CHARLES UNIVERSITY
FACULTY OF SOCIAL SCIENCES
Institute of Economic Studies

Hot hand bias in Czech sports betting market

Bachelor's thesis

Author: Jakub Čakan
Study program: Economics and Finance
Supervisor: PhDr. Jiří Kukačka, Ph.D.
Year of defense: 2024

Declaration of Authorship

The author hereby declares that he or she compiled this thesis independently, using only the listed resources and literature, and the thesis has not been used to obtain any other academic title.

The author grants to Charles University permission to reproduce and to distribute copies of this thesis in whole or in part and agrees with the thesis being used for study and scientific purposes.

During the preparation of this thesis, the author used ChatGPT 4, developed by OpenAI, and paraphrasing tool quillbot.com for assistance with drafting and refining the text and to enhance the clarity and flow of the narrative. After using these tools, the author reviewed and edited the content as necessary and takes full responsibility for the content of the publication.

Prague, April 30, 2024

Jakub Čakan

Abstract

This thesis investigates behavioral biases, specifically the "hot hand" bias, in the Czech sports betting market. Further, it explores two hypotheses: whether the Czech sports betting market efficiently incorporates all relevant information into the odds and the impact of the "hot hand" belief on bettor behavior. The study employs weighted and ordinary least squares estimation, respectively, revealing that while bookmaker's odds efficiently reflect comprehensive information, confirming market efficiency, bettors display significant "hot hand" bias. More precisely, it leads bettors to disproportionately favor teams on winning streaks, indicating an overreaction to recent team performances and an inefficiency on the part of bettors. Additionally, the thesis evaluates the profitability of betting strategies aimed at exploiting these biases. It does not find such strategies consistently yielding profits, highlighting the complex nature of betting markets and the difficulty of capitalizing on behavioral biases. This research enhances the understanding of behavioral biases in sports betting, illustrating the interaction between bookmaker precision and bettor irrationality within the Czech betting landscape.

JEL Classification	C31, G14, G17, G41
Keywords	market efficiency, sports betting, hot hand, behavioral bias
Title	Hot hand bias in Czech sports betting market

Abstrakt

Tato práce se zabývá behaviorálními vlivy, konkrétně přesvědčením "klam horké ruky", na českém trhu sportovního sázení. Dále zkoumá dvě hypotézy: zda český trh sportovního sázení efektivně zahrnuje všechny relevantní informace do kurzů a jaký vliv má přesvědčení o "klamu horké ruky" na chování sázejících. Studie využívá vážený odhad, respektive odhad pomocí obyčejných nejmenších čtverců, a odhaluje, že zatímco kurzy sázkových kanceláří efektivně odrážejí komplexní informace, což potvrzuje efektivitu trhu, sázkaři vykazují výrazné zkreslení "klamem horké ruky". Přesněji řečeno, vede sázkaře k neúměrnému zvýhodňování týmu na vítězných vlnách, což svědčí o přehnané reakci na nedávné výkony týmů a o neefektivitě sázkařů. Práce dále hodnotí ziskovost sázkových strategií zaměřených na využití těchto tendencí. Takové strategie tedy nepřináší konzistentní zisky, což poukazuje na komplexní povahu sázkových trhů a obtížnost využití behaviorálních vlivů. Tento výzkum rozšiřuje chápání behaviorálních vlivů ve sportovním sázení a ilustruje interakci mezi precizností bookmakerů a iracionalitou sázejících v českém sázkovém prostředí.

Klasifikace JEL	C31, G14, G17, G41
Klíčová slova	efektivita trhu, sportovní sazení, klam horké ruky, behaviorální vlivy
Název práce	Klam horké ruky v českém sportovním sázkařském trhu

Acknowledgments

I would like to express my gratitude especially to my supervisor PhDr. Jiří Kukačka, Ph.D., in particular for his knowledge, insightful advice, and willingness to assist.

In addition, I want to express my gratitude to my family and closest friends for their constant encouragement throughout my academic career, especially when it came to writing my thesis.

Typeset in LaTeXusing the IES Thesis Template.

Bibliographic Record

Čakan, Jakub: *Hot hand bias in Czech sports betting market.* Bachelor's thesis. Charles University, Faculty of Social Sciences, Institute of Economic Studies, Prague. 2024, pages 52. Advisor: PhDr. Jiří Kukačka, Ph.D.

Contents

List of Tables

List of Figures

Acronyms

ATF Away Team Favorite

ATS Away Team Streak

ATW Away Team Win

BS Brier Score

CLM Classic Linear Model

COVID-19 Coronavirus Disease

CZK Czech Crowns

D Draw

HTF Home Team Favorite

HTS Home Team Streak

HTW Home Team Win

LPM Linear Probability Model

MLB Major League Baseball

NBA National Basketball Association

NFL National Football League

NHL National Hockey League

OLS Ordinary Least Squares

U.S. United States

UK United Kingdom

WLS Weighted Least Squares

Bachelor's Thesis Proposal

Author	Jakub Čakan
Supervisor	PhDr. Jiří Kukačka, Ph.D.
Proposed topic	Hot hand bias in Czech sports betting market

Research question and motivation My main research question concerns several behavioral biases in sports betting market particularly in the Czech Republic.

The sports gambling market, similarly to other financial markets, comprises bettors with varying degrees of information (Paul and Weinbach 2011), creating a place for behavioral biases. As observed by Sauer (1998), well-informed bettors exist, developing different betting strategies, and systematic profitable opportunities are observed, but not persistent. Researchers developed several studies that approach bettorsâ€™ behavior on the gambling market. One of the behavioral biases in sports betting market was captured by Camerer (1989), who expanded the work of Gilovich, Valonne and Tversky (1985), observing a belief in hot hand in basketball matches, i.e., believing in making a shot after a sequence of shots made. Camerer (1989) observed this phenomenon in a basketball gambling market as well and proposed a prospective gambling strategy. Many authors such as Brown and Sauer (1993) or Woodland and Woodland (1994) commented on this fallacy and other papers (Paul and Weinbach 2011, 2014) developed more profound research.

This paper is aimed at individual bettors' decision-making, which is affected by exogenous individual performances of sport actors. Other areas of study concern, for instance, bettors' behavior as a reaction to betting odds. Those behavioral biases influence bookmakers in setting prices so they could yield greater profits, not acting like traditional market makers (Levitt 2004).

My thesis addresses the aforementioned hot hand fallacy and further study other behavioral biases comprising the overbetting of large favorites examined in research by Paul and Weinbach (2011) or more recent research on recency bias (Durand, Patterson and Shank 2021).

Contribution Existing research suggests that the aforementioned behavioural biases could be empirically measured and further discussed. However, all research papers conduct a study on American sports betting market, which significantly distinguishes from the European betting system. My contribution to existing research would consist of the necessary modification of the models that are sufficient for our dataset and applying such models on a unique Czech sports betting market dataset, providing exceptional evidence of such phenomenon.

Methodology My methodology builds on 3 research papers: I retrieve the econometric models from "Bettor Belief in the "Hot Hand": Evidence From Detailed Betting Data on the NFL.", "Bettor Misperceptions in the NBA: The Over Betting of Large Favorites and the "Hot Hand"." and "Behavioral biases in the NFL gambling market: Overreaction to news and the recency bias". Then, I perform those models on a Czech football league and the dataset would include every game from the last 5 seasons. I obtained the information from 2 sources: the first part of data set is gathered from a betting shop database (Tipsport) and includes the betting spread and particular betting odds on the main opportunities (home win, draw, away win). The second part is retrieved from a sport database (Livesport) and includes the results of all games. The joint database will then be extended by variables that would describe the current form/win streak of the team before a particular game and dummy variables that would account for game favorites and the team momentum. Consequently, I would form 3 hypotheses similar as in the aforementioned research papers, test the significances of proposed variables and the effects they have on the spread of bets.

Outline

1. Abstract
2. Introduction
    - why is my topic interesting
    - brief overview of existing knowledge
    - how I add to existing research
    - main results and what they mean
    - how is the thesis organized
3. Literature review and hypotheses
    - literature on behavioral biases in gambling markets
    - literature on hot hand fallacy

- what hypotheses will be tested
- motivation why is it reasonable to test them (why there should be a relationship)

4. Methodology

- relevant description of data
- why I use the independent and dependent variables I use, how they are measured
- how I perform tests

5. Results

- rejecting / not rejecting hypotheses
- my interpretation of the results

6. Conclusion

- broader interpretation of results
- implications for practice
- topics for further research

Core bibliography

1. Rodney J. Paul , Andrew P. Weinbach , and Brad Humphreys (2014). "Bettor Belief in the "Hot Hand": Evidence From Detailed Betting Data on the NFL." Journal of Sports Economics, Vol. 15(6) 636-649
2. Paul, Rodney J., and Weinbach, Andrew P (2005). "Bettor Misperceptions in the NBA: The Over Betting of Large Favorites and the "Hot Hand"." Journal of Sports Economics 6, 390-400.
3. Camerer, C. (1989). "Does the basketball market believe in the hot hand?" American Economic Review, 79, 1257-1261
4. Sauer, R. D. (1998). "The economics of wagering markets." Journal of Economic Literature, 36, 2021-2064.
5. Woodland, L. M., and Woodland, B. M. (1994). "Market efficiency and the favorite-longshot bias: The baseball betting market." Journal of Finance, 49(1), 269-280.
6. Robert B. Durand, Fernando M. Patterson, Corey A. Shank (2021). "Behavioral biases in the NFL gambling market: Overreaction to news and the recency bias." Journal of Behavioral and Experimental Finance 31.

