Aalto University
School of Business

When do betting odds best represent the actual outcomes?

Predicting NHL results based on moneyline odds movement

Master's Thesis
Henri Lahtinen
Aalto University School of Business
Information and Service Management
2019

Aalto University School of Business

Aalto-yliopisto, PL 11000, 00076 AALTO
www.aalto.fi

Maisterintutkinnon tutkielman tiivistelmä

Author Henri Lahtinen

Title of thesis When do betting odds best represent the actual outcomes? Predicting NHL results based on moneyline odds movement

Degree Master of Science (M.Sc.)

Degree programme Information and Service Management

Thesis advisor(s) Tomi Seppälä

Year of approval 2019 **Number of pages** 75 **Language** English

Abstract

The sports betting market has grown quickly over the past the few decades mostly due to the digitalization of the business. The bookmakers have moved online from the physical locations. The market has thus globalized, and the competition has increased, forcing the bookmakers to produce more accurate estimates about the events to keep up with the competition.

This study investigates the accuracy of NHL moneyline betting odds, in predicting the actual outcomes of games throughout the time that the betting events are open. The dataset covers three full seasons from 2015 to 2018. The odds are collected at 5 different time points for each game and the differences in the predictive power of time points is analyzed. Then, the odds and their movement is investigated further to see if there are profitable betting strategies to be found based solely on information about odds movement.

The tests related to the prediction accuracy of the odds reveal that there are no statistically significant differences between the prediction accuracy for different time points. Aggregate results are rather consistent though in showing that before the day of the game, the estimates implied by the odds aren't quite as accurate as they are during the game day. The regression tests further indicate that when the implied probability of a selection grows, the objective probability grows at a higher rate, meaning that there's a Favorite-Longshot Bias in the market. This means that betting on a more likely outcome yields better returns on average.

The tests for finding profitable betting strategies further enforce the notion of Favorite-Longshot Bias and subsequently all of the consistently profitable betting strategies, that are found, involve only betting on favorites. The data about the odds movement between time points and splitting the teams to favorites and underdogs reveals that betting on favorite teams who have had their odds rising in a given time point interval, yields a profit for 80% of the intervals. The margins are so small that none of the returns for profitable strategies are significantly larger than zero in a statistical sense but the difference to the average bookmaker margin is more significant. According to the analysis about different staking strategies, for this kind of betting system where no estimate is calculated individually for each game, simple staking strategies of betting a fixed amount or to win fixed amount, yielded the best balance of capital growth and risk.

The study concludes that there is little difference in predictive power of NHL moneyline betting odds at different time points throughout the life cycle of betting events. Based on the results it's clear that the bookmaker margin isn't allocated evenly between favorites and underdogs and there's an apparent Favorite-Longshot Bias in the market, which is in contrast with previous research about NHL moneyline odds. The bias logically leads to favorites being the side that offers better returns and betting on favorites with rising odds offers returns that consistently beat the bookmaker margin and are also marginally profitable.

Keywords NHL, sports betting markets, odds movement, betting strategy, favourite-longshot bias

Aalto University School of Business

Aalto-yliopisto, PL 11000, 00076 AALTO
www.aalto.fi

Maisterintutkinnon tutkielman tiivistelmä

Tekijä Henri Lahtinen

Työn nimi When do betting odds best represent the actual outcomes? Predicting NHL results based on moneyline odds movement

Tutkinto Kauppatieteiden maisteri (KTM)

Koulutusohjelma Informaatio- ja palvelujohtaminen

Työn ohjaaja(t) Tomi Seppälä

Hyväksymisvuosi 2019 **Sivumäärä** 75 **Kieli** englanti

Tiivistelmä
Urheiluvedonlyöntimarkkinat ovat kasvaneet huomattavasti viime vuosikymmeninä pääasiassa alan digitalisoitumisen ansiosta. Vedonlyöntiyhtiöt ovat siirtyneet kivijalkakioskeista toimimaan internetissä. Markkina on tämän myötä globalisoitunut ja kilpailu kiristynyt, pakottaen vedonlyöntiyhtiöt tuottamaan tarkempia arvioita urheilutapahtumista pysyäkseen kilpailussa mukana.

Tämä tutkielma tarkastelee NHL:n moneyline-vedonlyöntikertoimien tarkkuutta otteluiden lopputuloksia ennustettaessa läpi koko vedonlyöntikohteiden aukioloajan. Aineisto kattaa kolme täyttä NHL-kautta vuosilta 2015-2018. Kertoimet kerätään viitenä ajankohtana jokaiselle ottelulle ja analysoidaan kertoimien kykyä ennustaa lopputulosta eri ajankohtina. Kertoimia ja niiden muutoksia tutkitaan sen jälkeen tarkemmin ja testataan, onko mahdollista löytää voitollisia vedonlyöntistrategioita, jotka pohjautuvat pelkästään mainittuihin muuttujiin.

Ennustustarkkuuteen liittyvät testit paljastavat, että eri ajankohtien välillä ei ole tilastollisesti merkittäviä eroja. Testit kuitenkin melko johdonmukaisesti osoittavat, että otteluiden avauskertoimet ennustavat lopputuloksia hieman heikommin kuin ottelupäivän kertoimet. Regressiotestit näyttävät lisäksi, että kaikissa ajankohdissa, kun kertoimien osoittama subjektiivinen todennäköisyys kasvaa, todellisen lopputuloksen objektiivinen todennäköisyys kasvaa myös, mutta suhteessa enemmän, tarkoittaen että markkinoilla vallitsee suosikki-altavastaaja-harha. Tällöin vedon lyöminen suuremman todennäköisyyden kohteesta tuottaa keskimäärin paremmin.

Testit voitollisten vedonlyöntistrategioiden löytämiseksi antavat lisätukea löydökselle suosikki-altavastaaja-harhan olemassaolosta ja kaikki löytyneet, säännöllisesti voitolliset, strategiat pohjautuvatkin vetojen lyömiseen suosikkien puolesta. Aineisto kerroinmuutoksista ajankohtien välillä ja aineiston jaotteleminen suosikkeihin ja altavastaajiin paljastaa, että vetojen lyöminen sellaisten suosikkien puolesta, joiden kerroin kasvaa tietyllä ajankohtien välisellä intervallilla, tuottaa voittoa 80% intervalleista. Saavutetut tuottomarginaalit ovat niin pieniä, että voitollisten strategioiden tuotot eivät ole tässä otoksessa tilastollisesti suurempia kuin nolla. Kun ottaa huomioon vedonlyöntiyhtiön marginaalin, ovat tuotot kuitenkin merkittävämmin keskimääräistä suurempia. Analyysi eri panostusstrategioista osoittaa, että yksinkertaiset panostusstrategiat (esim. tasapanostus tai tasavoittopanostus) toimivat parhaiten tässä testattuun vedonlyöntijärjestelmään, jossa todennäköisyysarviota ja vedon odotusarvoa ei lasketa erikseen joka ottelulle.

Tutkielman perusteella voidaan sanoa, että NHL:n moneyline-kertoimien ennustustarkkuudessa on hyvin vähän eroa ajankohtien välillä läpi kohteiden aukioloajan. Tulosten perusteella on selvää, että vedonlyöntiyhtiöiden tuottomarginaali on epätasaisesti jaettu suosikeiden ja altavastaajien välillä ja markkinoilla on suosikki-altavastaaja-harha, mikä on päinvastainen löydös verrattuna aiempiin tutkimuksiin NHL:n moneyline-markkinoista. Harha loogisesti johtaa siihen, että suosikkien lyöminen on vedonlyönnillisesti tuottavampaa kuin altavastaajien. Suosikit, joiden kertoimet ovat nousseet, palauttavat paremmin säännönmukaisesti paremmin kuin vedonlyöntiyhtiön tuottomarginaalin verran ja ovat jopa marginaalisesti voitollisia.

Avainsanat NHL, urheiluvedonlyöntimarkkinat, kerroinmuutokset, vedonlyöntistrategia, suosikki-altavastaaja-harha

Table of Contents

List of tables

List of figures

1. Introduction

The sports betting market has grown quickly over the past few decades mostly due to the digitalization of the business. The bookmakers have moved online from the physical locations. The market has thus globalized and the competition has increased, forcing the bookmakers to produce more accurate estimates about the events to keep up with the competition.

The increased competition and the bettors' access to several bookmakers has decreased the amount of discrepancy in odds provided by different bookmakers. Also, the bookmaker companies have begun competing with the amount of bookmaker margin included in the odds. Traditionally the margin for events with even odds has been 4.77%, i.e. the odds for two events with equal perceived probabilities has been 1.909 (betting 1 unit of money yields a profit of 0.909 units if the bet wins). Now, several bookmakers have decreased their margins by about a half and the margins for e.g. spread markets of US major sports are between 2-3% at the most popular bookmakers. The smaller the margin of the bookmaker is, the less there is margin for error in the odds setting. The closing odds for some companies that offer small margins and accept high stakes without restricting winning players, can be considered to be very representative of the true probabilities of the event outcomes. The odds can move though, sometimes significantly, between when they're released by the bookmaker and by the starting time of the event when the betting event is closed. For pari-mutuel betting markets, e.g. horse racing, there's research confirming that the closing odds are clearly the most representative of the true probabilities due to large volume of so-called sharp bettors placing their bets at the very last minute before closing of the betting event. For fixed-odds betting markets, e.g. NHL moneyline bets, limited research exists though. It's interesting to find out if it's possible to find profitable betting strategies that involve only reading the movement of the odds.

1.1. Purpose and contribution

The purpose of this study is to investigate when the odds for NHL matches are the most representative of the true probabilities of the outcomes of the games and to investigate if there are profitable betting strategies to be exploited only based on the movement of the odds. I collect moneyline odds data for 3988 NHL matches from three seasons (2015-2018), including the odds for both home and away teams to win. Moneyline means betting on the ultimate winner of the match, as in NHL (and in hockey generally), if the regular time ends in a draw, there will be an overtime and possibly a penalty shootout to always determine a winner for the match. The odds are collected at 5 different time points for both the home and away team for every match. These time points are the opening of the betting event, 12 AM Eastern time on game day, 1 hour before the scheduled start of the game, 15 minutes before the scheduled start of the game and the closing of the betting event.

The part researching the representativeness of the odds aims to determine at which of the 5 time points, the odds most accurately represent the true outcomes of the matches.

The part investigating the possible betting strategies, analyzes the movement of the odds and utilizes statistical analysis to discover if there are profitable betting strategies to be exploited using only information about the odds and their movement.

There's very little previous research available for fixed-odds betting markets about the movement of the odds and about when the odds most accurately represent the true probabilities of the outcomes of the matches. For pari-mutuel betting markets, e.g. horse racing, this kind of research exists.

1.2. Motivation

The motivation for this study is that there's very little research available about the movement of the odds for fixed-odds betting markets.

I find the topic highly interesting and relevant personally. I've been thinking about the optimal time to place bets for NHL matches and haven't found any actual research done on the topic. Also, as the odds sometimes move significantly between the opening and closing of a betting event, it's interesting to see if there are betting strategies to be exploited purely based on the movement of the odds.

1.3. Main findings

The statistical tests on the implied probabilities at different time points clearly indicate, that there are no statistically significant differences between the time points in how well the odds represent the actual outcomes. This implies that throughout the life cycle of the betting events, the odds quite equally represent the information that is available at each moment. Although there are no significant differences, nearly all of tests indicate that the opening odds are the worst at representing the actual outcomes.

When assessing the profitability of betting strategies related to odds and their movement, it becomes very clear that betting on favorites yields significantly better results than betting on underdogs, which means that there's a Favorite-Longshot Bias (FLB) present in the market, which is a very common phenomenon for betting markets in various sports. It means that as the implied probability of a win gets higher, the profitability of the bet increases, i.e. the objective probability grows also but at a higher rate. This is an interesting finding as it's completely opposite compared to research done by Woodland and Woodland (2001), Gandar et al. (2004) and Paul and Weinbach (2012) that have all found that NHL is a market where there's actually a Reverse Favorite-Longshot Bias present (RFLB), meaning that on average underdogs yield better returns than favorites.

The highest returns based on odds movement in this study were yielded by betting on favorites that had their odds go up, which showed a profit for 8 out of 10 time intervals. The highest profit was found for rising favorite odds from 12 AM of the game day to the closing of the betting events. This interval yielded a profit of 3.17%. None of the profit figures were statistically higher than zero though but 6 of the 8 profitable time intervals were statistically significant at 10% level compared to the average bookmaker margin of -2.4% built in the odds, which is a result that should be achieved in the long run by making random selections to bet on.

1.4. Limitations

There are a few limitations to this study. First of all, the study assumes that the bookmaker margin has been allocated evenly on both sides of the betting event. This is the only way to calculate the implied probabilities in a consistent way. There's no way of actually knowing though how bookmakers allocate the margin between different outcomes.

