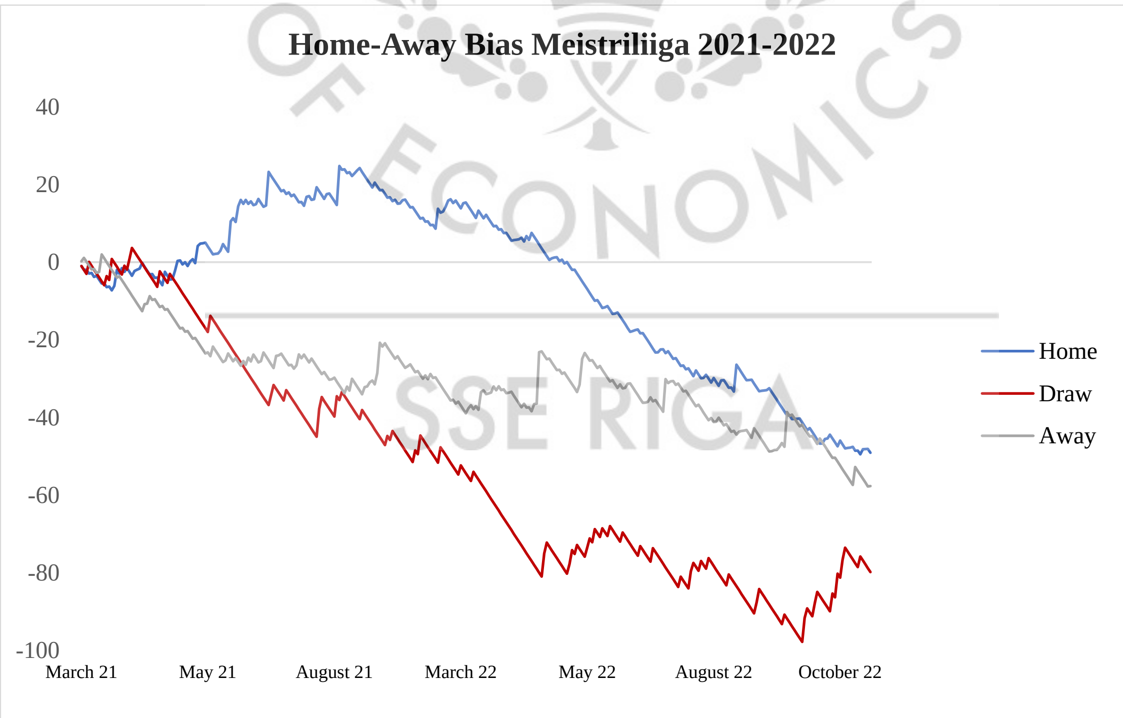


These bar graphs show the average return per 1-unit stake on each of the favorites and longshots decile ranges. Favorites accounting for ranges 70-80%, 80-90% and 90-100%. Longshots accounting for 0-10%, 10-20% and 20-30%. Created by the authors.

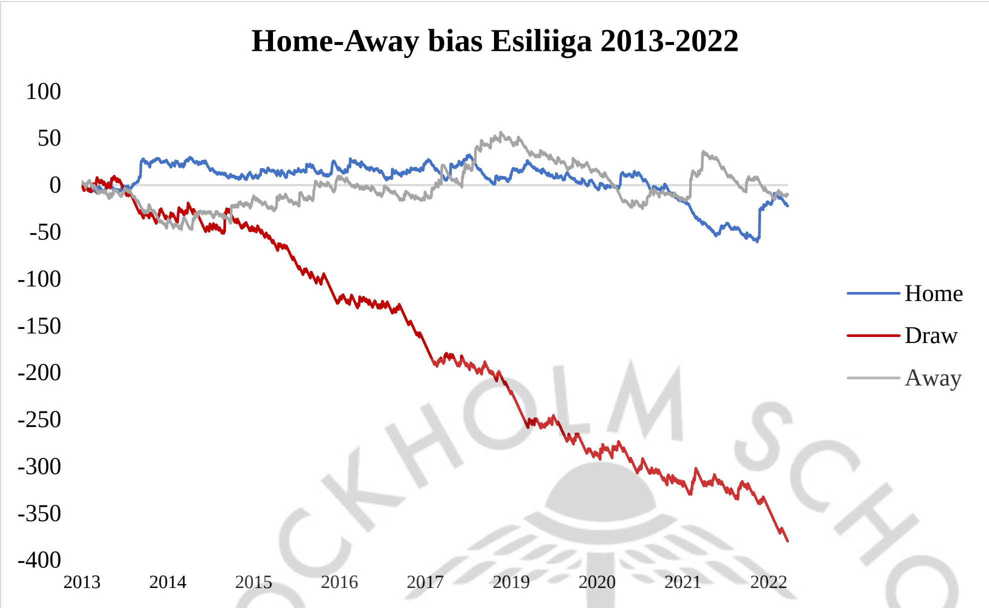
Appendix 8.1. Home-away bias simulation in Meistriliiga during 2013-2022.