Author

Supervisor

Chapter 1

Introduction

In recent years, the popularity of gambling has risen among Czech population, as they bet approximately 759 billion Czech Crowns (CZK) in gambling markets in 2022. CZK $118,8$ billion out of this betting volume was put on odds betting, accounting for $15,6\%$ share of bets, majority of those related to sports betting. The volume of winning bets paid out in this odds betting market amounted to CZK $106,8$ billion, generating a surplus of CZK 12 billion going to Czech bookmakers (Novinky.cz 2023). In this manner, we could conclude that betting shops in Czech Republic are a profitable business opportunity. However, the profit realization could result from different strategies employed.

Differently from casino's game of chance such as roulette or machine slots, where the expected gross profit known as "house edge" is statistically calculated (Forbes 2018), the price setting by bookmakers affects their returns and can lead to long-term losses if prices are set incorrectly. Levitt (2004) proposes several bookmakers' strategies that lead to profitability. As the first option, he expects bookmakers to be effective in predicting the actions of bettors. Therefore, they are not required to possess any special ability to predict the actual result of sporting events. Further option, according to Levitt (2004) study, concerns the fact that when it comes to correctly forecasting game results, bookmakers should consistently outperform gamblers. This means that, on average, they profit at least from commissions paid by the bettors on each bet. Finally, Levitt suggests the bookmakers possess both of those abilities, resulting in an even better expectation of increasing profits.

Numerous studies (Pankoff 1968; Sauer et al. 1988; Goddard and Asimakopoulos 2004) focused on the scenario where bookmakers efficiently predict the match outcomes and tested this hypothesis in different sports betting markets. Based on their realizations, other researchers (Woodland and Woodland 1994; Paul et al. 2014) proceeded with further investigations that concerned the fact that the inefficiencies in the betting market arise from the irrationality of bettors. They often face various behavioral biases in their judgement which prevent them from assessing the correct risk of a given bet. Therefore, bettors are unable to predict the sport's outcomes more accurately than bookmakers, resulting in overall loss. Several biases, such as favourite-longshot bias (Quandt 1986) or gambler's fallacy (Kahneman and Tversky 1972) were introduced, however one of the best applicable phenomena was first observed in basketball matches by Gilovich et al. (1985).

The "hot hand" hypothesis states that players who make a shot have a higher chance of hitting the next one than those who miss. This hypothesis is supported by the majority of basketball viewers according to the Gilovich et al. (1985) study. Shortly after, the hypothesis was implemented in the sports betting market. It indicates that bettors believe teams on winning streaks to win the next game more than the bookmakers do. Consequently, it convinces them that they could take advantage of the mispricing on behalf of the bookmaker. Unsurprisingly, the reality was found to be the opposite (Paul et al. 2014).

This thesis builds on the studies of Paul et al. (2014). and Goddard and Asimakopoulos (2004). Furthermore, we assume that even though the distortion of prices may be present due to the fact that bookmakers seek abnormal profits, they primarily target the most accurate predictions of sport outcomes in order to ensure their profitability. Consequently, we argue that the bias persists on the side of the bettors. Even though there might be bettors who could outperform bookmakers in the long-run, majority of them are unable to correctly evaluate the risks of each bet due to the presence of such biases. Our goal is to observe the "hot hand" bias. Not only we found this phenomenon to be fascinating and straightforward, but we are firmly convinced that the study should be expanded to include the European betting market. Our focus concerns other behavioral biases as well; however, due to the limitations of our data collection, it has prevented us from studying multiple biases in this thesis.

Our contribution to the existing literature consists of several parts. Most importantly, we perform our analysis on a unique Czech betting dataset representing the Czech sports betting market, which operates differently from the American and English markets. To elaborate on that, we combine the test of efficiency of the betting odds together with the test of the presence of behavioral biases both on the bookmaker's side and on the bettor's side in one work. Additionally, we adopt models from different studies that focused on American and English sports betting markets, respectively. Consequently, we adjust these models to mitigate the limitations of our dataset. Therefore, we conduct a complex analysis on a separate betting market, where we tend to support the proof of market efficiency on the bookmaker's side and validate the presence of behavioral bias on the bettor's side. This altogether can be considered as a support for the cause of the bookmaker's profitability presented in the beginning.

Hence, we test these two hypotheses:

> Hypothesis #1: Czech sports betting market is efficient, and all important information is included in the assessed probabilities.

> Hypothesis #2: Czech bettors act according to the "hot hand" hypothesis, which is reflected in the betting distribution.

We structure the remaining parts of the thesis as follow: In the section 2 we review the existing literature on the previously mentioned topics, in the section 3 we present the dataset we were working with, in section 4 we describe the statistical methods behind the regressions performed, in section 5 we present the results of our tests and in section 6 we conclude our findings and decide whether we reject or not the working hypotheses.

Chapter 2

Literature review

2.1 Sports Betting Markets

As in the definition by Teall (2022), "the sports betting market resembles a financial market that has a large number of participants, possessing different levels and types of information, that are driven by their financial incentives to logically respond to the best information available in a sports competitive market with a vision to monetize their knowledge". As Sauer (1998) states, even though this market represents a very small part of the economy, it raises interesting opportunities in economic analysis. It is supported by the fact that it does not have to deal with problems concerning pricing issues, creating a simple financial market for economic study. The horse racing market was the subject of the first studies on sports betting markets. Griffith (1949) conducted an early study to test the validity of the socially determined odds and found evidence in favor of the study. The study also revealed that participants overvalued long-odded horses and undervalued short-odded horses, which are bets with lower odds. Subsequent research in other sports betting markets in the years that followed mainly examined the market's efficiency and its relation to other betting and financial markets.

2.1.1 Market Efficiency

The term "market being efficient" was first used by Fama (1965) in his theory of random walks in capital markets. In his subsequent study, Fama (1970) defines an efficient market as one in which prices "fully reflect" all available information, which could imply the independence and uniform distribution of subsequent price changes. Three information subsets are used by Fama (1970) to categorize price adjustments: weak-form efficiency, which is based on historical price data; semi-strong-form efficiency, which is concerned with price adjustments in relation to other publicly available data; and strong-form efficiency, which is concerned with the question of whether certain investors or groups have monopolistic access to information that is essential for price formation. The first study examining the sports betting market efficiency was published not long after the random-walk theory was developed. Pankoff (1968), for instance, examined the American football betting market and discovered evidence that it was not efficient. Pankoff (1968) was supported by other research that refuted the efficient football betting market theory, including investigations by Sauer et al. (1988) and Zuber et al. (1985). Studies on other sports betting markets, however, seemed to have different results. Evidence of the racetrack betting market was found to be unexpectedly efficient by Ziemba and Thaler (1988), Goddard and Asimakopoulos (2004) and Graham and Stott (2013) reached a similar finding in a UK football betting market. Nevertheless, the efficiency of these gambling markets is still being the subject of further studies.

2.1.2 Difference from Other Financial Markets

Although researches carried out by Sauer (1998) suggested that the sports betting industry was comparable to other financial markets, Levitt (2004) comes to a different conclusion. According to Levitt (2004), bookmakers are not conventional market makers but rather have better predicting abilities than gamblers and routinely take advantage of bettor biases by establishing prices that are different from the market-clearing price. Levitt (2004) used his own dataset, which he obtained from a controlled experiment in which participants attended National Football League (NFL) game betting competition, to illustrate this observation. This dataset was special because it included information on the total number of bets made on both sides of the wager, which was not available in earlier research and allowed for a deeper analysis of participant behavior. The original claims that bookmakers manipulate prices to take advantage of their greater talent and generate seemingly larger profits than they might if they behaved like traditional market makers are eventually provided with evidence by Levitt (2004). This study therefore encouraged additional research conducted by Paul et al. (2014), for instance, to examine behavioral biases in the sports betting markets that had previously been looked into.

2.2 Behavioral Biases

As noted by Teall (2022), numerous studies have discovered evidence of persistent weak-form inefficiencies in sports betting markets. Additionally, as noted by Durand et al. (2021), behavioral finance has discovered a plethora of behavioral and cognitive patterns that impact the rationality of people's financial decision-making. It has not been established, nevertheless, that specific behavioral biases are evident in every sports betting market. In their prospect theory, Kahneman and Tversky (1979) demonstrated how people value things differently and see outcomes as gains or losses rather than as the ultimate condition of wealth. Additionally, individuals have a tendency to overvalue tiny probabilities, providing some support for the hypothesis that an inclination for taking risks stems from an incapacity to cope with losses or realize anticipated rewards. Kahneman and Tversky (1979) provides support for those who prefer to bet on non-favorites and leaves room for additional research.