Second, it's assumed that the bettor can always place the bet on the closing odds. This is relevant especially for the part researching betting strategies based on odds movement as several of the strategies involve tracking odds movement from one time point to the closing of the betting events. In reality, placing a bet at closing odds of Pinnacle, the bookmaker whose odds are used for this research, isn't always possible. The NHL betting events for Pinnacle are closed right at the time of the opening faceoff which is usually about 8 minutes after the scheduled starting time of the match for standard regular season matches. Without watching the match in real time, it's impossible to estimate exactly when the betting event will be closed, causing the bettor to either risk for the line to move after placing the bet or missing the closing of the market. The line movements in the very final minutes before the closing of the betting events are mostly minor though.

1.5. Structure

The thesis is organized as follows. The next chapter provides the necessary fundamentals of sports betting in general. Chapter 3 is a review of previous literature of using betting odds and their movement as a predictor of outcomes for sporting events and about the efficiency of the betting market. At the end of Chapter 3, also the hypotheses for the empirical part of the study are presented. Chapter 4 introduces the data, methodology, and assumptions employed in the study. The results of the study are presented in Chapter 5 and discussed further in Chapter 6. Finally, Chapter 7 concludes and Chapter 8 highlights ideas on future research.

2. Foundations of sports betting

Sports betting is a form of gambling where a person wagers money on that there will be a certain outcome in a sports event. Sports betting is different from other forms of gambling because in e.g. casino games, the probability of winning is known in certainty but in sports betting the probability of winning is subjective. This gives market participant a possibility to bet on sports with positive expected value if the subjective probabilities of the bettor are more accurate than the odds offered by the bookmaker. This isn't possible in other forms of gambling.

Sports betting markets are a type of prediction markets where traded assets are related to outcomes of sporting events. It has been established in several studies that prediction markets are very accurate at assessing the probability of a given event (e.g. Alberola and Garcia-Fornes, 2012). This chapter firsts gives a foundation to understand what motivates people to participate in sports betting, what features are characteristic for the sports betting markets and how different market and bet types differ from each other. After that, the basic concepts related to the mathematical side of sports betting, like odds, value, profitability and staking, are presented. The structure of the chapter mostly follows the work of Tiitu (2016).

2.1. Motivations for sports betting

The motivation that people have to participate in sports betting has been explained in academic literature with the help of Expected Utility Theory (EUT). The theory, originally proposed by Bernoulli already in 1738, and its variations have been used to analyze the behavior of bettors and there are two distinct major types of bettors based on their behavior. The first group are the recreational bettors who bet on sports to add excitement to watching their favorite team play or watching or following sports in general. Recreational bettors usually don't put too much emphasis on or analyze what the expected value of their bets is but they go on a gut feeling or based on very simple statistics, e.g. betting on higher ranked team no matter the odds. More often than not, recreational bettors are losing money in the long run because of this. It can be argued though that this doesn't bother them too much as sports betting never was really about maximizing wealth for them but a pleasurable pastime instead. For example, Busche and Hall (1988) have recognized this behavior of viewing sports betting as a hobby.

The other group are the people that are serious about trying to maximize their wealth and beat the betting market. But even beating the market doesn't automatically mean being profitable as sports betting for the bettor is a negative sum game and the total ROI for odds provided by the bookmakers is at most around 98%. For example, Ali (1977) has conducted research about the behavior of sports bettors under the assumption that they're participating in the market in order to maximize their wealth. It's been studied though that only a very small amount of people, a few percent, that do sports betting, do it profitably on a consistent basis. But when a bettor is able to find an edge over the market, the possibility of decent capital gains from funds invested into sports betting are boosted by the fact that the bankroll is regularly rolled over, increasing the ROI on the original bankroll. If a bettor is able to find a 2% edge for 365 bets per year, one per day, a very conservative approach of staking 1% of the original bankroll per each bet yields an annual profit of 7.3% on the original investment in the bankroll. Especially the number of bets where there is a perceived edge though, may be a lot higher, as in a single regular season in US major sports alone, there are 1271 NHL games, 1312 NBA games, 256 NFL games and as many as 2430 MLB games.

2.2. Features defining the online sports betting markets

The first online bookmaker was Intertops that opened their online sports betting site in 1996 (Domeneghetti, 2014). Since then, the number of online bookmakers has grown significantly and currently the leading Sportsbook rating site Sportsbook Review (SBR) lists 175 of them on their site (Sportsbook Review, 2019). The list isn't exhaustive though as it excludes a lot of smaller bookmakers and especially many sites that are only available regionally.

The expansions of the market has naturally resulted in increased competition that has forced the bookmaking companies to take measures to respond to the competition. One consequence of this has been that the margins that the bookmaker charge have decreased over time. Lower margins mean that there are more opportunities for profitable betting strategies in the market as the errors in the implied probability distribution need to be smaller in order for there to be positive expected value selections available.

2.2.1. Types of online sports betting markets

There are three different types of sports betting markets. The bookmaker system and the pari-mutuel system have existed already before the era of online sports betting but the third one, exchange betting, has only emerged within the last decade because it's heavily powered by ICT.

The bookmaker system

The most used form of organizing sports betting markets is the bookmaker system. In this system, the bookmakers set fixed betting odds that are published depending on the event, as late as the day of the event to as early as even years before the event occurs. Bettors choose their bets at these odds and the bookmaking company acts as market maker and automatically takes the opposite side of the bet, i.e. carrying the counterparty risk. In theory, bookmakers might accept unlimited volume of betting at the odds published, but in practice, each event has a maximum amount that can be wagered at given odds. The maximum amount can depend on the event, the limits are usually larger for more notable events, and also on the customer profile of the customer placing the bet. The bookmakers also have the possibility to adjust the odds that have been already published based on e.g. new information or based on betting activity of the event. Still, the bettor's claim is always tied to the odds that were given at the time of placing the bet.

The pari-mutuel system

In the pari-mutuel system, the stakes on all possible outcomes of a sporting event are added together and after the bookmaker's profit margin has been applied, the winnings are distributed to the winners according to their relative stakes. The odds fluctuate based on the proportion of bets on each outcome at the observed point in time and the claim of an individual bettor is not fixed before the start of the event but is dependent on all the incoming stakes until the market is closed right before the start of the event. So, compared to the bookmaker system, the odds that a bettor receives are not yet determined at the time of placing the bet. In addition, the system is risk-free for the operator of pari-mutuel betting as the profit margin is applied to each outcome equally before paying out the wins, whereas this is not usually the case for the bookmaker system. Essentially, at pari-mutuel betting markets, the bettors are effectively competing against each other but the profit margin of the organizer tilts makes beating the market more difficult. The pari-mutuel system for betting is still common in horse racing, but for other sports it's becoming less important compared to the bookmaker and betting exchange systems. In Finland though, the national gambling and sports betting operator Veikkaus, has e.g. a popular multi-sport correct score betting market, Tulosveto, and it's variant Moniveto, that utilizes the pari-mutuel system. Also, a significant part of academic research of betting markets, and especially on the efficiency of them, has focused on pari-mutuel betting. (Franck et al., 2013)

The exchange betting system

Exchange betting has emerged lately as a new and popular betting market system. The emergence has been revolutionary for the sports betting industry as the betting exchanges are originally inspired by electronic financial exchanges and rely heavily on ICT and the arrival of online betting. Betting exchange allows the bettor to place wagers like in bookmaker system, but the counterparty of the bet is another bettor who has chosen to opposite outcome of the event, instead of a bookmaker. Whereas in bookmaker and pari-mutuel betting market settings bettors can only bet on a given outcome to occur, which is called 'back' in the exchange market, in the exchange setting they can also bet against a given outcome to occur, called 'lay' bets. Instead of carrying the counterparty risk themselves and charging for that by offering odds with a profit margin built-in, betting exchanges only act as an intermediary for the bets and charge a commission for that. Sometimes, though, the exchanges may assume some of the counterparty risk and place bids on odds themselves to attract action on certain events. Exchange betting has been researched a lot since its introduction, by e.g. Smith et al. (2006, 2009), Gil and Levitt (2007), Franck et al. (2010, 2013) and Croxson and Reade (2014). Betfair is currently the most popular betting exchange in the world, processing around seven million trades per day. (Laffey, 2005; Franck et al., 2013; Croxson and Reade, 2014)

2.2.2. Other key features

Other essential features of online sports betting, but also of traditional sports betting, are the types of bets offered, different sports available for betting, types of odds and the timing of betting in relation to the start of the event, which is an area of interest for this study. This study is about ice hockey, for which the preferred type of bet in Finland has been the 1X2 where the bettor is placing a wager on one of the teams to win or that the match will end in a draw after regular time. This has been due to the fact that 1X2 is the standard type of bet for Pitkäveto, which is the most popular sports betting product of the Finnish gambling and sports betting company Veikkaus (Veikkaus, 2018). Internationally for ice hockey, and especially for NHL in the United States, the most popular type of bet is moneyline which means betting on the ultimate winner of the match, including overtime and the possible penalty shootout if the match ends in a draw (Drafkings, 2019). Besides 1X2 and moneyline bets there are several other types of bets available for hockey. That includes handicap bets, the most common of which, -1,5 / +1,5 handicap, is often called “puck line”. Over/under bets on combined total amount of goals in a match are popular as well, as is betting on the amount of goals that a single team scores in the match. Many bookmakers also offer a bet called “Grand Salami” for NHL, which means betting on whether home or away teams will score more goals combined in all of the matches played that day.

The bets can be either single or multiple (also called parlay). Single bets have only one selection and if that’s correct, the bet will win. Multiple bets include several selections that all need to be correct in order for the bet to be paid out as a winner. Although theoretically multiple betting could increase the expected value of the bets, if a bettor was able to make constant selections with positive expected value, in practice every sophisticated bettor almost exclusively places single bets. This is to decrease variance and to minimize the effect of possible misjudgments in analysis which results in negative expected value bets. Single bets on moneyline are also the focus of this study.

Nowadays bets are offered on a very wide range of sports, on both domestic and international events. Betting has also expanded to other than sporting events, for example political elections and the annual Eurovision song contest are events that attract the betting public and thus bookmakers offer odds for those. Globally and in most countries, association football is the most popular sport in terms of betting volume (Finnigan and Nordsted, 2010; Oikonomidis and Johnson, 2011, 204; Bitcoin Chaser, 2017). This study focuses on ice hockey though which is a significant betting market in Finland due to the sport being extremely popular in the country. According to Bitcoin Chaser (2017), ice hockey, basically NHL, is also the sixth most bet on sport in the United States after the biggest major sports NFL, NBA and MLB but also soccer and boxing/martial arts (e.g. UFC).

The traditional way of betting has involved placing the wagers before the start of an event but the emergence of online betting has also introduced live betting during the events as a popular form of betting. Live betting already constitutes a very significant part of bookmakers' gross margins (Church-Sanders, 2011; Croxson and Reade, 2011). Compared to pre-match betting, the situations are changing quickly and sometimes also dramatically as the event is live in-play (Kotlyar and Smyrnova, 2012). Especially bets placed before the start of an event, but also live bets, can also have longer horizon for settlement than just one match or part of a match. Probably the best example is that for most of the leagues and tournaments that are available for betting, there are usually outright markets available, meaning that bets can be placed on who will ultimately win the league at the end of the season or win the tournament.

Finally, there's a difference between fixed odds and variable odds. Fixed odds mean that the potential cash flows, i.e. winnings, are determined at the time of placing bet. Changes in odds after the bet is placed, don't have any effect on the bet. Variable odds on the other hand mean that the final odds and potential winnings aren't known at the time when the bet is placed, usually not until the betting has closed for the event. Of the betting market systems presented in section 2.2.1, the bookmaker system and betting exchanges offer fixed odds whereas odds in the pari-mutuel betting markets are variable. This study utilizes fixed betting odds provided by a bookmaker at various time points during the lifecycle of betting events.

2.3. Odds, probabilities, and the bookmaker margin

In sports betting, odds represent the amount of winnings the bookmaker will potentially pay out in relation to the stake of the bettor (see, e.g., Buchdal, 2003, 11). Odds reflect the offered probability of the outcome of the event. The higher the odds are, the lower is the offered probability and vice versa. There are different formats for presenting the odds and the usage of them depends on the geographical location and also on the betting market. In Northern Europe, the most common type is decimal odds, where the odds are the inverse of the offered probability of the outcome. Formally, when using decimal odds for a sporting event with n mutually exclusive outcomes, the offered probability of the j:th (j = 1,2,…,n) outcome is inversely related and defined as

$$\theta_j = \frac{1}{\delta_j}, \tag{1}$$

where θ_j is the offered probability and δ_j are the decimal odds quoted for the outcome. Especially in United Kingdom and Ireland, fractional odds are used. Fractional odds represent the amount of potential net payout of the bet if it wins and it's the inverse of the statistical definition of odds. In statistics, odds represent the ratio of probabilities for an outcome to happen and not to happen (Fulton et al., 2012). For j:th outcome of an event statistical odds can be defined as

$$\mu_j = \frac{\theta_j}{1-\theta_j}, \tag{2}$$

where μ_j are the statistical odds. Fractional odds for j:th outcome of an event can then be written as the inverse of these statistical odds

$$\nu_j = \frac{1}{\mu_j}, \tag{3}$$

where ν_j is the fractional odds. Throughout this study, the term 'odds' doesn't refer to its statistical definition, but it refers to betting odds published by a bookmaker to represent the potential payout of a winning bet.