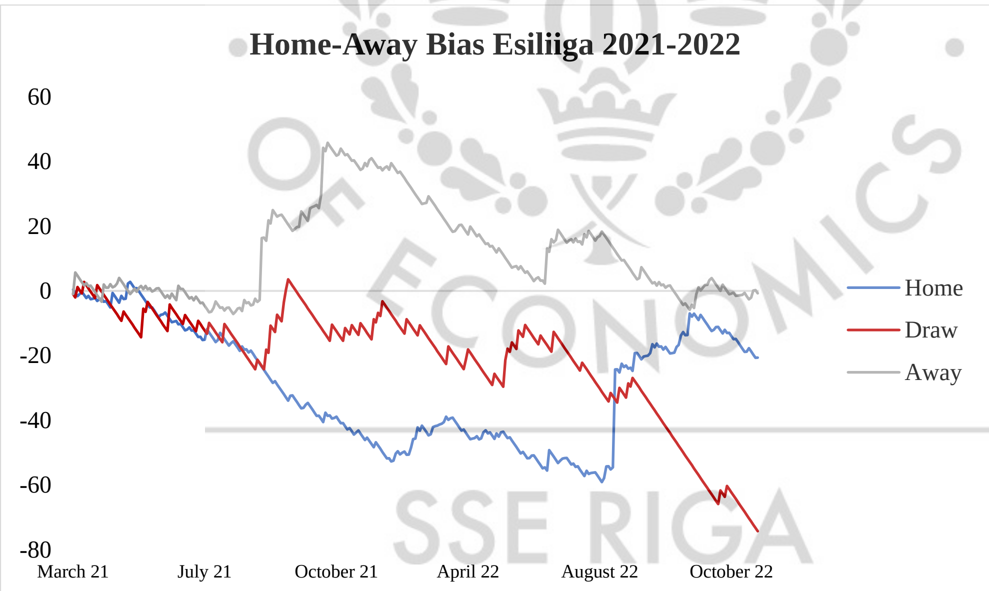
Appendix 8.2. Home-away bias simulation in Meistriliiga during 2021-2022.



Appendix 8.3. Home-away bias simulation in Esiliiga during 2013-2022.



Appendix 8.4. Home-away bias simulation in Esiliiga during 2021-2022.



These graphs show the cumulative return of home-away bias simulation where a 1-unit stake is being bet on home win, draw and away win. Created by the authors.

Appendix 9. Esiliiga home score of each team for game weeks 1-18 in 2022. Created by the authors.

Esiliiga GW 1-18					
Home team	**Games played**	**Goals**	**Average goals**	**Goals against**	**Average goals against**
Elva	9	15	1.67	12	1.33
FC Alliance	9	5	0.56	16	1.78
Flora U21	8	18	2.25	11	1.38
Harju Jalgpallikool	9	31	3.44	9	1.00
Levadia U21	9	22	2.44	9	1.00
Nomme Utd	9	12	1.33	17	1.89
Paide Linnameeskond U21	9	24	2.67	15	1.67
Parnu	9	6	0.67	15	1.67
Tulevik	9	6	0.67	25	2.78
Viimsi JK	8	26	3.25	11	1.38
Total	**88**	**165**	**1.88**	**140**	**1.59**

Appendix 10. Esiliiga away score of each team for game weeks 1-18 in 2022. Created by the authors.

Esiliiga GW 1-18					
Away team	**Games played**	**Goals**	**Average goals**	**Goals against**	**Average goals against**
Elva	8	11	1.38	15	1.88
FC Alliance	9	12	1.33	5	0.56
Flora U21	9	18	2.00	18	2.00
Harju Jalgpallikool	9	18	2.00	31	3.44
Levadia U21	9	22	2.44	22	2.44
Nomme Utd	8	16	2.00	12	1.50
Paide Linnameeskond U21	9	26	2.89	24	2.67
Parnu	9	3	0.33	6	0.67
Tulevik	9	4	0.44	6	0.67
Viimsi JK	9	10	1.11	26	2.89
Total	**88**	**140**	**1.59**	**165**	**1.88**

Appendix 11. Meistriliiga team attacking and defensive strengths for game weeks 1-18 in 2022. Created by the authors.