2.2.1 Favourite-longshot Bias

Quandt (1986) was the first to notice the occurrence of favourite-longshot bias, and subsequent research has focused further. To define, when the expected return for betting favorites is higher than the expected return for betting on teams with a lower probability to win (longshots), the bias emerges. Quandt (1986) discovered proof that this is the equilibrium of the market result in horse race wagering. Favourite-longshot bias, however, can also be understood in other ways. For instance, Woodland and Woodland (1994) found that most racetrack betting markets exhibit an underbet on favorites and an overbet on longshots. However, they discovered that the results in gambling marketplaces varied. This study claims that bettors in Major League Baseball (MLB) exhibit a relatively opposite bias, leading to the development of a so-called reverse favourite-longshot bias. With a larger data set, Woodland and Woodland (2003) revised their study and furthermore, they were unable to identify any differences from their original findings. They further expanded on their investigation and presented additional data from the National Hockey League (NHL) betting market (Woodland and Woodland 2001). However, Cain et al. (2000) provided evidence from the United Kingdom (UK) football betting market to support the initial idea, and Rossi (2011) joined other research in supporting this concept by adding evidence from the Italian football betting market to this topic. In any case, a number of sources indicate that those abnormalities are common in gambling markets and present some lucrative chances for wagering (Cain et al. 2000).

2.2.2 Gambler's Fallacy

Gambler's fallacy was first identified by Kahneman and Tversky (1972) and subsequently verified by Clotfelter and Cook (1991) and Ayton and Fischer (2004). It was first noticed at the Monte-Carlo casino in the early 20^{th} century. This psychological bias arises from the widespread belief that an event, such as an odd number in a roulette game, reduces the likelihood that it will occur in the following game, even though the event is known to be objectively independent of the previous trial. Kahneman and Tversky (1972) found evidence for this observation's validity for local parts of the sequence since it was demonstrated to apply universally to a sequence of occurrences. While Ayton and Fischer (2004) confirmed the hypothesis within the gambling market, Clotfelter and Cook (1991) discovered evidence of a similar fallacy in another gambling market lottery play. However, no literature has been found to study this kind of bias in any sports betting market.

2.2.3 Hot Hand Fallacy

The hypothesis was first subject to a study of perception of sport performance observed in basketball by Gilovich et al. (1985). They define this hypothesis as the idea held by spectators that players who have made successive shots in the past have a higher chance of scoring the next shot than players who are not on a "streak". They not only demonstrated that viewers generally hold this idea, but they also discovered evidence that actual results defy this belief, with shot outcomes turning out to be roughly independent or even somewhat negatively autocorrelated. Miller et al. (2014), on the other hand, conducted updated research on this subject using more comprehensive data and reached a conclusion that challenges the economic significance of the hot hand fallacy and refutes the theory developed by Gilovich et al. (1985). They also confirmed that greater shooters have a propensity to sustain a "hot streak". However, this phenomenon piqued the interest of researchers, who conducted a similar study on professional darts by Ötting et al. (2020). Although the cognitive bias in question is not applicable to professional darts, their findings statistically strongly support the "hot hand fallacy".

Hot Hand in Mutual Funds

Scientists searched for uses in other areas, and the mutual fund industry was also exposed to this phenomenon. The “hot hands strategy” was introduced by Hendricks et al. (1993) and refers to a method of selecting funds that exhibit consistent outperformance in the short term. They discovered that this approach works well on their dataset since choosing these funds can greatly beat average mutual funds. Furthermore, Sirri and Tufano (1998) discovered behavioral pattern evidence suggesting these top-performing funds draw a disproportionate amount of cash and investors. Nevertheless, Rabin (2002) eventually linked this discovery to the idea of overinference. Consequently, Huber et al. (2010) related these findings to the well-known “hot hand fallacy” and investigated the degree to which both individuals’ and groups’ investment choices are influenced by this cognitive bias.

Hot Hand in Sports Betting Market

Eventually, it was thought that this phenomenon would not just apply to sports betting but also have parallels with the mythical “hot hand fallacy” in the gambling industry. This was noted by Camerer (1989), who addressed the finding that, while actual outcomes may not represent the “hot hand fallacy”, betting odds may, and suggested the profitable tactic of betting against teams that are on a winning streak. He gave evidence, however, not sufficient to support his hypothesis. In response to this study, Brown and Sauer (1993) contended that actual outcomes also reflect the mythical “hot hand”, although they did not reject Camerer (1989) hypothesis. Building on these findings, Paul and Weinbach (2005) validated the hypothesis that betting against teams on a run is a winning strategy in the National Basketball Association (NBA) betting market. Furthermore, in the instance of the baseball betting market, they reached the same conclusion (Losak et al. 2023). Alternatively, in an effort to explicitly assess bettor preferences in this market, Paul et al. (2014) looked into the possibility of bias in the percentage of bets made on the teams in the NFL betting market. They repeatedly demonstrated the existence of the mythical “hot hand” bias in sports betting markets by finding evidence in their experiment that this impression leads to a disproportionate distribution of bets.

2.2.4 Other Biases Observed in Literature

Researchers did not hesitate to explore other biases that might arise in betting markets, even though the aforementioned biases were the focus of other investigations across the markets. Forrest and Simmons (2013) looked for evidence in the Spanish football betting market to suggest that the percentage of supporters supporting each club during a game appears to affect the match odds, giving bettors who support the more popular team better terms. They reached the opposite finding, disproving Levitt (2004) assertion that bookmakers manipulate odds to take advantage of bettors' preferences and drive up the cost of more popular wagers. Additionally, Durand et al. (2021) conducted a thorough analysis of the NFL betting market in an effort to identify instances that would skew the odds. They came up with the hypothesis that when fans learn that a club's starting quarterback is not playing, they will bet less on that team. In addition to seeing that bookmakers take use of these preferences to produce larger returns, which is consistent with Levitt (2004) premise, they successfully validated their hypothesis. The existence of biases in odds setting and active betting cannot be ruled out, despite the fact that it has been extensively researched whether individuals perceive odds correctly (Andersson and Nilsson 2015) and supported that subjective and objective probabilities are similar (Quandt 1986).

2.3 Market Efficiency and Hot Hand Hypothesis Experiments

Regarding Fama (1970) hypothesis of markets being efficient, he did not come to a conclusion in favor of it, despite the weak-form tests of the market efficiency model appearing to substantially support the hypothesis. Not to mention, researchers did not hesitate to draw such conclusions in sports betting markets. Two significant problems were posed by Pankoff (1968): first, whether or not players in gambling markets make decisions based on potentially exploitable patterns; and second, whether or not prices in those markets fairly reflect the intrinsic values of those markets. Following a number of investigations into the answers to those concerns, Zuber et al. (1985) performed a straightforward regression to test for market efficiency. It included subjective values (bookmakers' initial odds predictions) as independent factors and objective values (actual results) as dependent variables. In order to explore the "hot hand" hypothesis further, Brown and Sauer (1993) modified the model by adding dummy variables that accounted for the win and lose streaks of both visitors and homeowners. These models served as the foundation for additional research that tested similar hypotheses in several sports betting markets.

2.3.1 Area of Study

There are a number of variations in the tests conducted in the various research, some of which have European or United States (U.S.) origins. The primary distinction is in the kinds of odds that are employed in a study. Predicting the difference in points between rival teams is the focus of the majority of U.S. studies conducted on a point-spread betting market. We could anticipate that the opening point spread will be the most accurate unbiased forecast of the game's result if the information at hand is handled effectively (Zuber et al. 1985). However, money line odds characteristic of the American betting market were employed by Woodland and Woodland (1994). The odds next to a favorite indicate how much you would have to bet in order to win \$100. For instance, the moneyline of +150 means person has to bet \$100 in order to win \$150. In case of moneyline of −150, on the other hand, person must bet \$150 to win \$100 (Forbes 2024). They then convert the odds into implied probabilities for their market efficiency model, which makes the test easier to understand. Then, European decimal odds and British fractional odds separate the betting market in Europe. In fractional odds, the profit that the winning bet would earn is displayed by multiplying the fraction by the stake amount. On the other hand, decimal odds display a sum that includes the initial bet. One number, representing the amount a successful bet on a \$1 wager would collect, is displayed for decimal odds (Athletic 2022). Forrest and Simmons (2013) employ the implied probability resulting from decimal odds in their model examining the presence of attendance bias in bets, while Cain et al. (2000) use fractional odds for specific match outcomes to analyze the presence of favourite-longshot bias.

2.3.2 Data

Different sets of data are needed for the behavioral bias and market efficiency evaluations, which were carried out on various sports betting markets. In order to test the market efficiency hypothesis in the U.S. football betting market, the early research (Pankoff 1968; Zuber et al. 1985; Sauer et al. 1988) used point spreads, primarily requiring the spreads set by bookmakers and real point spreads. The initial studies on the "hot hand" existence were primarily conducted in the basketball betting market (Camerer 1989; Brown and Sauer 1993; Paul and Weinbach 2005). In these studies, the analysis could be based on the NFL point spread market efficiency, but they also needed data on home and away streaks. A different set of data was required by Paul et al. (2014), where they inspected the behavioral pattern in percentage bets of bettors placed on matches in NFL, and this dependent variable was much needed in regression. Forrest and Simmons (2013) employed attendance as a measure to account for supporter bias in the European football betting market, and used implied probability and actual outcome as independent and dependent variables, respectively.