In United States, an odds system called 'US odds' or 'Moneyline odds' is used. In the system, the outcomes that have an offered probability of more than 50%, are accompanied by a minus (-) sign. The minus odds represent the amount that needs to be wagered in order to gain a net win of $100. The odds for outcomes that have an offered probability of 50% or less, are accompanied by a positive (+) sign. Those odds represent the potential net winnings for every $100 wagered. The US odds can be converted into decimal odds with different formulas, depending whether the odds are minus or plus odds. The minus odds are converted into decimal odds with

$$\delta_j = \frac{100}{-\gamma_j} + 1, \tag{4}$$

and plus odds with

$$\delta_j = \frac{\gamma_j}{100} + 1, \tag{5}$$

where γ_j is the US odds quoted for the outcome. For this study, odds are presented in decimal format but the data collection was partially done in US odds and then converted into decimal odds using the above formulas as the odds information wasn't readily available in decimal format for the first season of data, 2015-2016, and for the first 61 matches of the 2016-2017 season.

The bookmaker margin can be obtained by adding up the offered probabilities θ_j of all possible outcomes so that

$$M = \sum_{j=1}^{n} \theta_j - 1, \tag{6}$$

where M is the bookmaker margin.

Based on information about the bookmaker margin, the implied probabilities for the outcomes can be converted from the offered probabilities. To do that, the offered probabilities must be scaled so that they sum to 1, i.e. by removing the bookmaker margin. The implied probabilities represent the bookmaker's true view about the probabilities only when assuming that the bookmaker margin has been distributed equally between the outcomes. The implied probability of the j:th outcome is defined as

$$\rho_j = \frac{\theta_j}{1+M}, \tag{7}$$

where ρ_j is the implied probability. Similarly, the fair odds, with no bookmaker margin, can then be calculated as

$$\vartheta_j = (1+M)\delta_j, \tag{8}$$

where ϑ_j refers to the scaled, decimal odds for the outcome j. In this study, the favorite-underdog status is determined based on implied probability. A favorite is a team that has an implied win probability of >50%, calculated as presented in Eq. (7) and correspondingly an underdog is a team that has an implied win probability of <50%.

While the implied probabilities can be viewed as probability set by the bookmakers and the market for an outcome to happen, the objective probabilities are something that can never be known for certain in sports betting and the estimates are not revealed by bookmakers. Each sporting event is unique and the factors affecting the outcomes are stochastic so the estimates can only be validated by large enough sample of events to see if the outcomes are in line of the estimates. Lately, advanced statistics about sporting events have helped in evaluating the validity of estimates produced, regardless of the actual outcome of the event. For example, in association football and ice hockey, expected goal (or xG) models provide information about how many goals a team would've scored on average with the shots they took (Opta, 2019). Although these models have their drawbacks too, especially with smaller sample sizes they provide a far more accurate measure of the goodness of the estimates than relying solely on the outcomes of the events that can be heavily influenced by random occurrences and variance.

2.4. Making money in online sports

Making a long term profit out of sports betting in the bookmaker market requires that one is able to make constant picks with positive expected value, i.e. able to gain an edge over the bookmakers. Besides the positive value picks, money management has to be considered to be able to maximize the profits. In the following, I present three methods to gain an edge over the bookmakers and four money management strategy categories. Besides these methods, there are unsustainable methods for profiting, like exploiting the clear errors of bookmakers in odds setting, which often quickly leads to limiting and closing of the accounts at bookmakers and sometimes also to confiscation of winnings. Also, there are downright illegal ways of profiting, like match fixing, which has to be considered to be one of the greatest threats for the sports betting markets. These unsustainable or illegal ways of profiting are excluded here.

2.4.1. Value betting

Value betting is a sports betting strategy where the bettor seeks to gain an edge over the bookmaker by only betting on outcomes that, based on the bettor's analysis, have a higher objective probability than offered probability. Buchdal (2003, 42–53) describes successful betting to constitute of understanding and managing probabilities and states that the only way to overcome the profit margin that the bookmakers charge, is value betting. To measure the value of a bet, the edge or expected value (EV) must be calculated. Based on the analysis done about the probabilities of the outcomes for an event, if the expected value of a selection exceeds 1, a bet is profitable in long term and can be considered a value bet. Put formally, a bet on a given outcome j is a value bet if

$$r_j = \frac{\pi_j}{\theta_j} > 1, \qquad (9)$$

where r_j is the expected value (also known edge), π_j is the objective probability, and θ_j is the offered probability of the outcome.

Employing value betting method usually means producing estimates about the probabilities for outcomes of a betting event with statistical models and then identifying the value bets according to the model by using Equation (9), replacing the objective probability with the probability given by the model. When a value bet is found, a sharp bettor will want to place the bet on a trustworthy bookmaker that offers the highest odds for the chosen outcome. Placing the bet with the highest odds isn't that straightforward of a strategy though if sports betting is a serious longer term investment for the bettor. For even semi-professional bettors, there are only few bookmakers though that will accept their bets in long term due to most companies eventually limiting the stakes of constantly winning players. This means that especially for tracking purposes, odds have to be taken from a bookmaker that will accept their bets long term to maintain a realistic view of the results. This study is based on the concept of value betting in the sense that it tries to establish the time when the odds are the most efficient, i.e. best reflect the objective probabilities for the match, and find value bets based on that information.

2.4.2. Sports arbitrage

Individual bookmakers always have a positive profit margin built into their odds, but sometimes there are situations on the market where the odds of two or more bookmakers have such a contrast that it results in a negative profit margin on aggregate. Arbitrage refers to the practice of making a risk-free profit by exploiting price differentials between different markets and sports arbitrage refers to betting on sports so that a profit is guaranteed for an event, regardless of its outcome, by combining the odds of different bookmakers. Formally, a bet is an arbitrage if

$$\frac{1}{\sum_{j=1}^{n}\frac{1}{\delta_j}} - 1 > 0, \qquad (10)$$

where n is the number of outcomes and δ_j is the decimal odds quoted for the j:th outcome. The occurrence of sports arbitrage opportunities depends on two factors: the discrepancy in odds between bookmakers and the profit margins built in the odds. The bigger the discrepancy of the odds is between two or more bookmakers and the smaller the profit margins are, the more there are opportunities for arbitrage.

Although it may seem that sports arbitrage offers relatively easy risk-free profits, there are some practical challenges that have to be carefully taken into account. For example Franck et al. (2013) have addressed some of these challenges. They mention that the main restrictions are related to cancellation of bets and stake limits. Franck et al. observed ten bookmakers operating in the United Kingdom and concluded that all of them had a Terms & Conditions statement that the bookmakers reserve the right to cancel or reduce the stake of an already-placed bet based on technical difficulties, suspicion of fraud or suspicion of arbitrage betting. Also, the maximum stakes of a user may be limited if the bookmaker identifies one as an arbitrageur. Franck et al. (2013) in their study identified two bookmakers that explicitly stated that they do this to customers identified as arbitrageurs. This is a significant challenge because sports arbitrage opportunities are also often either very small in terms of negative bookmaker profit margin, or they're available for only a short amount of time before one or both of the bookmakers update their odds. As the negative profit margins are often small, the bet amounts must be significant to make arbitrage betting worthwhile. On top of these challenges, an arbitrageur needs accounts at a very large number of bookmakers as the arbitrage opportunities don't usually involve two large bookmakers but at least the other one is a smaller bookmaking company. Depositing and maintaining large enough funds on the accounts of smaller bookmakers isn't completely risk-free and sometimes withdrawals may cause difficulties from such smaller companies. Sometimes also the rules for the settlement of bets may differ between bookmakers, creating risk for the bettor taking advantage of arbitrage situations. The most common rule differences are related to the settlement of bets where the match in question is finished before the expected full time due to for example one of the players retiring in the middle of a tennis match or a football match being forced to be abandoned before full time.

Sometimes sophisticated bettors may use arbitrages to their advantage if they occur due to moving odds. A sophisticated bettor can have a view that a certain selection has value at current odds and places a bet on that selection. Usually, if the bettor is correct about the value, the odds will fall over time towards the start of the event. If the odds movement is large enough, the bettor may be able to make a risk free profit by betting also on the other selection(s) of the event. This kind of strategy is mostly seen only as a way of managing risks though and reducing variance as usually the bookmaker margin of the other selection(s) will diminish the expected value of the original bet.

2.4.3. *Promotions and bonuses*

The third way of making money from sports betting is by utilizing bonuses, free bets and other promotions that some bookmakers offer to attract new and maintain their existing customers. A strategy known as matched betting involves having an account at a bookmaker that offers these kind of promotions and another account at another bookmaker or at a betting exchange. By placing a bet with the bookmaker handing out the incentive and at the same time placing a bet against that outcome, a guaranteed profit is made regardless of the outcome of the event.

Another way to take advantage of the promotions, is to treat them like value bets. By simply placing the bets with e.g. bonus money or enhanced odds, the expected value of the bet is often positive and it can be estimated in more detail, especially when the promotion involves enhanced odds, by using Equation (9). This will significantly increase the variance compared to matched betting but it can be argued that the actual expected value is higher because the bettor will save the bookmaker margin paid to the counterparty bookmakers or betting exchanges for the counterbets.

The supply of promotions and bonuses is limited though, as most of the bookmakers that do offer bonuses or other promotions, offer them only to new customers and they can be only used once. And also, bookmakers that have recurring promotions, may rule out players, that constantly make a profit from the bonuses, from participating in them.

2.4.4. *Principles of money management*

Besides gaining a mathematical advantage over bookmakers and the market, making money out of sports betting requires good management and trying to optimize the staking. It means setting aside a bankroll dedicated for sports betting and having a staking plan, how much to invest on each bet with a positive expected value. The goal should be to maximize the returns and reduce the risk of bankruptcy to an acceptable level.

There are four categories for staking strategies: fixed staking, variable staking, progressive staking and percentage staking. Fixed staking means always betting the same amount regardless of the odds or expected value of the bet and regardless of the recent development of the bankroll. In variable staking, the bettor changes the size of the bet based on a criteria that can vary. A simple example of variable staking is betting to always gain the same net profit with every winning bet. Progressive staking refers to increasing or decreasing the stake based on the result of the previous bet. The most common progressive staking plans involve increasing the stake after a loss and reducing it after a win. This may give an illusion of steadier returns compared to for example fixed staking, but the downside is that the risk of significant loss increases if the bettor hits a streak of losing selections. One of the most famous this kind of progressive staking plans is called Martingale, introduced by Ville (1939), which involves doubling the bet amount after every loss. Lastly, percentage, or proportional, staking means staking a certain percentage of the current bankroll with each bet, instead of staking a percentage of the original starting bankroll.

2.4.5. Kelly criterion

Kelly criterion is a capital growth model that maximizes the expected value of the logarithm of wealth on a bet-by-bet basis. It's a percentage staking strategy that has qualities of also variable staking as the bet amount varies from one bet to another based on the expected value of the bets even if the bankroll before each bet was the same. Kelly criterion was first introduced by Kelly (1956) and his work has been later expanded and proved by Breiman (1961) and Algoet and Cover (1988). For sports betting, Gramm and Ziemba (2008, 320–321) highlight the three most significant qualities of the model: (1) it maximizes the growth rate of capital, (2) it minimizes the expected time to reach any specific level of wealth, and (3) in the long run it will almost surely perform better than any other essentially different betting strategy. Also when considering financial markets, where investors must decide how large proportion of capital to invest after identifying a perceivably profitable investment opportunity, the methodology is just as useful; the Kelly-optimal investment strategy has been used in calculations of optimal portfolio weights in allocations problems, whether concerning multi asset or worldwide asset allocations (MacLean and Ziemba, 2006).

Buchdal (2003, 155–156) states that the Kelly-optimal staking offers an optimum risk-reward ratio, allowing the bankroll used for betting to grow at the maximum rate for minimum risk, when used repeatedly over a period of time. With Kelly staking, the stake of a bettor is always defined as a percentage of the bankroll, so a growing bankroll means stakes get larger and vice versa. For a given size of the bankroll, the size of the stake is dependent on both the odds available and the perceived expected value of the given bet. Formally, the stake size according to the Kelly-optimal staking plan for placing a bet on the outcome j, which has been found to have positive expected value, i.e. is a value bet, is defined as

$$k_j = \frac{r_j - 1}{\delta_j - 1}, \tag{11}$$

where kj is the size of the stake recommended by the Kelly criterion as a proportion of the bankroll, rj is the perceived expected value of the bet as defined in Eq. (9) and δj is the offered odds quoted for the outcome. The main drawback of the Kelly criterion is that the model may suggest stakes that are in some cases very large, if the perceived expected value of the bet, that the stake is calculated for, is high. Individual bets may then be too large to be acceptable, although being a percentage staking model, a bettor can never go completely bust using Kelly criterion. Especially, if the probability estimates are flawed, the risk of over betting is high. Most of the time bettors who apply the Kelly methodology will eventually have an increasing bankroll but it does not exclude the possibility of losing one's wealth in the case that either perceived value bets aren't such in reality, or a huge value bet loses. In the context of only one bet, there isn't a definitive way to make a distinction between the two. A method to avoid these kind of oversized bets is to use fractional Kelly staking where the calculated optimal growth rate is compromised but security is increased by betting only a fraction of the optimal stake calculated by the original Kelly formula. This allows the bettor to achieve close-to-optimal capital growth rate with increased security as the risk aversion is then higher than zero (MacLean et al., 1992; 2010).