Meistriliiga GW 1-18				
	Home		Away	
Team	Attacking strength	Defensive strength	Attacking strength	Defensive strength
Flora	1.66	0.35	2.04	1.87
Kalju	1.05	0.43	1.36	1.18
Kuressaare	0.80	0.86	0.69	0.80
Legion	0.74	1.29	0.68	0.83
Levadia	1.54	0.35	1.47	1.54
Narva	0.69	1.07	0.58	0.69
Paide	1.79	0.95	1.21	1.79
Parnu JK Vaprus	0.62	1.84	0.69	0.55
Tallinna Kalev	0.42	1.55	0.70	0.33
Tammeka	0.55	1.47	0.69	0.55

Appendix 12. Esiliiga team attacking and defensive strengths for game weeks 1-18 in 2022. Created by the authors.

Esiliiga GW 1-18				
	Home		Away	
Team	Attacking strength	Defensive strength	Attacking strength	Defensive strength
Elva	0.89	0.84	0.86	1.00
FC Alliance	0.30	1.12	0.84	0.30
Flora U21	1.20	0.86	1.26	1.07
Harju Jalgpallikool	1.84	0.63	1.26	1.84
Levadia U21	1.30	0.63	1.54	1.30
Nomme Utd	0.71	1.19	1.26	0.80
Paide Linnameeskond U21	1.42	1.05	1.82	1.42
Parnu	0.36	1.05	0.21	0.36
Tulevik	0.36	1.75	0.28	0.36
Viimsi JK	1.73	0.86	0.70	1.54

Appendix 13. Esiliiga goal expectancies of home vs away teams for game week 19 in 2022. Created by the authors.

Esiliiga GW 19			
Home team	Goal expectancy	Away team	Goal expectancy
Elva	1.33	Nomme Utd	1.68
Harju Jalgpallikool	4.90	Paide Linnameesko	1.82
FC Alliance	0.20	Parnu	0.37
Flora U21	0.80	Tulevik	0.38
Levadia U21	3.77	Viimsi JK	0.70

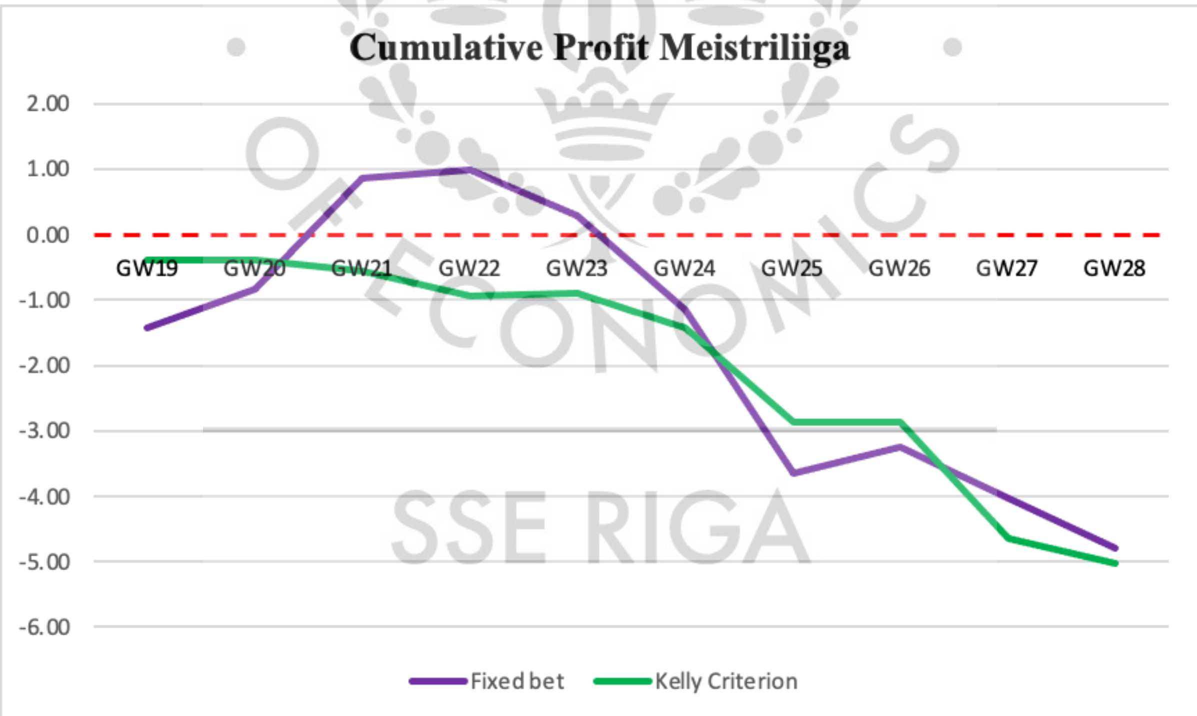
Appendix 14. Esiliiga goal distribution matrix for a matchup between Harju Jalpallikool and Paide Linnameeskond U21 from game week 19 in 2022. Created by the authors.