2.3.3 Models

Ordinary Least Squares (OLS) regression was utilized for the studies on the market efficiency hypothesis and "hot hand" hypothesis conducted on point spread betting marketplaces. Despite the fact that the dependent variable is logically constrained by 0 and 1, research studies about biases in percentage bets (Paul et al. 2014; Durand et al. 2021) also employed OLS regressions with robust standard errors. Goddard and Asimakopoulos (2004) and Forrest et al. (2005) focused their studies on odd-setting betting markets instead and therefore utilized Weighted Least Squares (WLS) regression to account for the heteroscedasticity in their analysis. Next, Forrest and Simmons (2013) employed the probit model for their regression, because their dependent variable is a dummy that gains one if the wager wins. A comparable logit model was employed by Rossi (2011) to examine the bookmakers' predictability because his dependent variable was equal to one in the event that the winning side prevailed. Additionally, Cain et al. (2000) discovered that poisson and negative binomial distributions were useful describers when attempting to explain goal-scoring processes through fixed odds utilizing such regressions.

Chapter 3

Data

3.1 Dataset

One of the largest Czech betting companies, Tipsport, provided us with proprietary information, which is what makes this analysis unique. Unfortunately, restrictions apply to the availability of these data, which were used under license for the current study and are therefore not publicly available. The data is, however, available from Tipsport upon reasonable request. Dataset collection includes 1342 observations of cross-sectional data of all games played in the Czech football league between the seasons 2018/19 and 2022/23. It consists of closing decimal odds (ODD^r) determining odds for three main opportunities: Home Team Win (HTW), Draw (D), and Away Team Win (ATW). Furthermore, the percentage distribution of bets is enclosed (Percentage bet), meaning how many bettors bet on each of the main betting opportunities (HTW, D, and ATW).

The second set of data was retrieved from the official site of the Czech football league (Fortunaliga.cz 2024) which included the actual results of all matches. However, it was crucial to first clean up both parts of the dataset and correctly link both parts to each other. It was necessary due to the fact that the dataset acquired from Tipsport contained closing odds on matches that were postponed and therefore the odds were listed repeatedly. Most of those games were postponed during the Coronavirus Disease (COVID-19) pandemic.

It is important to note that between seasons 2020/21 and 2021/22 a new league format was established, bringing more matches between equally strong teams. That also gives new perspective to our analysis. A more thorough description of the league format is attached in Appendix A. In addition, it is essential for our investigation to include every game since we look into the impact of earlier games. Whereas match outcomes prior to a specific game are closely related to that, they cannot be disregarded in order for our research to be completed. Moreover, it is crucial to remember that certain teams were included even though they were relegated or promoted.

3.2 Data Processing

For our hypothesis #1 we generated implied probability odds derived from the odds set by Tipsport bookmakers. The decimal odds used in this thesis, however, have to be interpreted differently from the fixed and money-line odds utilized in the English and American betting market, respectively. The implied probability odds could be considered equivalent to the reciprocal of decimal odds (Forrest and Simmons 2013). To further analyze, there are two factors that determine the decimal odds ODD^r. It includes the implied probability odds ϕ^r that a particular event will result in an outcome, in our case either HTW, D, or ATW. The second item concerns a positive margin called a take-out rate that allows bookmakers to profit λ. The inverse equation for implied probability odd could be formally modeled as follows (Andersson and Nilsson 2015):

$$\phi^r_{i,j} = \frac{1}{ODD^r_{i,j} \times (1 + \lambda)} \tag{3.1}$$

Therefore, it is crucial to modify the odds by the bookmaker's margin to properly interpret implied probability odds which should add up to 1.

To further test our hypothesis #1, we created three dummy variables r = HTW, D, or ATW. These variables are implemented as dependent variables in the three respective models discussed in methodology section. The particular variable equals 1 if the match ends with a result of the given outcome r. If it does not result in that outcome, the variable takes 0. An example can be seen in the table below:

Table 3.1: Match outcome dummy variables

MATCH	**SCORE**	**HTW**	**D**	**ATW**
Slavia Praha - Olomouc	2:1	1	0	0

For the sake of hypothesis #2, we created dummy variables for the scenario in which there is an Away Team Favorite (ATF) in that match based on the odds assigned to that match. Similarly, Home Team Favorite (HTF) dummy was created.

Following that, we had to create variables for the winning and losing Home Team Streak (HTS) and Away Team Streak (ATS). The real streak was tracked when the team had either won its previous run of games (winning streak) or when the team had drawn or lost the previous consecutive games (losing streak). In case of a winning streak the variable takes on positive values, and for losing streaks it takes on negative values. However, in the regression concerning hypothesis #2, the first games needed to be omitted as for them the streak could not be computed. In order for our dataset to be considered relevant, we also had to exclude games in which a team re-entered the league (i.e., promoted in following seasons).

Next, we present the summary of descriptive statistics in table 3.2:

Variable	Min	Median	Mean	Max	Std.
ODD^{HTW}	1.050	2.155	2.730	18.20	1.945
ODD^{D}	2.860	3.650	4.127	17.20	1.385
ODD^{ATW}	1.150	3.550	5.080	42.00	4.646
ϕ^{HTW}	0.053	0.446	0.456	0.921	0.188
ϕ^{D}	0.056	0.264	0.249	0.334	0.051
ϕ^{ATW}	0.023	0.271	0.294	0.835	0.165
Percentage bet HTW	0.002	0.611	0.537	0.996	0.377
Percentage bet D	0.002	0.106	0.160	0.625	0.154
Percentage bet ATW	0.001	0.111	0.303	0.995	0.349
HTS	-19.00	-1.000	-1.390	10.00	3.215
ATS	-20.00	-1.000	-0.941	11.00	3.345

Table 3.2: Descriptive statistics

We can see that all variables are within the expected range. Odds have a lower bound bigger than 1, implied probabilities and percentage bets are within the range $(0; 1)$ and home and visiting's team streak take on both negative and positive values. It is also observable that winning streaks are not that common, as more than half of matches do not include teams on winning streaks. It follows from the fact that on average less than half of teams wins a particular game, given that the match could end with a draw which is in our case considered as a losing streak. Furthermore, we checked that for percentage bets and implied probabilities, the sum of the three outcomes always added up to 1. And on top of that, we investigated the binary variables, and for all match outcomes it covers all matches as well as the home/away favorites. We consider our dataset to meet the requirements for further investigation.

Chapter 4

Methodology

In this section, we summarize the methodology of this thesis based on Goddard and Asimakopoulos (2004) and Paul et al. (2014). For the first hypothesis, Goddard and Asimakopoulos (2004) suggests the usage of WLS estimation of the market efficiency model, while Paul et al. (2014) recommend OLS regression for their model as they correct the standard errors for heteroscedasticity by using White standard errors.

4.1 Betting Market Efficiency Model

Goddard and Asimakopoulos (2004) followed up on the work of Pope and Peel (1989), who estimated the efficiency of prices set by bookmakers for the fixed odds in English football. They proposed a simple test of weak-form efficiency hypothesis that builds on the regression of the actual outcomes of matches on the implied bookmaker's probabilities. For this purpose they use Linear Probability Model (LPM) which is modeled as follow:

$$r_{i,j} = \beta_0 + \beta_1 \phi^r_{i,j} + u_{i,j} \tag{4.1}$$

where dummy variable r and the implied probability odd ϕ^r are defined as in the Data Section. Other unexplained information is included in the random error term u. The necessary weak-form efficiency condition consists of a joint hypothesis: $\beta_0 = 0, \beta_1 = 1$. Since it holds for LPM that there the presence of heteroscedasticity is known, as a result, the estimators become inefficient. However, the OLS estimators are unbiased estimators of population parameters that determine the probability $p(x)$, and as a result, through fitted values of the

OLS estimation we can estimate the conditional variance of dependent variable as:

$$\hat{h}_i = \hat{y}_i \left(1 - \hat{y}_i\right) \tag{4.2}$$

Therefore, as suggested by Pope and Peel (1989) and Goddard and Asimakopoulos (2004), the estimation of the equation 4.1 using WLS is employed using fitted values in formula $\hat{r} \times (1 - \hat{r})$ to compute weights. However, to proceed with the WLS estimation of LPM, we need to account for the cases where the fitted values are not within the range: $0 < \hat{y} < 1$, because then the function $\hat{h}$ becomes negative and therefore the weights will be invalid. If we take that into consideration, we can substitute the values below 0 with a value 0.001 and values above 1 with a value 0.999. For all regressions, the number of over or under-fitting values lies below 10, which could be considered negligible for our case of 1342 observations. We then simply estimate the equation by WLS, using weights $\frac{1}{\sqrt{\hat{h}}}$.

Furthermore, we assume our regression model satisfies the assumption of random sampling by including every game and accounting for influences from previous matches through comprehensive variable, such as bookmaker's odds, which should aggregate all relevant public and private information. Additionally, the use of these odds as independent variables aids in fulfilling the exogeneity assumption, as they likely incorporate all factors influencing match outcomes, thereby mitigating potential endogeneity. Since heteroscedasticity is resolved by employing WLS estimation, we assume Classic Linear Model (CLM) assumptions required to ensure validity of the estimation to hold. We support this claim based on the existing research (Goddard and Asimakopoulos 2004; Forrest et al. 2005).