Formally, the fractional Kelly staking plan for placing a value bet on the outcome j is given by

$$f_j = \frac{k_j}{g}, \tag{12}$$

where fj is the size of the fractional Kelly stake as a decimal percentage of the bankroll, kj is the size of the standard Kelly stake determined in Eq. (11), and g is the applied divisor so that $g > 1$. One common choice is to apply the half Kelly staking plan, according to which $g = 2$, compromising the optimal growth rate by 25% (Thorp, 2006, 411–412). Kelly criterion will also be applied in this study, together with the basic fixed and variable staking plans.

3. Betting odds as a predictor and betting market efficiency

There are several ways for trying to predict the outcome of a sporting event. Nowadays there's a lot of information and statistics available and the tools one can use to assess the upcoming matches include recent scores, the league tables and statistics and power rankings, i.e. listings often published especially for US major sports leagues that try to incorporate in the analysis elements that aren't clearly visible in the results or the league table, like the quality of the opponents so far. Boulier and Stelker (2003) concluded for National Football League (NFL) though, that betting odds are a better measure for predicting outcomes than any of these statistics alone.

Tax and Joustra (2015) conducted a study of soccer matches in Dutch Eredivisie, the highest national league, using machine learning techniques to compare predictions based on publicly available data to ones based on betting odds. This publicly available data included factors such as previous performance, head-to-head performance, winning or losing streaks, home advantage, motivational factors, fatigue, travelling distance, club budgets, key player presence/absence and possible managerial changes. The best model created based on betting odds had higher prediction accuracy than the best model based on public data but the difference in the percentages of correct predictions wasn't large enough to be statistically significant. The study was also only predicting whether a game would be won by a team or not. It didn't consider the profitability of betting on the predicted winners proposed by the models.

This study focuses on NHL ice hockey moneyline bets, which is a fixed odds market, i.e. the bettor's potential winnings are determined based on the odds that were offered at the time of placing the bet. It doesn't matter if the odds change after placing the bet but a sophisticated bettor trying to maximize profits, will want to place the bets when they are as high as possible. That's why it's important to understand the basic mechanics behind the odds movement so that a best guess can be made about the optimal time to place the bet.

3.1. The difference of odds movement in pari-mutuel and fixed odds betting

Pari-mutuel betting markets are quite limited nowadays but for some sports they're still popular, horse racing being the best example. In pari-mutuel markets, the odds are variable, meaning that no matter when a bet is placed during the market being open for an event, the final odds and subsequently the potential winnings are only decided when the market is closed for the event. The odds are simply calculated based on the proportion of money bet on each outcome available for selection and by deducting the bookmaker margin from those fair odds. This leads to the fact that the momentary odds presented throughout the lifecycle of the market are the more volatile the earlier they are checked as there's less money bet in the market. A study by Gramm et al. (2016) about Australian horse racing events established that informed bettors and other smart money players in pari-mutuel markets place their wagers as late as possible. The reason for this is quite clear. The later a bet is placed, the easier it is to estimate the final odds and subsequently the edge that a bettor potentially has. For fixed odds betting, waiting until the final moments before the closing of a market isn't nearly as important as for pari-mutuel betting. As the odds provided at any given moment are the ones dictating the potential winnings if a bet was placed at that time, all that is needed for placing a value bet is a perceived edge over the bookmaker and a notion that the odds will not increase significantly as long as no new information affects the market. Usually if there's a sophisticatedly perceived edge in the market, it's wise to utilize the edge as soon as possible as the odds are likely to decrease on that selection. That's why many informed bettors in fixed odds markets place their bets very early when the odds represent mostly the initial views of the bookmaker and the money bet by the market hasn't affected the odds very much yet. The drawback of this method is that the stake limits can be quite low at the time of the opening of the events and the odds are more likely to move quickly and significantly if large wagers are placed on certain outcomes. The aim of this study is to overcome these drawbacks and the difficulty of making sophisticated predictions about the outcomes of events. This study attempts to establish a rough time point when the odds for NHL moneyline events are the most predictive of the objective probabilities based on historical data. This time point is also when the market is the most efficient, meaning that no simple betting strategy would beat the market, yielding a loss equal to the bookmaker margin.

3.2. The movement and setting of fixed betting odds

Although, when placing a bet with fixed odds, the bettor will receive the potential payout according to the odds that were offered at the time of placing the bet, it's important to keep in mind that the odds can move afterwards before the start of the event, possibly offering an opportunity to place the same bet with better odds. The odds movement can happen for different reasons that are presented next.

First, new information surfacing can cause the odds to be changed by bookmakers. If, for example, a news is released that the star player of a team will be missing for the night's game, it will affect the probability of each team winning and sometimes there may be an effect on the probable amount of goals or points to be scored in the game too. Usually new information is what causes the largest odds movements. An informed bettor incorporates the possibility of additional information surfacing into the analysis of the event. Second, betting activity is a natural cause of odds movement in the market. By default, bookmakers move the odds in order to reduce the risk related to the event by decreasing the odds of the outcome that's receiving an overly large share of the bets. In contrast, the odds of the other outcome or outcomes move up to attract bets on those and balance the action. The odds movements caused by betting activity are usually larger closer to the opening of the betting event since there's less money bet on the event at that time and individual bets may be larger in relation to the total amount wagered on the event. Towards the closing of betting and start of the event, the odds movement caused by betting activity tends to stabilize as the share of money wagered on each outcome isn't as likely to be dramatically changed by individual bets.

It has been established though that sometimes bookmakers become active participants in the market by taking a stand on a certain outcome although the betting dollars aren't evenly spread between the possible outcomes (Paul and Weinbach, 2007; Paul and Weinbach, 2008). Before Paul and Weinbach, also Levitt (2004) argued that bookmakers actually set odds to maximize their expected profits instead of aiming to make a risk-free profit by balancing the action between the possible outcomes. Paul and Weinbach researched bookmaker behavior for NFL and NBA and found that for NFL, bookmakers were exploiting

popular bettor biases, like heavily betting on big favorites, to maximize profits. For NBA, though, it was concluded that the bookmakers were setting the odds as a true forecast of the outcome of the games. If the forecast are accurate on average, this allows the bookmakers to earn a profit equal to the bookmaker margin in the long run. American football (NFL) and basketball (NBA) are so-called point spread markets in the US, meaning that instead of betting on the absolute winner of a match, a point spread is applied so that each team playing has an almost even chance of winning the game when the handicap is taken into account.
For NHL, betting on the ultimate winner of the game, the moneyline, is the most popular form of betting so Paul and Weinbach extended their research by also focusing on NHL (2012). They divided their dataset of 3 seasons into two groups: one where home team is the favorite to win and another where the road team is the favorite. They found that for both parts of the data the hypothesis of balancing the book could be rejected. Paul and Weinbach concluded that for NHL, bookmakers are mostly setting the odds as a forecast of the actual outcome the games. They proposed three reasons for this strategy. First one is that the long-run strategy of earning a profit equal to the bookmaker margin offers steady returns and utilizes the large volume of NHL games, 1230 per regular season until the 2017-2018 season when the Vegas Golden Knights joined the league as the 31st team, and 1271 since then. Second is that setting the odds as a forecast discourages informed bettors from joining the market and betting on games. This means that the bookmakers don’t have to compete with those informed bettors for the additional profits that exploiting the biases of betting public might make available. Third reason is that due to the relatively small size of NHL betting market, compared to for example NFL, the transaction costs of actively managing the odds to exploit the biases may be too high to be worthwhile for the bookmakers. This involves preventing informed bettors from exploiting the biased odds and trying to prevent uninformed bettors from losing too much so that they would stay in the market. This also relates to the increased competition in the market. People often have accounts on several bookmakers, so in case of biased odds, people may simply place their bets on a different bookmaker instead if the odds on a certain outcome are bad.

3.3. The accuracy of betting odds and odds movement in predicting actual outcomes

Sports betting markets are a type of prediction market where the traded assets are related to outcomes of sports events. Prediction markets have been found to be powerful predictors of the probabilities of future events (Alberola and Garcia-Fornes, 2012). As mentioned in the previous section 3.2, e.g. for NHL ice hockey, bookmakers have been found to also try to set the odds as best forecasts of the outcomes of the events.

One of the few studies focusing solely on evaluating betting odds movement as a predictor of actual game outcomes was done Odachowski and Grekow (2012). The study was done for 1X2 soccer markets and utilized the odds of Pinnacle, the same bookmaker as used for the odds extraction of this study. Odachowski and Grekow recorded odds changes for the last 10 hours before the start of the games and divided the data into four sampling periods, which were the whole 10 hours and the 10 hours divided equally to 3 periods of 3 hours 20 minutes each. The odds movement was analyzed by 24 standard features regarding the odds and their movement. Different prediction algorithms were used to classify the matches as either home wins, draws or away wins. Eventually draws were taken off the model, essentially making predictions for the 'Draw No Bet' market where in case of a draw, the initial stake is returned to the bettor. The Draw No Bet model was able to achieve a hit rate of 70.3% with a Bayesian network algorithm. For actual betting purposes the study didn't offer much though, as the predictions were in simple binary win-loss form, completely disregarding the underlying odds that were the basis of the models, so there's no way of knowing what were the winning probabilities for the predictions.

The accuracy of the odds is closely connected to efficiency of the betting markets, which has been studied rather widely. When assuming that bookmakers were setting the odds to balance the action on both sides of the betting event, the findings in favor of market efficiency were deemed to be a result of actions of the bettors but with the notion that the odds are set as forecasts of the actual outcomes, the possible findings in favor of market efficiency may rather be due to the excellent forecasting by the bookmakers. The studies about the efficiency of sports betting markets have yielded varying results, depending on the sport and dataset used. One of the most extensive studies so far was done by Tiitu (2016). The research data included the market average and highest 1X2 (Home win-Draw-Away win) odds of nearly 96,000 soccer matches played between 2009 and 2014. Tiitu found the association football betting market to be both statistically and economically[1] weak form inefficient. Tiitu's (2016) study about the efficiency of the sports betting market found that there existed a clear favorite-longshot bias. Favorite-longshot bias will be described in detail next chapter 3.4.

3.4. The favorite-longshot bias

Favorite-longshot bias (FLB) is a widely observed phenomenon in sports betting, gambling and financial markets. It was first discovered by Griffith (1949) and has since become the most known bias in the betting market literature. It means that the market participants tend to undervalue the favorites, i.e. more likely outcomes, and overvalue the longshots, i.e. underdogs or the less likely outcomes. When the implied probability of an outcome increases, the objective probability increases also but more than the implied probability. This leads to the fact that betting on favorites yields on average a better return than betting on longshots, i.e. the underdogs. It's often stated that the favorite-longshot bias means that the betting public overbets the favorites and underbets the longshots but as noted before, the assumption that bookmakers set the odds to balance the action on the event isn't accurate for many of the sports and fixed odds betting events. The statement is true for pari-mutuel markets though, e.g. horse racing, in which the FLB has been researched most extensively. Several studies have found signs of FLB in horse racing, e.g. Dowie (1976), Ali (1977), Snyder (1978), Hausch et al. (1981), Asch et al. (1982, 1984), Vaughan Williams and Paton (1997), Jullien and Salanié (2000), Gramm and McKinney (2009), Snowberg and Wolfers (2010) and Gramm et al. (2016).

Evidence for favorite-longshot bias has been found for several other sports too. There are several studies for association football that have provided indications of FLB, e.g. by Pope and Peel (1989), Paton and Vaughan Williams (1998), Cain et al. (2000, 2003), Deschamps and Gergaud (2007), Vlastakis et al. (2009), Koning (2012), Direr (2013), Nyberg (2014)

---

[1] Tiitu (2016) defines statistical weak form efficiency of the betting market to mean that the betting odds are unbiased estimators of the outcomes of the events when using only historical odds information. Economic weak form efficiency means that a bettor cannot earn profits by using this historical odds information.

and Tiitu (2016). Cain et al. (2003) also found indications of favorite-longshot bias for sports including boxing, cricket, greyhound racing and snooker. For tennis, Abinzano et al. (2019) found clear indications of FLB and particularly for matches that attract a lot of public attention, which usually also means a lot of betting activity on the favorite side of the event. Regardless of the sports though, most of the time the favorite-longshot bias is so small that it cannot be solely used to gain a profit on the betting market because of bookmaker margin. The reverse favorite-longshot bias (RFLB) is the opposite of FLB and means that betting on favorites yields less than betting on longshots. Although FLB is a prominent phenomenon for many sports, studies regarding the US major sports leagues, NFL, NBA, NHL and MLB, have indicated the opposite. Golec and Tamarkin (1991) and Borghesi (2012) detected RFLB in NFL, Paul and Weinbach (2005) in NBA, Woodland and Woodland (2001), Gandar et al. (2004) and Paul and Weinbach (2012) in NHL and Woodland and Woodland (1994, 2003) in MLB. As this study is centered on NHL ice hockey, the notion about RFLB is important for hypothesizing.