No. of goals	Paide Linnamees kond U21 goals	0	1	2	3	4	5	6
Harju Jalgpallikool goals	Probability	0.163	0.295	0.268	0.162	0.074	0.027	0.008
0	0.008	0.001	0.002	0.002	0.001	0.001	0.000	0.000
1	0.037	0.006	0.011	0.010	0.006	0.003	0.001	0.000
2	0.089	0.015	0.026	0.024	0.015	0.007	0.002	0.001
3	0.015	0.024	0.043	0.039	0.024	0.011	0.004	0.001
4	0.179	0.029	0.053	0.048	0.029	0.013	0.005	0.001
5	0.175	0.029	0.052	0.047	0.028	0.013	0.005	0.001
6	0.140	0.023	0.042	0.038	0.023	0.011	0.004	0.001

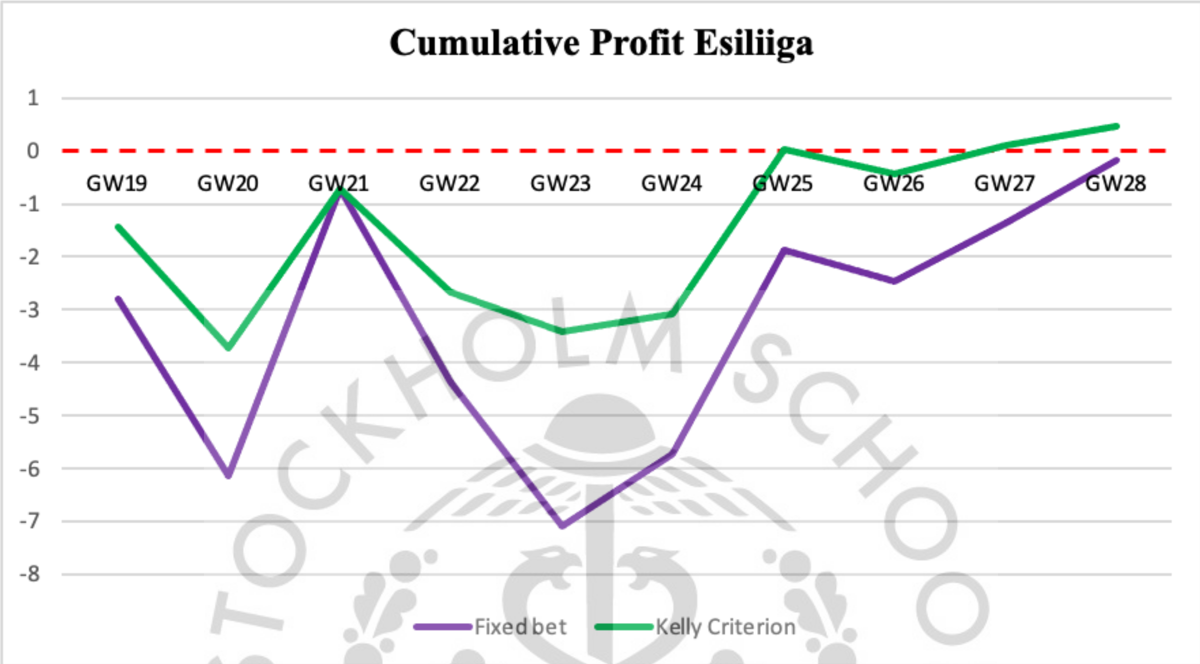
Appendix 15. Esiliiga Poisson model outcome probabilities for home vs away teams from game week 19 in 2022. Created by the authors.

Esiliiga GW 19						
		Probability				
Home team	Away team	Win	Draw	Loss	Match score	Correct prediction
Elva	Nomme Utd	0.31	0.24	0.45	3:1	No
Harju Jalgpallikool	Paide Linnameeskond	0.80	0.10	0.10	4:1	Yes
FC Alliance	Parnu	0.13	0.61	0.26	2:1	No
Flora U21	Tulevik	0.43	0.41	0.16	4:0	Yes
Levadia U21	Viimsi JK	0.89	0.08	0.04	1:2	No

Appendix 16. Meistriliiga cumulative profit graph for the whole test period of game week 19-28 in 2022. Created by the authors.



Appendix 17. Esiliiga cumulative profit graph for the whole test period of game week 19-28 in 2022. Created by the authors.



8. Acknowledgments

We would like to thank Morten Hansen, who served as our thesis advisor, for his support, constructive criticism, and directions while we wrote our thesis. Additionally, we would like to express our gratitude to the professors and students who took the time to read early drafts of our thesis and provide us with helpful comments. Lastly, we would like to point out that in this thesis AI-based software (ChatGPT) was solely utilized to help us better manage our time by helping us to find the right R-Studio commands quicker. Our football match outcome prediction model was the only portion of the code that was created with the help of AI.