Additionally, we expand the existing model that includes all relevant information for our "hot hand" hypothesis. The proposed extended model would look as follows:

$$\begin{aligned} r_{i,j} = \beta_0 + \beta_1 \phi^r_{i,j} + \beta_2 ATF_{i,j} \\ + \beta_3 HTS_{i,j} + \beta_4 ATS_{i,j} + u_{i,j} \end{aligned} \tag{4.3}$$

which includes dummy variable for away favorite distinction and the recent forms of home and visiting teams. We estimate the equation using WLS estimation with the same procedure described as before and we test the joint hypothesis: $(\beta_2 = 0, \beta_3 = 0, \beta_4 = 0)$. Finding the insignificance of these variables should imply that they do not add any relevant information besides those that are already included in bookmaker's odds. This model extension was inspired by Goddard and Asimakopoulos (2004), who offered an expansion on weak-form efficiency tests by incorporating assessed probability $p_{i,j}$ derived from their forecasting model:

$$r_{i,j} = \alpha_r + \beta_r \phi^r_{i,j} + \gamma_r \left(p^r_{i,j} - \phi^r_{i,j} \right) + u_{i,j} \tag{4.4}$$

This model is estimated using ordered probit regression and includes several relevant variables for evaluating the probabilities of match outcomes such as position of the team in a table or result of last recent match. The proposed test was similar: $\alpha_r = 0, \beta_r = 1, \gamma_r = 0$. Unfortunately, based on our dataset collection, we were unable to produce such a complex forecasting model. However, the case whether bookmakers include information on team's recent performances properly is investigated.

To test the weak-form market efficiency and its expansion, we use an F-test, for which the F-statistic is defined as follow:

$$F = \frac{\left(R^2_U - R^2_R\right)/q}{\left(1 - R^2_U\right)/\left(n - k - 1\right)} = \frac{\left(SSR_R - SSR_U\right)/q}{\left(SSR_U\right)/\left(n - k - 1\right)} \tag{4.5}$$

where q is the number of restrictions, n number of data points, k number of independent variables, and R^2_U and R^2_R are the R-squared of the unrestricted and restricted model, respectively. Similarly, SSR_U and SSR_R are the sum of squared residuals for unrestricted and restricted model. Since our sample is large enough, and assuming other CLM assumptions to hold (as described when evaluating WLS estimation), we could consider the F statistic to be asymptotically F distributed with $(q, n - k - 1)$ degrees of freedom. Moreover, it is important to note that to test the joint hypothesis for two model estimated using WLS, we have to estimate them using the same weight computed for the unrestricted model. For our extended model test: $(\beta_2 = 0, \beta_3 = 0, \beta_4 = 0)$, we use formula with R^2. In case of the simple weak-form test of efficiency: $\beta_0 = 0, \beta_1 = 1$, we prefer the formula using SSR as for the restricted model the SSR_R is simply computed as: $\left(r_{i,j} - \phi^r_{i,j}\right)^2$.

Furthermore, for the descriptive measure of accuracy of a probability forecast we follow Forrest et al. (2005). They incorporated the Brier Score (BS) in their analysis, inspired by findings of Brier (1950). In case of bookmaker's implied probabilities for home win, we demonstrate the BS as follow:

$$BS = \sum_{i=1}^{N} \left(HTW_i - \phi_i^{HTW} \right)^2 / N \tag{4.6}$$

where $HTW = 1$ if the match resulted in a home win and 0 otherwise, N is the number of matches and ϕ^{HTW} implied probability set by a bookmaker for a HTW. There are equivalent definitions for draws and away wins. BS is comparable to the mean square error of a set of probability forecasts. BS always lies within the scale 0 to 1; the smaller BS is within this scale, the more accurate the probability forecasts are. The original definition formed by Brier (1950) accounts for multi-category forecasts, which is applicable to our situation as well. The original formulation is in the followed manner:

$$BS = \sum_{j=1}^{R} \sum_{i=1}^{N} \left(r_{i,j} - \phi_{i,j}^{r} \right) / N \tag{4.7}$$

where additional information includes R, the number of possible classes in which the event can fall, in our case three (HTW, D, and ATW). Variables r and ϕ^r were previously described.

4.2 Hot Hand Model

After testing the first hypothesis, where we aim at evidence that important information is included in the odds, in the second hypothesis, we test the accuracy of bettor's predictions. The original hypothesis defined by Paul et al. (2014) states that if bookmakers accurately set betting odds as a forecast of game outcomes, bets on teams should win with the bookmaker's estimated probability. If bettors who believe in the "hot hand" (over-betting teams on winning streaks) participate in the market, accurate betting odds may result in imbalanced betting. Therefore, the percentage of bets on home and visiting teams should provide clear evidence of bettor trust in the "hot hand", e.g. teams on winning streaks should attract a higher percentage of bets and teams on losing streaks should attract a lower percentage of bets.

Therefore, based on Paul et al. (2014) we assemble a model estimated by OLS regression as follow:

$$\begin{aligned} \textit{Percentage bet } HTW_i = \beta_0 + \beta_1 ODD_i^{HTW} + \beta_2 ATF_i \\ + \beta_3 HTS_i + \beta_4 ATS_i + \epsilon_i \end{aligned} \tag{4.8}$$

where the percentage of bets made on the home team in each football game is the dependent variable. Using indicator variables such as odds for the win of a home team, streaks of home and visiting teams of varying lengths, as well as other explanatory variables known to effect betting percentages, such as the distinction of home/visiting favorite, we explain the variation in the percentage of bets on the home team. More precisely, the away favorite dummy variable is included based on the findings of Levitt (2004) who proved that the willingness of bettors to bet on strong visiting teams is not likely captivated by bookmaker's odds due the alleged existence of an imbalanced book. All other variables influencing the proportion of bets placed on the home team are included in the random error term ϵ.

In their original study, Paul et al. (2014) formed an expanded model where they separated winning and losing streaks and expressed them in the form of dummy variables:

$$\begin{aligned} \textit{Percentage bet } HTW_i = \beta_0 + \beta_1 ODD_i^{HTW} + \beta_2 ATF_i \\ + \beta_3 \left(HW \textit{ streak} \right)_i + \beta_4 \left(AW \textit{ streak} \right)_i \\ + \beta_5 \left(HL \textit{ streak} \right)_i + \beta_6 \left(AL \textit{ streak} \right)_i + \epsilon_i \end{aligned} \tag{4.9}$$

However, due to the small number of observations for higher streaks and in order to avoid the possible multicollinearity between many independent variables, we decided to simplify the model by creating just two variables that account for the accumulating streaks that the home and visiting teams have in a given match. Eventually, our goal is to find whether "hot hand" bias generally occurs.

Additionally, we expand Paul et al. (2014) findings by regressing the percentage bets placed on an away team on similar independent variables, although adjusted for a case when an away team is studied, to prove the universal application of the "hot hand" hypothesis:

$$Percentage\ bet\ ATW_i = \beta_0 + \beta_1 ODD_i^{ATW} + \beta_2 HTF_i + \beta_3 HTS_i + \beta_4 ATS_i + \epsilon_i \tag{4.10}$$

Apart from the model used in hypothesis #1, the dependent variable is a continuous variable constrained by the range $(0; 1)$ and therefore the form of heteroscedasticity of the variance is unknown. But since it might be presumed, we permit the random error term's variance to be heteroscedastic, following Paul et al. (2014). We confirm its presence by plotting the fitted values of the model from equation 4.8 on its residuals:

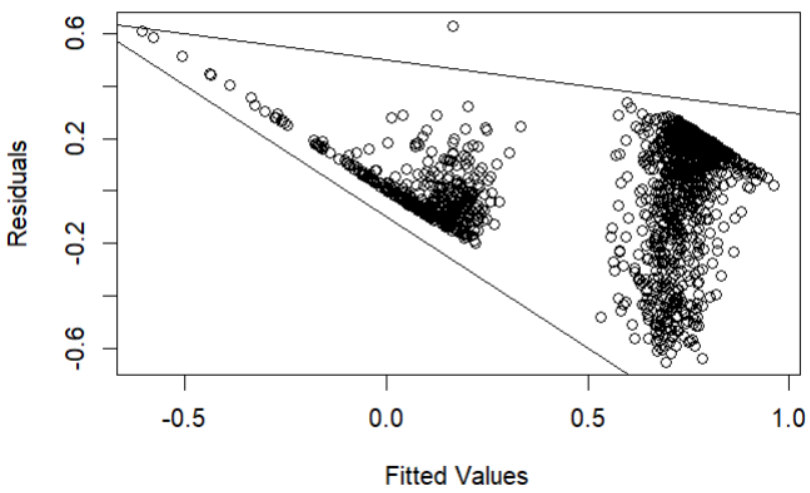

Figure 4.1: Graph plot of fitted values on residuals from equation 4.8

Paul et al. (2014) used the usual White standard errors method to adjust the standard errors for heteroscedasticity. As supported by the graph plot in figure 4.1, we found this correction to be relevant for our analysis as well. Fortunately, thanks to the large sample the robustness of errors should not be diminished in order to achieve efficient estimators of the parameters.

Similarly to the first model, we assume our regression model satisfies the assumption of random sampling by including every game and accounting for influences from previous matches through comprehensive variables, such as bookmaker's odds, which aggregate all relevant public and private information. Moreover, incremental information that possibly could not be reflected in the odds is included by streak variables. Additionally, the use of these independent variables aids in fulfilling the exogeneity assumption, as they likely incorporate all factors influencing match outcomes, thereby mitigating potential endogeneity. Since heteroscedasticity is resolved by incorporating robust standard errors, we assume CLM assumptions required to ensure validity of the estimation to hold. We support this claim based on the existing research (Paul et al. 2014).