3.5. Hypotheses

This section presents the hypotheses of this study. As described above, for pari-mutuel markets, the accuracy of closing odds has been compared to earlier odds and the closing odds have been found to be the most accurate predictors of actual outcomes. Very few similar studies have been conducted for fixed odds betting markets, and this combined with the notion about the accuracy of closing odds for pari-mutuel markets, we get the following hypothesis:

Hypothesis I: The closing odds are the most accurate predictor of the actual outcome for NHL ice hockey matches.

The second area of interest is if there are any profitable betting strategies that can be created only based on the movement of the betting odds. This is an area that also hasn’t been studied before as such but based on the notion that the market always has a bookmaker margin built in the odds, especially as the odds are from a single bookmaker. As there is no certainty that historical events predict future events, the hypothesis is phrased to refer to only historically profitable betting strategies. We hypothesize the following:

> *Hypothesis II: NHL betting market doesn't allow for historically profitable betting strategies by only analyzing the odds movement*

4. Data and methodology

This chapter presents the dataset used to conduct the empirical part of this study, the methods used to analyze the data and the underlying assumptions that have been applied in the analysis to limit the amount factors to consider.

4.1. Data

The data of this study consists of moneyline odds data for 3,988 NHL games, which covers all the regular season and playoffs games available for betting during three seasons, 2015-2018. Collecting the odds data was a very significant part of the total workload of this study. The odds data has been obtained from sbrodds.com that is one of the only services to provide full history of the odds movement throughout the lifecycle of the betting events. The odds were provided by the bookmaking company Pinnacle. Pinnacle is the bookmaker that most of the professional bettors in Finland use as it offers high bet limits on events and low margins and never restricts winning bettors by limiting the stakes or blocking accounts. That’s why their odds can be considered to be very accurately set to represent the actual probabilities as there’s little room for error. The odds information has been collected for both teams of every game at five different time points:

- The first, i.e. opening odds for the game
- 12 a.m. Eastern Time (ET) on the day of the game
- 1 hour before the scheduled start of the game
- 15 minutes before the scheduled start of the game
- The last, i.e. closing odds for the game

The reasoning behind choosing these time points for the analysis is following. Opening and closing odds are rather self-explanatory as they are the first and last odds quoted for the event. 12 a.m. Eastern Time was chosen because at that time, all the information that arises by the morning of the game day, is incorporated into the odds. Also, betting on NHL is a big market in Europe, especially Northern Europe, and at that point it’s 6 p.m. Central European Time (CET, e.g. Sweden) and 7 p.m. Eastern European Time (EET, e.g. Finland), meaning that a lot of the sophisticated bettors there have done their analysis on the night’s games and have placed their wagers already. By 12 a.m. (ET), teams playing in the East Coast of the

United States or Canada, may also be done with their morning skates, which isn't true for teams playing in the West Coast. That shouldn't be a significant factor for the prediction accuracy of the odds though. The timeslot of 1 hour before the scheduled start of the game was chosen because then information about lineups and starting goalkeepers is out at latest. The most usual lineup information that significantly affects the odds is related to injuries to key players and even more often related to the playing of second-string goalkeepers. 15 minutes before the scheduled start of the game the pre-game warm-ups are finished and for example the final decisions about late fitness tests for injured players have been made.

Table 1 presents the data with home/away split for wins and both mean and median offered odds at the opening and closing of the event. From the table it can be seen that while in 2016-2017 and 2017-2018 seasons the home team win rate has been nearly identical, 2017-2018 being very slightly ahead by 0.1 percentage point difference in win rate. In 2015-2016 though, home teams won about 3 percentage points less, 52.7% of the time. Also, the mean and median odds reflect the fact of this order as 2015-2016 has the lowest offered average probability of home team wins for both opening and closing of mean and median odds and 2017-2018 season has the highest.

Table 2 presents the bookmaker margin for each season. The differences are small, the gap between the smallest and largest average margin is only 0.03 percentage points but the findings are interesting though. The average bookmaker margin has actually increased from the opening to the closing of the betting events in each of the season, contrary to the intuitive idea that as the odds are first published, there's less information on the market and thus the bookmaker margin would be higher. Also, the average closing bookmaker margin has increased each season, which is in contrast with the general development of the online betting market where competition has been increasing and reducing the bookmaker margins have been the main strategy of competing for most of the bookmakers. On the other hand, Pinnacle has established itself as a bookmaking company that offers low margins already before the first season of this study, so it may be that they have been slightly increasing their margins over time.

Table 1
Match outcomes and odds with home/away split

This table presents a summary of match outcomes for each season and with home/away split. The first two columns give the number of wins in a given season for both home and away teams and the frequency of the wins. The third and fourth column list the mean odds at both opening and closing of the events. The fifth and sixth column list the opening and closing median odds.

	Wins	Win frequency	Mean opening odds	Mean closing odds	Median opening odds	Median closing odds
2015-2016						
Home	694	0.527	1.819	1.835	1.769	1.787
Away	624	0.473	2.221	2.207	2.180	2.150
Total	1318	1.000	2.020	2.021	1.952	1.952
2016-2017						
Home	734	0.558	1.812	1.821	1.760	1.750
Away	581	0.442	2.245	2.242	2.190	2.200
Total	1315	1.000	2.029	2.032	1.950	1.950
2017-2018						
Home	757	0.559	1.808	1.805	1.750	1.750
Away	598	0.441	2.245	2.256	2.200	2.200
Total	1355	1.000	2.027	2.031	1.950	1.950

Table 2
Bookmaker margins

This table presents the mean bookmaker margin for each of the seasons at both opening and closing of the betting events. The margins are calculated as defined in Eq. (6).

	Number of matches	Mean Opening bookmaker margin	Mean Closing bookmaker margin
2015-2016	1318	2.386 %	2.396 %
2016-2017	1315	2.381 %	2.403 %
2017-2018	1355	2.393 %	2.413 %

4.2. Methodology

The study is divided into two parts, comparing the predictive accuracy of the odds at the different time points and researching if there are profitable betting strategies to be exploited only based on the odds and their movement. To assess the predictive accuracy of the odds, Brier score is first used to get an overview about the prediction accuracy. Then linear and logistic regression are used in order to further assess and compare the prediction accuracy of the odds at different time points. To obtain results about how to utilize the information about odds and their movement in betting, some simplistic analysis is done about returns when naively betting based on simple principles and about returns when odds movement between time points is considered. For the tests of prediction accuracy, the offered odds by the bookmaker have been converted into implied probabilities as presented in Eq. (7). For the testing of betting strategies, the offered decimal odds are used and when needed, offered probabilities are converted from the odds as in Eq. (1).

4.2.1. Odds grouping

To help assessing the accuracy of betting odds in predicting the actual outcome of the games, for some of the Brier score tests and linear regression, the data is divided into groups after sorting them based on implied probability. In sports betting research, there are two widely established ways of grouping odds, introduced originally by Ali (1977) and Snyder (1978). One is to form the group in such way that each group includes roughly the same amount of observations. The other is that groups are formed so that one group includes all the observations that have an implied probability falling in a certain range. The first is used for both Brier score and linear regression, the latter only for Brier score.

4.2.2. Brier Score

Brier score is a test that evaluates predictions as forecasts of actual outcomes, introduced originally by Brier (1950). The score is calculated individually for each forecast-event pair as a squared error of the probability forecast (range from 0 to 1) and the actual event outcome (binary 0 or 1) and the average of the scores describes the score of a subset of forecast-event pairs. The metric was originally formulated by Brier to be applicable for multi-category forecasts, but a simpler form can be used for binary events, such as in this study. Formally Brier score for binary events can be defined as

$$BS = \frac{1}{N} \sum_{t=1}^{N} (f_t - o_t)^2 \tag{13}$$

where BS is Brier Score, N is the number of forecast-event pairs for which the score is calculated, f_t is the probability forecast of a win and o_t is the actual outcome of the event instance t (0 if a loss, 1 if a win).

4.2.3. Linear regression

I run linear regressions to further explore the relation between the implied and objective probabilities at different time points. The data is binary in nature because it includes one team winning and one team losing each game so linear regression wouldn't produce very good results if run simply on individual games. That's why I estimate the regressions using two different groupings on the odds, 80 groups with 49-50 games in each and 40 groups with 99-100 games in each. The games are divided to the groups based on the implied probabilities as described in section 4.2.1. The regression is run for all five of the time points and for every sort-out based on implied probability of every time point, meaning that there's a total of 25 regressions estimated. The other possible methods would've been simply choosing one time point as a baseline for sorting the data or running a regression only once for each time point, sorted based on probability at that time. Running the regressions with different sort-outs doesn't have the risk of sorting based on a single time point affecting the results as the regressions are run sorted based on all the time points. Also, the regressions attempt to model the odds movement by running the regressions with every sort-out for all of the time points. That way there could be information about, for example, when are the predictions the most accurate for the highest or lowest implied probability teams based on opening odds. The estimated models can be written as

$$\bar{\pi}_h = \alpha + \beta \bar{\rho}_h \tag{14}$$

where $\bar{\pi}_h$ is the objective probability, i.e. the win rate for games in odds group h and $\bar{\rho}_h$ is the average of implied probabilities for games in odds group h. If the implied probability perfectly predicts the objective probability, then $\alpha = 0$ and $\beta = 1$. This means that implied probability is an unbiased estimator of objective probability. If $\beta > 0$, the implied probability predicts the objective probability in the correct direction, i.e. when implied probability increases, so does objective probability. If $\beta < 0$, the market predicts the objective probability in the wrong direction (Tiitu, 2016). If then $\beta > 1$, the implied probability is growing faster than objective probability, thus implying that there is a favorite-longshot bias in the market at that time point. If $0 < \beta < 1$, the opposite holds, which implies that there's a reverse favorite-longshot bias in the market. The differences between time points are evaluated based on R2, α and β values of the regressions. This kind of approach for analyzing regression results has been used by for example Koening (2012) and Tiitu (2016). It must be noted that for very low and very high implied probabilities, this strategy can't be applied with linear regression as it's possible that the result of the regression is an objective probability of < 0 or > 1, which is not feasible. In this study, the implied probabilities range in between 0.22 and 0.78 and thus this doesn't pose a problem here. Figure 1 illustrates an example about a regression that has α < 0 and β > 1, that would indicate that there's an FLB in the market. It can be seen that with low implied probability values the objective probability is less than it would be based on the hypothesis of α = 0 and β = 1. Eventually the straights cross each other and with high implied probability values, the objective probability is higher than based on the hypothesis.

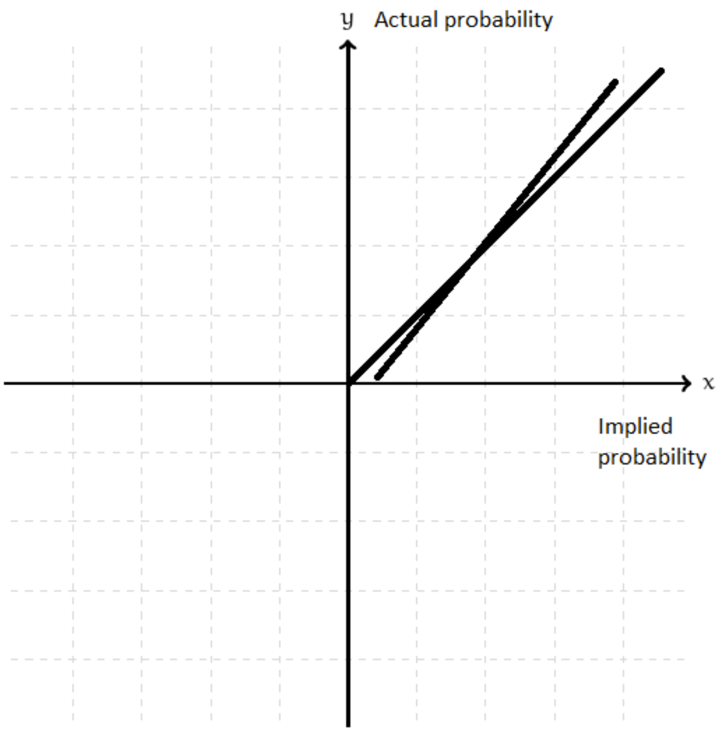

Figure 1. An example of a regression where α < 0 and β > 1 against a regression where α = 0 and β =1

4.2.4. Logistic regression

Besides linear regression, I use logistic regression to assess the relations between implied and objective probabilities. The regression is run separately for each time point. Logistic regression and logistic models in general are meant for assigning probabilities to binary events, as is the case here. One of the teams always wins a game (1) or loses it (0). The regression estimates for variable Y the probability that $Y = 1$.

Based on the output of logistic regression analysis, the resulting probability of $Y = 1$ can be formulated as

$$P(Y = 1) = \frac{1}{1+\exp\left(-(\beta\rho_j - \alpha)\right)} \tag{15}$$

where ρ_j is the implied probability of j:th outcome for the betting event.

Like other methods of assessing the relation between the implied and objective probabilities, logistic regressions have been used in previous academic research mostly for the purposes of evaluating the efficiency of the betting markets. These studies include Pope and Peel (1989), Forrest and Simmons (2000), Vlastakis et al. (2009), Hvattum and Arntzen (2010), Koning (2012), Nyberg (2014) and Tiitu (2016). All of these studies mentioned handle the efficiency of the soccer betting markets.

4.2.5. Returns of betting based on odds and their movement

The returns of betting based on odds and their movement are analyzed in two parts. First, returns of basic simple strategies are calculated to obtain a baseline for further analysis of the data. These simple strategies include betting solely on home or away and favorite or underdog teams.

Then, the returns are calculated for each possible interval between the 5 time points, forming a total of 10 intervals. The returns are grouped based on whether the odds have been moving up or down for a team in a given interval, and further, returns are also reported based on

favorite/underdog status of the teams. The returns that can be considered to be somewhat consistent throughout the sample are taken into analysis of staking strategies.
So called unsophisticated staking strategies applied will consist of staking a fixed 1% of the original bankroll on each bet, staking to gain a net win of 1% of the original bankroll with every winning bet and staking 1% of the current bankroll at the time of placing each bet. Kelly criterion, as described in section 2.4.5, is applied as a sophisticated staking strategy. Besides the original Kelly staking, fractional Kelly is applied using 2, 3, 4, 5 and 10 as the divisors.

4.3. Assumptions

This section presents the underlying assumptions for carrying out the tests in the study. The first and second one are related to the analysis of prediction accuracy and third and fourth are related to betting strategies.

4.3.1. Bookmaker odds setting

It’s an underlying assumption of this study that the odds are set by the bookmaker as a prediction of the actual outcome of the event. Bookmakers do not intentionally set inefficient odds to exploit known bettor biases but aim to maximize their profits in the long run by providing accurate odds. This strategy for odds setting is beneficial for bookmakers as discussed by Paul and Weinbach (2012). Setting the odds as a forecast allows the bookmaker to earn the bookmaker margin in the long run without having to pay the transaction costs of changing the odds to neutralize the risk for individual events. Also, when odds are set as a best forecast of the actual outcome of the event, there’s less incentive for informed bettors to enter the market. Whether the bookmaker is setting the odds by trading to balance the books as suggested by traditional literature regarding sports betting (see e.g. Zuber et al., 1985; Sauer et al., 1988; Dare and MacDonald, 1996; Gray and Gray, 1997; Gandar et al., 2002) or by the ability to set the odds as a best forecast of the result, isn’t itself relevant for this study though.

4.3.2. Bookmaker margin distribution

Besides the assumption above about setting the odds as a true forecast, in order to be able to calculate the implied probabilities for different outcomes, this study assumes that the bookmaker margin for events is distributed evenly between the possible outcomes of the event. This means that for every outcome the implied probability can be calculated from the offered probability by dividing the offered probability with the bookmaker margin as in Eq. (7). If the bookmaker margin was unevenly distributed between the outcomes or if the odds were set to exploit biases in the market, it would make it significantly more complicated to assess the implied probabilities behind the offered odds.

4.3.3. A bet can be placed at closing odds

The latter empirical part of this study attempts to find profitable betting strategies only by utilizing information about odds and their movement. Several of these strategies involve placing the bet at closing odds, i.e. at the last quoted odds before the start of the game. The policies between bookmakers vary regarding the closing of NHL betting events. Some bookmakers close the events for betting at the announced start time which is most of the time at exact hours or half past (e.g. 7 PM EST or 7.30 PM EST). Pinnacle, which is the bookmaker used for this study, closes their events right before the opening faceoff the games, which is usually 6-10 minutes after the scheduled starting time.

For simplicity, this study assumes that a bet can always be placed after the last odds movement but before the closing of the betting event. In reality, this is not always possible as a bettor would have to know very accurately when a game is going to start in order to always hold off on placing the bet until the very last moments before the closing of the betting event. For reliability of this study this is a minor assumption as the very late line movements are usually small.

4.3.4. Bettor taxation

When testing for profitable betting strategies, this study assumes that the bettors don’t have to pay taxes for their betting activities or especially winnings. This holds for at least people that are residents of countries within the European Economic Area and that are placing bets at Pinnacle, or any other bookmaker that holds a gaming license within the EEA. Pinnacle, as many other bookmakers, are located and hold a gaming license at Malta. If living outside the EEA or using a bookmaker that has doesn't have license in the EEA, the effects of possible taxes must be incorporated into the calculations when assessing the profitability of a given betting strategy.

5. Results

This chapter presents the results of the analysis of prediction accuracy and betting strategies as introduced in the section 4.2 about methodology. The results will be followed by discussion in Chapter 6. First part of the results is focused on the prediction accuracy and the second part on researching possible profitable betting strategies.

5.1. Tests of prediction accuracy

In the following, I will perform the tests to compare the prediction accuracies of implied odds at different time points. As described in chapter 4, I first use Brier score to assess the prediction accuracies between time points and with different splits and groupings. Then, I will perform linear and logistic regression to further assess the prediction accuracy of the odds at different time points.

5.1.1. Brier score

The data was split in four ways to calculate the Brier score. First, by home and away team and second by favorite and underdog status. After that I divided the data into odds groups in two different ways, by equal odds intervals and to deciles.

The results, presented in Table 3, indicate that for home-away split, the odds get slightly more accurate between the opening and 12 am of the game day for all of the seasons, which is expected as the odds are adjusted based on early action in the market and possible new information. After 12 AM though, the prediction accuracy remains stable on aggregate until the closing of the betting. There's variance between seasons as in 2015-2016 the 12 AM odds have been less predictive of the actual outcome than odds at 1 hour and 15-minute marks and at closing of the event. In 2016-2017 it's been the other way around and 12 AM odds have been the most accurate at predicting the actual outcome. In 2017-2018 the predictive power has stayed almost exactly the same from the 12 AM until the closing of the betting.

For favorite-underdog split the results are more interesting. As expected, for all of the seasons, the opening odds are the least predictive of the actual outcome. The closing odds aren't the most predictive odds though as in aggregate the odds 15 minutes before the start of the game are the most predictive of the actual outcome of the matches. This holds for 2015-2016 and 2017-2018 seasons individually too but in 2016-2017 the odds at 12 am ET on game day were the most predictive. The differences are very minor in general though. None of the results for individual seasons are even close to being statistically significant and the same holds on the aggregate level as the lowest p-value is 0.68. The Brier scores for favorite-underdog split are presented in Table 4.

Table 3
Brier scores with home/away split

Table presents the Brier Scores calculated as in Eq. (13) for a split based on home/away team status. The scores have been calculated for every time point separately, per season and on aggregate. The best values for each row are bolded.

	Open	12:00 AM	1 hour	15 mins	Close
2015-2016	0.2434	0.2428	0.2423	0.2423	**0.2422**
2016-2017	0.2386	**0.2377**	0.2383	0.2381	0.2381
2017-2018	0.2385	0.2382	0.2382	**0.2381**	0.2383
Total	0.2401	**0.2395**	0.2396	**0.2395**	**0.2395**

Table 4
Brier scores with favorite/underdog split

Table presents the Brier Scores calculated as in Eq. (13) for a split based on favorite/underdog status of the teams. The scores have been calculated for every time point separately, per season and on aggregate. The best values for each row are bolded.

	Open	12:00 AM	1 hour	15 mins	Close
2015-2016	0.2432	0.2427	**0.2420**	**0.2420**	0.2421
2016-2017	0.2384	**0.2375**	0.2378	0.2377	0.2379
2017-2018	0.2382	0.2379	0.2379	**0.2375**	0.2382
Total	0.2399	0.2394	0.2392	**0.2391**	0.2394
Number of games	3894	3919	3902	3922	3932

Closing odds not being the most predictive is in contrast with previous research, mainly done for pari-mutuel betting events, where the final odds have been found to be significantly more predictive than earlier odds (e.g. Gramm et al., 2016). Also, as the amount of information increases all the time until the closing of the betting event, it would be a very intuitive inference that the closing odds are the most accurate prediction of the actual outcome of the event as all the information that ever is available during the lifecycle of the betting event should be incorporated in the odds at that point. For example Tiitu (2016) bypasses the question of prediction accuracy at different time points in his extensive study of betting market efficiency for soccer by stating that "intuitively, closing odds should be the most efficient odds amongst the pre-match odds at different points in time." He points out that closing odds could be assumed to best reflect the news, statistics and the sentiment of betting public regarding the game in question. The data here shows that this intuitive inference doesn't unambiguously hold true for NHL moneyline betting although the found differences in Brier scores for home-away and favorite-underdog splits are not statistically significant enough to make any other inferences either.

For odds interval groups, the closing odds are again the best predictor only for the group that includes the teams with an implied probability of 0.70 or more. The group had only 95 games in it at the Close time point. Brier score improves between odds groups at each time point except for the 0.70- group at Open, 12 AM and the 15 mins time points. This is most likely due to the small sample of games in the 0.70- group and the difference is not statistically significant. From the scores it can be seen that even the group with implied probabilities in the range of 0.50-0.55 has Brier scores lower than 0.25 which is the constant score for a 50-50 probability event. This means that there's predictive power in the odds compared to the actual game outcomes. The Brier scores for odds interval groups are presented in Table 5.

Table 5
Brier scores for odds interval groups

Table presents the Brier Scores calculated as in Eq. (13) for a split into groups based on the implied probability of teams winning at the given time point. The scores have been calculated for every time point separately. The best values for each row are bolded.

Implied Probability	Open	12:00 ap.	1 hour	15 mins	Close	n
0.70-	0.2250	0.2194	0.2079	0.2177	**0.2003**	95
0.65-0.70	0.2196	0.2157	**0.2135**	0.2139	0.2202	325
0.60-0.65	**0.2301**	0.2306	0.2350	0.2347	0.2320	816
0.55-0.60	0.2430	**0.2415**	0.2420	0.2419	0.2431	1348
0.50-0.55	0.2486	0.2487	0.2479	**0.2471**	0.2475	1404

To form equal sized groups, the games were divided into deciles. First eight of the deciles have 399 games in the group and the last two have 398. The games are sorted based on the implied probability and a lower decile number indicates higher implied probability for a team to win. The results for these deciles are very scattered. Odds 15 minutes before the start of the match are the most accurate predictor for 4 of the 10 groups while other time points are the best predictor for a maximum of 2 groups. There's a clear trend that the Brier score decreases, i.e. improves, the higher the implied probability is. When moving from a higher decile number to a lower one, the Brier score improves in 34 out of the 50 cases. As the odds are sorted based on the implied probability at the closing of the event, it's interesting to note that the predictions for games in Decile 10 have a Brier score higher than 0.25 (which is the fixed score of predicting an occurrence of 50-50 probability distribution) at Open, 12 AM, 1 hour and Close time points. All the scores for the 10 deciles are presented in Table 6.

Table 6
Brier Scores for deciles

Table presents the Brier Scores calculated as in Eq. (13) for deciles based on the implied probability of teams winning at the given time point. The lowest implied probability for every decile is reported in the column Lower threshold. The scores have been calculated for every time point separately. The best Brier score for each decile is bolded.

Decile #	Lower threshold	Open	12:00 ap.	1 hour	15 mins	Close
Decile 1	0.6512	0.2152	0.2142	**0.2137**	0.2138	0.2139
Decile 2	0.6217	0.2331	0.2327	**0.2318**	0.2319	0.2322
Decile 3	0.6020	**0.2316**	0.2320	0.2326	0.2327	0.2324
Decile 4	0.5861	0.2341	0.2342	0.2338	0.2345	**0.2335**
Decile 5	0.5704	0.2438	0.2439	0.2442	**0.2437**	**0.2437**
Decile 6	0.5570	0.2469	**0.2467**	0.2490	0.2488	0.2493
Decile 7	0.5420	0.2462	**0.2450**	0.2455	0.2455	0.2458
Decile 8	0.5284	0.2470	0.2480	0.2471	**0.2465**	0.2468
Decile 9	0.5173	0.2493	0.2478	0.2483	**0.2477**	0.2479
Decile 10	0.5000	0.2542	0.2511	0.2501	**0.2498**	0.2501

5.1.2. Linear regression

Linear regression was performed on odds groups with two different sized groupings. After sorting the data based on the implied probabilities, the matches were divided into 40 and 80 groups. The regressions provide information about the relative ability of the odds to predict actual match outcomes but also suggests contradicting results compared to previous research on NHL betting market in general. As mentioned in section 3.4, Woodland and Woodland (2001), Gandar et al. (2004) and Paul and Weinbach (2012) had found evidence of Reverse Favorite-Longshot Bias (RFLB) in NHL betting market. The results of the linear regression indicate though that the coefficient β is significantly larger than one for all of the time points, which would mean that there's actually a Favorite-Longshot Bias (FLB) in the market. The results for regression run with 40 groups showed varying results based on by what time point the odds were sorted on. The lowest β was at the 15-minute time point twice and once at the Open, 1 hour and Close points. The lowest R^2 was twice at the Close and once at the Open, 12 AM and 15-minute points. The scattered results make interpretations based on the regression difficult but by simply looking at the average α, β and R^2 values for these five regressions, the Close time point has the α closest to 0, β closest to 1 and highest R^2. The averages of $\alpha's$, $\beta's$ and R^2's for regressions with 40 groups are presented in Table 7. The complete tables with OLS estimates of the said regressions can be found in Appendix A.