4.3 Economic Tests of Profitability

Investigating the inefficiencies in the betting market directly can also be done by figuring out ex post the returns that different betting strategies could have yielded (Goddard and Asimakopoulos 2004). Camerer (1989) discovers evidence in NBA betting market that bets made on teams that are currently winning are more likely to result in losses than wins. Consequently, he suggests a betting strategy that runs against these teams. If we take into account the "hot hand" hypothesis, and given the profitability of bookmakers, the opposite strategy could yield an interesting result. In addition, Woodland and Woodland (2011) discovered consistent evidence in MLB and NHL that, in relation to their chances of winning, bettors tend to overbet favorites and this is found profitable. Following that, we explore the tendency to bet on match favorites. We follow these strategies and investigate whether using these information would result in the realization of profit. However, if the market efficiency hypothesis is not rejected, an attempt to exploit such biases should not be successful. This will be shown in the gross average season returns on CZK 100 bet on match outcome that is predicted by the strategy to be winnable. We assume a bet on each match is made.

Chapter 5

Results and Discussion

5.1 Weak-form Efficiency Tests

In the first section, we present the results of testing hypothesis #1, in which we question whether the assessed probabilities by bookmakers reflect the true probabilities of a match outcome and contain all past relevant information. We tested it on decimal odds provided by betting shop Tipsport for Czech football league (Fortuna liga) matches between seasons 2018/19 and 2022/23.

First, we can have a look at the comparison of probabilities between the implied probabilities set by Tipsport bookmakers and the average results from matches. We follow up the conversion of a decimal odd to implied probability from equation 3.1. In table 5.1, in the first 3 columns we can see the average assessed probability for a home team win, draw and a visiting team win in a match, respectively. In the last 3 columns we could find the average percentage distribution of home team wins, draws and visiting team wins. For demonstration, the probabilities were shown for each season and altogether. In terms of average figures, the distortion of the implied probabilities from actual results does not seem to be considerably high, as except for the season 2018/19, the numbers do not deviate by more than 3 percentage points. Goddard and Asimakopoulos (2004) used a similar descriptive table to demonstrate the differences in assessing probabilities for different seasons and various levels of competitions. Interestingly, he concluded that better forecasts can be made towards the end of the season. However, from our point of view, it is not visibly distinguishable in which season the bookmakers provided the most accurate probabilities.

Interestingly, this feature could be studied through another descriptive mea-

Table 5.1: Comparison of assessed probabilities and actual outcomes

Season	ϕ^{HTW}	ϕ^{D}	ϕ^{ATW}	HTW%	D%	ATW%
All seasons	45.62%	24.95%	29.43%	45.59%	24.44%	29.66%
2022/23	46.04%	24.49%	29.47%	43.12%	24.64%	32.25%
2021/22	45.84%	25.01%	29.15%	47.50%	25.71%	26.79%
2020/21	44.08%	25.20%	30.72%	41.18%	26.47%	32.35%
2019/20	46.11%	24.97%	28.91%	48.33%	25.00%	26.67%
2018/19	46.35%	25.08%	28.57%	50.83%	19.58%	29.58%

sure of accuracy, the Brier Score. Table 5.2 summarizes in the first 3 columns the Brier scores for home team win, draw and away team win, respectively. The scores are computed based on equation 4.6 and presented for each season separately. The absolute Brier Score is featured in the last column and follows the equation 4.7. However, it could be deduced that the absolute score is just a summation of the three previous scores. As the accuracy of predictions is represented by a single number, it is easily comparable. In this manner, we could conclude that for bookmakers the most accurate season was during 2021/22, as the absolute Brier Score is the lowest of all seasons. To support this fact, Brier Scores for home team win and visiting team win are lowest in that season as well.

In the case of a comparison of individual betting opportunities, we could interpret the results in such way that across all seasons, forecast that the home team will win was the least accurate. Nevertheless, all values approximate around 0.2, which indicates reasonable predictions. Forrest et al. (2005) used this measure for comparison of performances of various bookmakers with their built forecasting model, showing significant trends of outperforming the odd setters. We did not implement the measure for this purpose, but to rather provide an affirmation for our weak-form efficiency test regression analysis.

Table 5.2: Brier scores for different seasons

Season	BS HTW	BS D	BS ATW	Absolute BS
All seasons	0.1999	0.1805	0.1815	0.5621
2022/23	0.2118	0.1822	0.2027	0.5968
2021/22	0.1855	0.1824	0.1658	0.5333
2020/21	0.1928	0.1923	0.1876	0.5728
2019/20	0.2009	0.1838	0.1728	0.5575
2018/19	0.2110	0.1581	0.1767	0.5459

Next, we follow up on Goddard and Asimakopoulos (2004) with their weak-form efficiency model. Tables 5.3. and 5.4. present the WLS regression results of equations 4.1 and 4.3 respectively. Estimates of the former equation are depicted in the first table. As it continues, the three columns represent the main opportunities: the home team win, draw and visiting team win. When interpreting intercepts of the estimated models, we can conclude that only for the home team win, it significantly differs from 0. In case of the implied probability variable for individual models, all estimates are significantly different from 0. However, if we apply similar student's t-test for $H_0 : \beta_1 = 1$, again only β_1 for implied probability of a home team win will be significantly different from 1. Third row represents the computed F-statistics for the null hypothesis, that concerns the simple weak-form efficiency test.

Table 5.3: Weak-form efficiency: regression-based tests (Part 1)

	HTW		D		ATW
Constant	-0.0649^{***}	Constant	-0.0591	Constant	-0.0021
	(0.0186)		(0.0422)		(0.0207)
ϕ_i^{HTW}	1.1534^{***}	ϕ_i^{D}	1.21^{***}	ϕ_i^{ATW}	1.0056^{***}
	(0.0408)		(0.1732)		(0.0661)
F_1	3.88^{**}	F_1	0.90	F_1	0.01
Observations	1342		1342		1342
R^2	0.374		0.036		0.147
Adjusted R^2	0.374		0.035		0.146

Notes: Standard errors of estimated coefficients are shown in parentheses.
Tests based on equation 4.3
F_1 is an F-test for $H_0 : \{\beta_0, \beta_1\} = \{0, 1\}$
$^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$

As the individual F-tests could have indicated, we do not have enough evidence to reject the null hypothesis for draws and visiting team wins, as the F-statistics are very low, leading to a very high p-value for the null hypothesis rejection. Therefore, we can assume implied probabilities of those opportunities to be weak-form efficient. In case of a home team win the F-statistic 3.88 implies that at 5% level of significance, we would have rejected the null hypothesis. However, at 1% the null hypothesis is not rejected. Hence, the weak-form efficiency could be considered for the home team win implied probabilities as well. These findings are important for our further analysis. As of now, we could assume that there is no bias present in the bookmaker's odds, which is consistent with our hypothesis #2.

We continue to investigate the presence of biases in the implied probability odds in equation 4.3., whose estimates are provided in table 5.4. Apparently, the significance of individual estimates for intercepts and implied probabilities stays the same. Nonetheless, parameters for variables, which represent possible biases in the implied probabilities, do not significantly differ from zero. What is more, in the manner of joint F-tests, none of the tests have a strong F-statistic, supporting the previous claim. On top of that, comparisons of the R^2 and adjusted R^2 from the previous tables signal no major information added to the models. As a result, we can conclude that the implied odds contain past relevant information including the level of winning and losing streaks of teams. Consequently, with these additional tests, the implied odds satisfy the weak-form efficiency hypothesis.

Table 5.4: Weak-form efficiency: regression-based tests (Part 2)

	HTW		D		ATW
Constant	-0.1102^{**}	Constant	-0.0537	Constant	-0.0210
	(0.0475)		(0.1861)		(0.0261)
ϕ_i^{HTW}	1.2415^{***}	ϕ_i^{D}	1.1441^{***}	ϕ_i^{ATW}	1.0496^{***}
	(0.0723)		(0.1861)		(0.1192)
ATF	0.0309	ATF	0.0372	ATF	-0.0059
	(0.0380)		(0.0273)		(0.0434)
HTS	0.0003	HTS	-0.0008	HTS	-0.0018
	(0.0036)		(0.0037)		(0.0037)
ATS	0.0061	ATS	0.0007	ATS	-0.0054
	(0.0036)		(0.0035)		(0.0035)
F_2	0.82	F_2	0.49	F_2	0.58
Observations	1342		1342		1342
R^2	0.386		0.038		0.153
Adjusted R^2	0.384		0.036		0.151

Notes: Standard errors of estimated coefficients are shown in parentheses.
Tests based on equation 4.3
F_2 is an F-test for $H_0 : \{\beta_2, \beta_3, \beta_4\} = \{0, 0, 0\}$.
$^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$

Our findings partially coincide with Goddard and Asimakopoulos (2004), who found English football betting market to be weak-form efficient. Nevertheless, they identified evidence of inefficiency in the last period of the season, contrary to the overall efficiency. Robbins (2023) comes with an additional support for existence of weak-form efficiency, as for four major leagues in the U.S. studied, only NBA betting market was found to contain odds bias.

Summary

According to our studied sample, the Czech football betting market is found to be weak-form efficient and therefore odds are not subject to studied "hot hand" bias. This conclusion is backed by similar evidence from English and American betting market.