Table 7
Average values of regressions with 40 groups

This table presents the mean α's, β's and R^2's of the linear regressions with 40 groups for each time point. The best value for each row is bolded.

	Open	12:00 ap.	1 hour	15 mins	Close
α	-0.1275	-0.1123	-0.0939	-0.0906	**-0.0897**
β	1.2278	1.2029	1.1692	1.1634	**1.1619**
R^2	0.8238	0.8224	0.8236	0.8233	**0.8244**

For the regressions run with 80 groups, the results were similarly scattered. Close time point had two highest R^2's and one lowest β as did also the Open time point. 15 minutes before the game time had two lowest β's, 1-hour point had one lowest β and 12 AM point had one highest R^2. When looking at the average α's, β's and R^2's for both grouping methods, Close point has the α closest to 0 and β closest to 1 for both and also the highest R^2 for the regression of 40 groups. The highest average R^2 for the regression 80 groups was at the Open time point. The averages of α's, β's and R^2's for regressions with 80 groups are presented in Table 8. The OLS estimates of the said regressions can be found in Appendix B.

Table 8
Average values of regressions with 80 groups

This table presents the mean α's, β's and R^2's of the linear regressions with 80 groups for each time point. The best value for each row is bolded.

	Open	12:00 ap.	1 hour	15 mins	Close
α	-0.1242	-0.1085	-0.0898	-0.0866	**-0.0855**
β	1.2216	1.1960	1.1618	1.1562	**1.1542**
R^2	**0.6651**	0.6624	0.6625	0.6623	0.6630

5.1.3. Logistic regression

Logistic regression was run individually on each time point. Based on Wald statistics, implied odds at each of the time points were statistically significant for the model predicting the actual outcomes. The Cox&Snell R^2 values though are quite low but similar between the time points. The highest Wald statistic figures and Cox & Snell R^2's, indicating better explanatory value, are at 12 AM and 15 minutes points. The differences to 1 hour and Close points are very minor once again though. The gap for Open time point is slightly bigger, but still relatively small also. Exp(β), which is the exponentiation of β, a measure that estimates how much an increase of 1 percent in the implied probability affects the objective probability, again indicates that there's a Favorite-longshot bias present in the market as the values vary between 1.048 and 1.050. The Exp(β) values closest to 1 are at Open, 1 hour and Close time points but the differences between all of the time points are very small as all of the values are within 0.002 of each other. The results for logistic regressions are presented in Table 9.

Table 9
Results of the binary logistic regression tests for individual time points

This table presents the maximum likelihood estimates for coefficients α and β of the logistic regression tests based on which the probabilities for each Y = 1 can be calculated as in Eq. (15). The regression has been run separately for odds at each time point. Exp(β) is the exponentiation of coefficient β, indicating the rate of change for the logistic model when the underlying implied probability changes. Wald's statistic refers to the Wald's chi-square test statistic for the estimated model, indicating the significance of the input variables. Cox&Snell R^2 is another goodness of fit measure based on the log likelihood for the model compared to the log likelihood for a baseline model. The p-values of coefficients are reported in the parentheses. The best values for Exp(β), Wald and Cox & Snell R^2 have been bolded.

	Open	**12:00 ap.**	**1 hour**	**15 minutes**	**Close**
α	-2.374	-2.484	-2.394	-2.402	-2.379
	(0.236)	(0.236)	(0.230)	(0.229)	(0.228)
β	0.047	0.049	0.047	0.047	0.047
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Exp(β)	**1.048**	1.050	**1.048**	1.049	**1.048**
Wald	120.072	**130.417**	128.550	**130.585**	129.477
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Cox & Snell R^2	0.031	**0.034**	0.033	**0.034**	0.033
n	3988	3988	3988	3988	3988

5.2. Returns of betting strategies based on odds and their movement

This section displays the results of analysis on the betting strategies based on the odds and/or their movement. The methods were presented in section 4.2.5. First, returns are presented on aggregate level for betting only on home/away or favorite/underdog teams to get an overall view of the market. Then, an analysis is done about the returns of placing bets based on odds movement between different time points. For all the consistent and potentially positive betting strategies, the returns are calculated with Kelly staking, unit win staking and variable staking in addition to the one-unit fixed staking, that's used as a basis for all of the return calculations. For this part of the study, offered odds and probabilities are used instead of fair odds and implied probabilities.

5.2.1. Returns on simple splits

To get an overview about the market in terms of betting returns, the data was again first split based on home and away teams and also based on favorites and underdogs. Betting blindly on home or away teams yields a profit quite close to the average bookmaker margin of 2.4% as can be seen in Table 10 that presents all the results of simple splits. The largest difference in aggregate between the returns for home and away teams was the 0.008 percentage point gap when betting with the opening odds. Betting on away teams offers better returns at all of the time points but that's mostly due to the 2015-2016 when away teams clearly outperformed the expectations and yielded a profit between 0.01153 and 0.02185 depending on the time of betting. In the 2016-2018 seasons home teams outperformed away teams in terms of betting returns. Based on the results it can be interpreted that the market estimates the advantage of playing at home quite well.

There's a clear difference between the returns on favorites and underdogs. The returns on favorite teams are better than the returns on away team for 14 of the 15 time points in the three seasons. The only time when the returns on underdogs exceed the returns on favorites is the opening odds in the 2016-2017 season where betting on underdogs yields a return of -0.02342 and betting on favorites yields a return of -0.03532. The odds movement up until 12:00 am (EST) on game days already turns this around in this season also and the returns are -0.05579 and -0.00755 at that point, respectively. Betting on every favorite at all of the time points of every season yields an average return of -0.00221, meaning that a bettor has an ROI of 99.8% by betting blindly on favorites.

In a sample of almost 4000 games, this further supports the notion made in section 5.1.3 that during these three seasons, there has been a favorite-longshot bias present in the market. Also, this raises a question about the assumption of section 4.3.2 that the bookmaker margin is be distributed evenly across both sides of the event, which in turn would significantly affect the analysis of predictive ability in section 5.1 as the implied probabilities would change in that case. According to the data it seems that the bookmaker margin hasn't been distributed evenly for both sides of the betting events but it rather appears that the implied probability of the favorite odds is quite an accurate predictor of the outcome and almost all of the bookmaker margin has been allocated to the underdog side of the events.

Table 10
Returns of betting on Home/Away and Favorite/Underdog teams

This table presents the returns for betting simply on every Home/Away or Favorite/Underdog team in the sample. The returns are calculated separately for every time point and presented season-by-season and on aggregate.

	Open	12:00 AM	1 hour	15 mins	Close
2015-2016					
Home	-0.06340	-0.05836	-0.05778	-0.05744	-0.05802
Away	0.02185	0.01411	0.01153	0.01168	0.01243
Favorite	0.00502	0.00185	-0.00655	-0.00048	-0.00502
Underdog	-0.04640	-0.04605	-0.03966	-0.04524	-0.04051
2016-2017					
Home	-0.01356	-0.01383	-0.01174	-0.01121	-0.01052
Away	-0.04503	-0.04932	-0.04731	-0.04845	-0.04939
Favorite	-0.03532	-0.00755	-0.00536	0.00434	-0.00690
Underdog	-0.02342	-0.05579	-0.05391	-0.06415	-0.05312
2017-2018					
Home	-0.01492	-0.01612	-0.01745	-0.01782	-0.01705
Away	-0.04372	-0.04438	-0.04134	-0.04135	-0.04091
Favorite	0.00821	0.00409	0.00824	0.00075	0.00058
Underdog	-0.06707	-0.06477	-0.06717	-0.06007	-0.05861
Total					
Home	-0.03049	-0.02933	-0.02889	-0.02873	-0.02843
Away	-0.02248	-0.02668	-0.02583	-0.02617	-0.02608
Favorite	-0.00727	-0.00050	-0.00110	0.00153	-0.00373
Underdog	-0.04580	-0.05563	-0.05375	-0.05652	-0.05086

5.2.2. Returns of betting based on odds movement

The return figures for betting on odds that have risen or fallen from one time point to another support the notion made in previous section, that betting on favorites is more profitable than betting on underdogs, which holds true for all of the time intervals for both rising and falling odds. Especially betting on favorites, whose odds have moved upwards, i.e. their offered and implied probability has decreased over time, seems to be a strategy that yields a positive net return. The net return is positive in 8 of the 10-time interval groups and over 1% positive for 6 of them. The groups include a lot of overlapping selections though, thus the returns of different time intervals are rather dependent on each other. In general, it seems that when basing betting decisions solely on odds movement, betting on odds that have moved up yield better returns than betting on odds that have been moving downwards, which holds true for 8 of 10 time intervals for favorites, 9 of 10 for underdogs and 8 of 10 on aggregate. This would imply that the market slightly overreacts to whatever is moving the odds, whether it's money bet on the market or additional information. Favorites, whose odds have fallen, offer better returns though than underdogs, who've had their odds increase, in 9 out of 10 interval groups. This further enforces the perception of favorite-longshot bias in the data and that FLB affects profitability more than odds movement. The returns for different time intervals are presented in Table 11.

Table 11
Returns of betting based on odds movement

This table presents the returns when betting on teams based on the movement of their odds in a given time interval. The intervals are based on columns 'From' and 'To'. 'Up' means that the team's odds have risen in the given interval, meaning that the implied probability derived from the odds has fallen and 'Down' means that the odds have fallen, i.e. implied probability has increased. Favorites are teams that have offered odds <1.95 and underdogs have offered odds of >1.95. Two teams may have identical offered odds of 1.95 and as in these situations there is no favorite or underdog, these occurrences are only included in the column 'Total'. 'Count' refers to the total amount of games that have had odds movement in the given interval.

		Up			Down			
From	**To**	**Favorite**	**Underdog**	**Total**	**Favorite**	**Underdog**	**Total**	**Count**
Open	12:00 AM	-0.00675	-0.04734	-0.03070	-0.00342	-0.05462	-0.02323	3467
Open	1 hour	**0.00397**	-0.02542	-0.01487	-0.02080	-0.06538	-0.03831	3813
Open	15 min	**0.01546**	-0.03909	-0.01569	-0.00936	-0.07966	-0.03961	3839
Open	Close	**0.00972**	-0.02450	-0.01028	-0.01855	-0.07765	-0.04355	3828
12:00 AM	1 hour	**0.01219**	-0.02850	-0.01062	-0.01950	-0.07184	-0.04227	3518
12:00 AM	15 min	**0.01119**	-0.04701	-0.02208	-0.00520	-0.06748	-0.03179	3670
12:00 AM	Close	**0.03169**	-0.02714	**0.00136**	-0.02143	-0.09834	-0.05754	3713
1 hour	15 min	-0.02703	-0.05698	-0.04251	**0.00995**	-0.02507	-0.00668	2026
1 hour	Close	**0.01730**	-0.02795	-0.00603	-0.02179	-0.07936	-0.04912	2910
15 min	Close	**0.02163**	-0.00153	**0.01247**	-0.04761	-0.07538	-0.06370	2265

When considering the time intervals for further analysis, the intervals starting one hour before the start of the game or later were omitted due to the results being inconsistent. The rising favorites are returning a loss of -2.70% from the 1-hour mark to 15-minute mark but from the 15 minute mark to the closing of the event the same strategy are showing a profit of 2.16%. Also, the interval from the opening of the betting to 12 AM of the game day was omitted due to negative returns. Thus, the six intervals for which Kelly staking is applied to, are Open to 1 hour/15 minutes/Close and 12 AM to 1 hour/15 minutes/Close. In theory, using the interval of 12AM to Close will yield the highest profits of 3.17%. It must be noted though that from a practical point of view, out of these time points, interval ending at the 15-minute point would be optimal for a betting strategy in the sense that Pinnacle's betting for NHL games is closed only right before the opening faceoffs, which is usually 6-10 minutes after the announced starting time. Most of the European bookmaking companies close the events already at the announced starting time, which means that if using Pinnacle's closing odds as the benchmark, the bettor won't have the possibility of utilizing all of the possible bookmakers for finding the highest odds available, which is what a sharp bettor would wish to do.