5.2 Hot Hand Model Results

This section builds on the Paul et al. (2014) model from equation 4.8. In table 5.5, we summarize the estimates of the regression. In the first column, we estimated the model on a dataset including all seasons. Intercept together with road favorite dummy feature very strong significance, and their values 0.82 and 0.49 respectively, indicate strong preference of bettors on match favorites. To interpret, if the visiting team is determined to be a favorite of the match, the percentage of bettors who would bet on a home would drop by more than 49 percentage points. This fact is consistent with Levitt (2004), who pointed out the possible bettor's preference of teams that declared a favorite while being a visiting team. Odd placed on a home team by a bookmaker shows strong significance as well. With an additional increase in odds by 1, the percentage of bettors preferring a home team drops by approximately 4 percentage points. Note that an increase in odds implies a drop in the probability of winning assessed by a bookmaker, which is in line with the resulting sign of the estimator.

Streaks of home and visiting teams present a strong significance in both cases. If we elaborate further, the numerical value for home team streak indicates that with an additional win in a row, the percentage of bettors betting on a home team increases by 1.5 percentage points, while additional loss or draw in a row represents a 1.5 percentage point decrease in bettor's preferences. Moreover, visiting's team form influences the bettor's preferences as well, as an additional win in a row represents a 1 percentage point decrease in

Table 5.5: Estimates of the hot hand model (Part 1)

	Dependent variable Percentage of Bets Placed on Home Team		
	All seasons	2018/19-2020/21	2021/22-2022/23
Constant	0.8169***	0.8086***	0.8354***
	(0.0142)	(0.0173)	(0.0252)
ODD^{HTW}	−0.0415***	−0.0369***	−0.0526***
	(0.0065)	(0.0074)	(0.0124)
ATF	−0.4918***	−0.4837***	−0.4884***
	(0.0188)	(0.0238)	(0.0311)
HTS	0.0154***	0.0185***	0.0120***
	(0.0018)	(0.0023)	(0.0028)
ATS	−0.0104***	−0.0140***	−0.0053**
	(0.0017)	(0.0022)	(0.0027)
Observations	1327	775	552

Notes: Standard errors of estimated coefficients are shown in parentheses.
$^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$

the percentage of bettors betting on a home, and 1 percentage point increase otherwise.

To summarize, bettors on a home team do not only consider recent performance of both engaged teams, but they also give more weight to teams on longer streaks. Altogether, it gives us evidence of prevailing hot hand bias in bettors preferences, supporting hypothesis #2. This evidence is consistent with the findings by Paul et al. (2014) who came up with two possible interpretations. One possible explanation is that bettors overestimate the degree to which the "hot hand" effect influences team performance, although there is a real "hot hand" effect that bookmakers appropriately reflect in odds. Another scenario is that both bookmakers and bettors mistakenly assume there are "hot hand" effects, but that bettors adhere to the myth more than bookmakers do. As we incorporated this investigation into the previous section, we could claim that the former statement holds.

In addition to that, we investigated separate periods. A new format was put into action between seasons 2020/21 and 2021/22, which brought more matches between balanced teams. Thus, we decided to further analyze the influence of the new establishment of a new format on the results of the estimation.

Columns two and three in table 5.5 represent the regressions of periods before the establishment of the format and after, respectively. The results indicate that bettors more emphasized team streaks before the establishment, as the estimates for the home team and visiting team streaks differ by 0.65 and 0.85 percentage points, respectively. Furthermore, estimates of the streaks after the establishment register drop in significance, although not substantial, as the ATS variable is statistically different from zero only at the 5% significance level. These findings might imply that the increased number of even matches may mean that bettors focus more on other aspects that determine their decision for a winning bet. Nevertheless, we could conclude that the "hot hand" effect is still prevalent.

Lastly, we comment on the regression of percentage bets placed on an away team:

Table 5.6: Estimates of the hot hand model (Part 2)

	Dependent variable Percentage of Bets Placed on Visiting Team
	All seasons
Constant	0.7377^{***}
	(0.0125)
ODD^{ATW}	-0.0094^{***}
	(0.0011)
HFD	-0.5651^{***}
	(0.0141)
HTS	-0.0081^{***}
	(0.0015)
ATS	0.0135^{***}
	(0.0015)
Observations	1327

Notes: Standard errors of estimated coefficients are shown in parentheses.
$^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$

It can be deduced that signs in the regression are interpreted correctly. Higher odds decrease the preference of bettors, as well as the situation when the home team is declared a favorite in a particular match. As in the regression from table 5.5, the strong preference of match favorites is evident. Moreover, positive home team streak lowers the preferences for a visiting team and a positive away team streak in turn has a positive effect on the dependent variable. In addition to that, all estimates are strongly significant which only strengthens our initial hypothesis and supports our claim about the complexity of "hot hand" hypothesis.

5.3 Test of Profitability Results

In the last section, we take a closer look at the profitable strategies investigated by Camerer (1989) and Woodland and Woodland (2011). Table 5.7 presents average returns for both implemented strategies in separate seasons and through the whole five seasons span per CZK 100 bet. For Camerer (1989) strategy, we find only season 2022/23 to generate positive returns, while Woodland and Woodland (2011) strategy reached profitability in the season before. Overall, we do not find either of the strategies consistently profitable, which is in contradiction with the evidence of Camerer (1989) and Woodland and Woodland (2011) in NBA, NHL, and MLB betting market, respectively. These observations may imply that there is no persistent bias incorporated in the odds that could be exploited by bettors. This is consistent with our evidence of weak-form efficiency.

Table 5.7: Performance of Camerer (1989) and Woodland and Woodland (2011) strategies

Season	Bets placed against teams on better streaks	Bets placed on match favourites
All seasons	−7.3	−2.6
2022/23	+16.2	−8.1
2021/22	−17.4	+2.8
2020/21	−16.0	−2.2
2019/20	−10.9	−5.0
2018/19	−7.8	−0.7

Note: Data are average pre-tax (gross) returns per CZK 100 bet.

Summary

With further investigation, we found evidence that bettors in the Czech football betting market are influenced by a hot hand belief, as it was projected in the distribution of bets. However, strategies proposed to battle these biases are not found to be profitable either.

Limitations

Concluding the discussion on the limitations of this thesis, it is evident that while the methodology employed and data utilized provide valuable insights into the efficiency of betting markets, they carry constraints that must be acknowledged. Although practical, using bookmaker's odds as a stand-in for market expectations presents biases that could distort perceptions of market efficiency. These odds may not purely reflect collective market beliefs but are as well influenced by the strategies of bookmakers aiming to balance books or manipulate market behaviors, as mentioned in the introduction chapter. Additionally, the exclusion of certain unpredictable factors such as player psychology, precise match conditions, and tactical changes emphasize a significant gap in the model's ability to incorporate all variables that influence match outcomes and betting behaviors.

Furthermore, the methodological limitations, particularly concerning endogeneity and the violation of regression assumptions, suggest caution in interpreting the results as a definitive evidence of market efficiency. The model's potential sensitivity to omitted variable bias and the dynamic nature of betting markets imply that findings could be rather specific to the dataset and model specifications used. Finally, limitations of this study are demonstrated by its dependence on static data, which does not capture real-time odds changes or provide deep market analytics, thus restricting a thorough understanding of the nuances in betting behavior. Nevertheless, it is important to recognize that these limitations, though notable, are inherently difficult to overcome and do not severely undermine the validity of the findings.

Chapter 6

Conclusion

This thesis examines the presence of "hot hand" bias in the Czech football betting market between years 2018 and 2023. Using WLS estimation (Goddard and Asimakopoulos 2004), we test the efficiency of bookmaker's side of market, which contradicts the presence of the bias. Consequently, the bias is explored on the side of bettors. For these purposes, we use data that include betting odds and bettor's distribution of odds for all matches of Czech football league from seasons 2018/19 to 2022/23.

After considering our findings for hypothesis #1, consisting of Brier Score evaluation and weak-form tests, we can conclude that we believe the Czech football betting market to be efficient. Therefore, all past relevant is included in the betting odds. We supported the hypothesis using the fact that we found no evidence of presence of "hot hand" bias that would be included in the bookmaker's odds. Thus, when variables indicating home and visiting team's recent performances were included in the model, they were determined statistically insignificant. This follows our hypothesis that bookmakers tend to reach profitability by being the most precise and consistent in terms of forecasting outcomes.

To elaborate on that, the result of this paper adds to the confirmation of the efficient betting market hypotheses for point-spread bets (Pankoff 1968), money-line bets (Robbins 2023) and fractional bets (Goddard and Asimakopoulos 2004). However, as pointed out by Sauer (1998), since efficient prices reflect features that a particular model implies, they depend on the information structure, behavioral claims, and limitations that define the model. Hence, it is better to interpret the market efficiency hypothesis as a highly useful benchmark.

Next, we look closely at the bettor's drivers of betting preferences. The bettor's side of market would be considered efficient if the distribution of bets perfectly reflects the bookmaker's odds (Paul et al. 2014). Nevertheless, our findings lead to a different conclusion. As expected, when a home team was accumulating streak of games in which they won, bettors have expressed overwhelmed confidence in that team although this information has already been accounted for by the bookmakers. This outcome was already shown in the previous regression. Similarly, if the home team did not win several matches in a row, it was reflected in the declining home team preference by the bettors. The recent performance of the visiting team, which influences the bettors betting on the home team, was found to be statistically significant as well, with opposite effects to be precise.

This altogether reaffirms our hypothesis #2 that states that bettors overreact to team performances regarding the "hot hand" belief, which may be one of the drivers of increase in profits for betting shops. In addition to that, we rejected the Camerer (1989) and Woodland and Woodland (2011) betting strategies. They acknowledged those biases and suggested betting against those beliefs; however, we found those strategies unprofitable. This outcome might partly result from the former proven market efficiency hypothesis. As we elaborated on Paul et al. (2014) findings, we proved that a similar adjusted model is applicable not only to a point-spread betting system, but to European decimal odd-setting system as well.

To conclude, we demonstrated the evaluation of efficiency of the Czech football betting market for both sides consisting in bookmakers and the bettors. The market seems to behave according to our predictions, as bookmaker's side appears to be efficient, while bettor's side expresses persisting bias highlighting teams on streaks. This thesis gives some of the numerous new areas that still need to be investigated with some encouragement. Other sports, which are subject to betting in Czech betting market, represent a new matter that is open to further investigation. To expand, more exhaustive research of market efficiency in the European decimal odds betting system would be appreciated in market efficiency literature. In addition, there is a great scope for exploring other behavioral biases that are harder to measure but can potentially have a large impact on bettor's decision making. Such variables could be used to expand our model.

Bibliography

Andersson, P. and H. Nilsson (2015). Do bettors correctly perceive odds? Three studies of how bettors interpret betting odds as probabilistic information. *Journal of Behavioral Decision Making 28*(4), 331–346.

Athletic, T. (2022). How to read decimal odds in sports betting. https://theathletic.com/2544857/2022/01/27/how-to-read-decimal-odds-in-sports-betting/.

Ayton, P. and I. Fischer (2004). The hot hand fallacy and the gambler's fallacy: Two faces of subjective randomness? *Memory & Cognition 32*, 1369–1378.

Brier, G. W. (1950). Verification of forecasts expressed in terms of probability. *Monthly Weather Review 78*(1), 1–3.

Brown, W. O. and R. D. Sauer (1993). Does the basketball market believe in the hot hand? Comment. *The American Economic Review 83*(5), 1377–1386.

Cain, M., D. Law, and D. Peel (2000). The favourite-longshot bias and market efficiency in UK football betting. *Scottish Journal of Political Economy 47*(1), 25–36.

Camerer, C. F. (1989). Does the basketball market believe in the "hot hand"? *The American Economic Review 79*(5), 1257–1261.

Clotfelter, C. T. and P. J. Cook (1991). *Selling hope: State lotteries in America.* Harvard University Press.

Durand, R. B., F. M. Patterson, and C. A. Shank (2021). Behavioral biases in the NFL gambling market: Overreaction to news and the recency bias. *Journal of Behavioral and Experimental Finance 31*, 100522.

Fama, E. F. (1965). The behavior of stock-market prices. *The Journal of Business 38*(1), 34–105.

Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. *The Journal of Finance 25*(2), 383–417.

Forbes (2018). How casinos use math to make money when you play the slots. https://www.forbes.com/sites/davidschwartz/2018/06/04/how-casinos-use-math-to-make-money-when-you-play-the-slots/.

Forbes (2024). What does the moneyline mean in sports betting? https://www.forbes.com/betting/guide/what-does-moneyline-mean/.

Forrest, D., J. Goddard, and R. Simmons (2005). Odds-setters as forecasters: The case of English football. *International Journal of Forecasting 21*(3), 551–564.

Forrest, D. and R. Simmons (2013). Sentiment in the betting market on Spanish football. In *Economics of Betting Markets*, pp. 118–126. Routledge.

Fortunaliga.cz (2024). Rozpis zápasů. https://www.fortunaliga.cz/rozpis-zapasu/2024?id_stage1&month=0&round=0&type=1.

Gilovich, T., R. Vallone, and A. Tversky (1985). The hot hand in basketball: On the misperception of random sequences. *Cognitive Psychology 17*(3), 295–314.

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

Graham, I. and H. Stott (2013). Predicting bookmaker odds and efficiency for UK football. In *Economics of Betting Markets*, pp. 99–109. Routledge.

Griffith, R. M. (1949). Odds adjustments by American horse-race bettors. *The American Journal of Psychology 62*(2), 290–294.

Hendricks, D., J. Patel, and R. Zeckhauser (1993). Hot hands in mutual funds: Short-run persistence of relative performance, 1974–1988. *The Journal of Finance 48*(1), 93–130.

Huber, J., M. Kirchler, and T. Stöckl (2010). The hot hand belief and the gambler's fallacy in investment decisions under risk. *Theory and Decision 68*, 445–462.

Kahneman, D. and A. Tversky (1972). Subjective probability: A judgment of representativeness. *Cognitive Psychology 3*(3), 430–454.

Kahneman, D. and A. Tversky (1979). Prospect theory: An analysis of decision under risk. *Econometrica 47*(2), 263–292.

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

Losak, J. M., A. P. Weinbach, and R. J. Paul (2023). Behavioral biases in daily fantasy baseball: The case of the hot hand. *Journal of Sports Economics 24*(3), 374–401.

Miller, J. B., A. Sanjurjo, et al. (2014). A cold shower for the hot hand fallacy. Technical report, Working paper.

Novinky.cz (2023). Lidé v Česku loni prosázeli přes 750 miliard, meziročně o 44,2 procenta více. https://www.novinky.cz/clanek/ekonomika-lide-v-cesku-loni-prosazeli-pres-750-miliard-mezirocne-o-442-procenta-vice-40438313.

Ötting, M., R. Langrock, C. Deutscher, and V. Leos-Barajas (2020). The hot hand in professional darts. *Journal of the Royal Statistical Society Series A: Statistics in Society 183*(2), 565–580.

Pankoff, L. D. (1968). Market efficiency and football betting. *The Journal of Business 41*(2), 203–214.

Paul, R. J. and A. P. Weinbach (2005). Bettor misperceptions in the NBA: The overbetting of large favorites and the “hot hand”. *Journal of Sports Economics 6*(4), 390–400.

Paul, R. J., A. P. Weinbach, and B. Humphreys (2014). Bettor belief in the “hot hand”: Evidence from detailed betting data on the NFL. *Journal of Sports Economics 15*(6), 636–649.

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

Quandt, R. E. (1986). Betting and equilibrium. *The Quarterly Journal of Economics 101*(1), 201–207.

Rabin, M. (2002). Inference by believers in the law of small numbers. *The Quarterly Journal of Economics 117*(3), 775–816.

Robbins, T. R. (2023). Weak form efficiency in sports betting markets. *American Journal of Management 23*(2).

Rossi, M. (2011). Match rigging and the favorite long-shot bias in the italian football betting market. *International Journal of Sport Finance 6*(4), 317–334.

Sauer, R. D. (1998). The economics of wagering markets. *Journal of Economic Literature 36*(4), 2021–2064.

Sauer, R. D., V. Brajer, S. P. Ferris, and M. W. Marr (1988). Hold your bets: Another look at the efficiency of the gambling market for national football league games. *Journal of Political Economy 96*(1), 206–213.

Sirri, E. R. and P. Tufano (1998). Costly search and mutual fund flows. *The Journal of Finance 53*(5), 1589–1622.

Teall, J. L. (2022). *Financial trading and investing.* Academic Press.

Woodland, L. M. and B. M. Woodland (1994). Market efficiency and the favorite-longshot bias: The baseball betting market. *The Journal of Finance 49*(1), 269–279.

Woodland, L. M. and B. M. Woodland (2001). Market efficiency and profitable wagering in the NHL: Can bettors score on longshots? *Southern Economic Journal 67*(4), 983–995.

Woodland, L. M. and B. M. Woodland (2003). The reverse favourite–longshot bias and market efficiency in MLB: An update. *Bulletin of Economic Research 55*(2), 113–123.

Woodland, L. M. and B. M. Woodland (2011). The reverse favorite-longshot bias in the NHL: Do bettors still score on longshots? *Journal of Sports Economics 12*(1), 106–117.

Ziemba, W. T. and R. H. Thaler (1988). Parimutuel betting markets: Racetracks and lotteries. *Journal of Economic Perspective 2*(2), 161–174.

Zuber, R. A., J. M. Gandar, and B. D. Bowers (1985). Beating the spread: Testing the efficiency of the gambling market for NFL games. *Journal of Political Economy 93*(4), 800–806.

Appendix A

Description of Fortuna Liga Playing Format Through Seasons 2018/19 to 2022/23

The playing format has undergone multiple adjustments during these seasons. The initial format concerned 16 teams playing each other home and away, accounting for 30 matches per team. However, a new format, which divides the league into three groups after the regular season, was established and planned to be implemented after season 2019/20. The COVID-19 pandemic disrupted the 2019/20 season, which was later completed using the traditional 30-game regular season schedule. Consequently, next season it was decided that league would be extended to 18 teams instead of the original number of 16 teams, with three teams relegating and only one team promoting from the second-tier national football league. Seasons 2021/22 and 20222/23 were played under the newly established format and after the regular stage of 30 games the first 6 teams competed for the title with additional matches between each other, middle 4 teams fought for the chance to get into the Europe competitions, and the last 6 teams competed for not being relegated.