The analysis on the returns of different staking plans reveals that for this kind of strategy where there are no individual profit estimates for each game, but the expected value is derived from the results, simple staking plans actually provide better and steadier returns than using Kelly criterion. This notion was further backed by the correlation figures between the perceived expected value based on the odds movement and the actual returns. The correlations are presented in Table 12. The highest correlations are at 12 AM to 1 hour, 15 minutes to Close and 12 AM to Close intervals with 0.2140, 0.1760 and 0.1668, respectively. This is far from the correlation figures of implied odds and actual results that ranged between 0.889 and 0.920 for different time points. So, although betting on odds that have risen offers higher returns than betting on falling odds, the higher returns aren't proportionate to the odds movement. Staking to win a certain amount on each bet yields the best return for all of the intervals starting at the opening of the betting events and staking a flat amount on each bet yields the best return for intervals starting at 12 AM. The returns for different staking plans for the most profitable interval, 12 AM to Close, are presented in Table 13. The return tables for the rest of the intervals that were analyzed further can be found in Appendix C.

Table 12
Correlations between the expected value of bets based on odds movement and the actual returns

This table presents the correlations of the expected value of the bets based on odds movement and the actual returns of the bets. The correlation is calculated for each interval separately. Positive returns based on odds movement imply that the odds at the start of an interval are considered to be representative of the true probability and rising odds offer a return that's higher than the profit margin of the bookmaker. The expected values are calculated like the perceived edge for value bets in Eq. (9).

From	To	Correlation
Open	12:00 AM	-0.0795
Open	1 hour	0.0548
Open	15 min	0.0258
Open	Close	0.0514
12:00 AM	1 hour	0.2140
12:00 AM	15 min	0.1486
12:00 AM	Close	0.1668
1 hour	15 min	-0.0886
1 hour	Close	0.0733
15 min	Close	0.1760

Table 13
Returns of rising favorites in 12 AM to Close interval with different staking plans

This table presents the results of different staking plans for betting on favorites with rising odds in the interval from 12 AM to closing of the betting events. Fixed staking 1% refers to staking 1% of the original bankroll for each bet, and win 1% means betting to win 1% of the original bankroll, regardless of the previous results. Percentage staking refers to betting 1% of the current bankroll for every bet. Kelly/X's refer to the Kelly criterion calculated as in Eq. (12) using the average return of the interval as the perceived edge. X is the divisor for the fractional Kelly. The best value of each column is bolded.

Staking strategy	ROI	Average stake	Ending bankroll	Lowest bankroll	Highest bankroll
Fixed staking 1%	**3.17 %**	1.00 %	1.4957	0.9310	1.5751
Win 1%	2.97 %	1.47 %	1.6804	0.8608	1.7962
Percentage staking	3.02 %	1.17 %	1.5535	0.9222	1.6834
Kelly/1	1.91 %	**5.80 %**	**2.7306**	0.5069	**4.0235**
Kelly/2	2.60 %	2.98 %	2.2076	0.7560	2.6657
Kelly/3	2.75 %	1.87 %	1.8072	0.8409	2.0469
Kelly/4	2.82 %	1.35 %	1.5963	0.8825	1.7519
Kelly/5	2.85 %	1.05 %	1.4705	0.9070	1.5837
Kelly/10	2.92 %	0.50 %	1.2266	**0.9546**	1.2727

From Table 13 it can be seen that for the 12 AM to Close interval fractional Kelly plans divided by 3 or less provide a higher ending bankroll than either flat/percentage staking or staking to win a fixed amount but that's due to higher average stakes. As there are up to 10 games in a day that fill the criteria for the strategy, staking any more than around 2% of the current bankroll per bet means that the bettor is potentially exposing over 20% of the whole bankroll at once, in one day, which is risky in the long run. Kelly/1 actually ended up with a profit of 173.06% over the three seasons but the average stake of 5.80% of the original bankroll, 4.65% adjusted to the bankroll at the time of the bets, is too high especially for a betting strategy with this kind of high volume, as longer losing streaks can lead to situations where for example one bookmaker account needs more funds but the rest of the funds are tied at other bookmaker account(s) and transferring the money takes so much time that bets are missed or they have to be placed with suboptimal odds. Kelly/1 also in this sample had a point where half of the original bankroll was lost.

Although betting 1% of the starting bankroll for each bet doesn't end up with the highest bankroll in the end, it offers the best return on each unit invested and it's one of the lower risk option of the staking plans observed here as it has the second highest low-point for bankroll only 2.36 percentage points behind Kelly/10 plan which has on average bet of 0.5% and a total return of less than half of that of 1% fixed stakes. The optimal staking plan balancing profitability and risk would seem to be betting a flat stake of around 2% of the original bankroll per bet. Fixed bets of 2% would've yielded a total bankroll growth of 99.14% in the three-season span.

6. Discussion

The results of the tests in the previous chapter indicate that in general there are little differences in the predictive power of NHL moneyline odds throughout the game day. Already Brier scores showed that the accuracy of predictions on aggregate is very even throughout the whole lifecycle of the betting events as none of the differences in the Brier scores between different time points were statistically significant. It was evident though that the opening odds on average were slightly less predictive of the final outcome of the games than the odds during game day. This is a very intuitive observation as the behavior of the market is an important factor when bookmakers and especially Pinnacle adjust their odds to reflect the best prediction of the outcome of the game. When the betting is opened for a game, there's no information about the market yet and the odds are only based on statistical data from the past, hence it's reasonable to assume that the predictions aren't quite as accurate as they're later after the odds have been already available for betting for a while. The findings are very different compared to the studies of pari-mutuel betting markets (Gramm and McKinney, 2009; Gramm et al., 2016) where the closing line is significantly the most representative of the actual results. The difference is logical though as for pari-mutuel markets it's wise for the sharp players to wait until the very last moments before the closing of the event to place the bet as at that point the most accurate estimation can be made about the potential payout, which is ultimately determined only after the event is closed. For fixed odds markets the bettor always knows the potential payout, no matter when the bet is placed, making the comparison of probability estimates and the current odds the most important factor in analysis. An idea about the probable direction of future odds movement is helpful though when determining the optimal time to place the bet.

Linear regressions on odds groups further supported the remark that there isn't a clear time point that's the most accurate at predicting game outcomes but it must be noted that on aggregate level for both odds groupings, the closing line had the β value that was closest to 1. An interesting finding was that even for closing lines, the β's were all significantly larger than 1, implying that there's a Favorite-Longshot Bias (FLB) in the Pinnacle NHL market as when the implied probability of a team winning increases by 1 percent, the objective probability increases by a minimum of 1.14%.

Indications of Favorite-Longshot Bias completely contradict with the findings in previous NHL research (e.g. Woodland and Woodland (2001), Gandar et al. (2004) and Paul and Weinbach (2012)), that all found that NHL was a market with Reverse Favorite-Longshot Bias, meaning that underdogs yield better profits than favorites. A possible reason for this is that the odds have been collected at different times and from other bookmakers than Pinnacle. Woodland and Woodland (2001) collected their data from two bookmakers operating in Nevada, and also the data is relatively old as it's from pre-online betting era, 1990-1996. Gandar et al. (2004) used the same data set as Woodland and Woodland but improved the quality of data by correcting the calculation of bookmaker commission. Paul and Weinbach (2012) used a dataset of three seasons between 2005 and 2008, collected from sports betting information provider called Sports Insights. The dataset included odds data from four US online bookmakers. It's possible that Pinnacle is setting their odds in a different way compared to US-based bookmakers or alternatively the market has changed in the time between when the data was collected for the earlier studies and for this study (7 years for the study of Paul and Weinbach, 19 years for the studies of Woodland and Woodland, and Gandar). It must be noted that also Pinnacle did actually operate in the United States and did so also partly during the time of Paul and Weinbach's period of observation, but they withdrew from the market at the beginning of 2007 due to changes in legislation.

The logistic regression also produced consistent results with earlier tests. The regressions were run individually for results at each time point and the Exp(β) values were very similar for all of the time points, all within 0.002 of each other. The values also implied again about the presence of Favorite-Longshot Bias as the range of them was 1.048 to 1.050. At every individual time point, the implied probabilities were highly predictive of the objective probabilities as the Wald statistic had a p-value of < 0.000 for all of the time points.

The analysis of betting returns based on simple splits between home and away teams, and favorites and underdogs, again highlights that the FLB exists, as betting blindly on favorites returned only a net loss of 0.22% on aggregate throughout the whole dataset of this study. As the average bookmaker margin was around 2.4%, it seems that over 90% of the bookmaker margin has been allocated in the odds of the underdog side. This also raises the question whether it was the best method to calculate the implied probabilities for the purposes of this study by allocating the bookmaker margin evenly to odds of both the favorite and underdog sides. In previous research though, there's no other widely adopted method for allocating the margin.

The analysis of betting based on odds movement revealed that betting on favorites that have their odds rising is clearly the most profitable of the strategies. There were 10 time intervals, 8 of which yielded a profit when applying the strategy of betting on favorites with rising odds between the time points. None of the profits for the intervals were statistically significant but 6 intervals were significantly higher than the average bookmaker margin of 2.4% at 10% level.

The only intervals that didn't show a profit were Open to 12AM and 1 hour to 15 minutes. The Open to 12 AM interval is logically explained by the odds not being the sharpest right when they're opened, and they should be moving in the right direction right after they're published. This is backed up by the fact that for that interval favorites whose odds have fallen are yielding a better return than favorites whose odds have risen. Also, as a whole, teams with falling odds have offered better returns than teams with rising odds at that interval. The other time interval that didn't show a profit, 1 hour to 15 minutes before the start of the game, makes the profit figures based on the odds movement within the last hour before the opening faceoff very two-pronged. The 1 hour to 15 minutes interval was the worst of all intervals for favorites that had their odds rising. The interval yielded a loss of 2.70% across the 2026 matches that had odds movement during that time frame. From the 15 minute mark up until the opening faceoff was the second best interval though, showing a profit of 2.16%, which lifts the aggregate interval of 1 hour to Closing of the betting event to be profitable with an ROI of 1.70%. The 1 hour to 15 minutes interval is the worst for not only rising favorites but also for rising underdogs and for all rising teams combined. The difference in returns between that interval and other intervals around the said time frame, is clearly an abnormality in the data. The reasons for the finding and the exact time interval when the different behavior of the market happens would need to be researched further to get concrete answers. As such it seems though that there's a time between the 1-hour mark and 15-minute mark before the start of the games when the market is working in a different way than at other times. It's only guessing, but it's possible that some significant NHL bettors, individual or institutional, place their bets around the same time, less than an hour but more than 15 minutes before the start of the games. These entities in the market are by default very well informed, i.e. sharp, and can identify the value in the market. The sums that these big bettors are staking, are significant and Pinnacle also has a system in place for monitoring the behavior of successful bettors so that they can draw insights from these bets and incorporate them with their own analysis. These factors would lead to the odds decreasing as a result of the bets placed by these kind of big bettors but as there's already so much money in the market that near to the start time of the games, the line moves may not be as drastic as they would be earlier during the lifecycle of the betting events. These kinds of well-informed bets reducing the odds, could be at least a part of an explanation for the outlier bad returns of the 1 hour to 15 minutes interval.

The intervals that showed a profit for favorites with rising odds were taken up for analysis of staking strategies. It appeared that Kelly criterion, the standard staking model of most of the sophisticated bettors, wasn't optimal at least for this dataset. That could've been anticipated though as Kelly calculates the optimal stake based on the odds of the selection and based on the perceived value calculated for the particular event. The strategy in this study didn't include calculating the perceived value individually for each selection that fit the criteria, and that was also not advisable based on the correlations of the expected value derived from odds movement and the actual results, which showed that although rising odds provided positive returns, the amount of rise in the odds didn't correlate very significantly as higher returns.

7. Conclusions

This thesis was done to fill the void in academic research of odds movement in fixed odds sports betting markets. There have been several studies about the effect of late money and odds movement for pari-mutuel sports betting markets where the odds of each bet are decided only after the event is closed and all the bets are placed. These studies have found clear results that the offered probabilities, based in the odds, move towards the actual probabilities when the closing of the event gets closer. The fixed odds betting markets have been mostly ignored though.

The dataset comprised of Moneyline (the ultimate winner of the game including possible overtime and shootout) betting odds data for 3 National Hockey League (NHL) seasons, 2015-16, 16-17 and 17-18 and was collected from bookmaking company Pinnacle. The odds for both teams were collected at 5 different time points: when the betting events were opened, at 12 AM Eastern time on game day, 1 hour before the scheduled start of the match, 15 minutes before the scheduled start of the match and at the closing of the events.

The tests showed that there aren't statistically significant differences in the predictive ability of the odds between the different time points. Brier scores were able to make hardly any distinction between the time points, but linear regression was the test that gave the most logical results as on aggregate, the β of the regressions got closer to 1 as the start of the game got closer. The significant finding in this dataset though was not directly related to the movement of the odds. Several studies, including Woodland and Woodland (2001), Gandar et al. (2004) and Paul and Weinbach (2012), found that in NHL the market has a Reverse Favorite-Longshot Bias, meaning that betting on underdogs, the less likely outcomes, yields better returns than betting on favorites. The results of this study showed the complete opposite though as according to both linear and logistic regression tests, when the implied probability was rising, objective probability rose too and at a higher rate. This is a finding consistent with other sports betting markets, e.g. horse racing (Snowberg and Wolfers, 2010), soccer (Tiitu 2016) and many others. For US major league betting markets though, Reverse Favorite-Longshot Bias has been found constantly.

There was a clear best strategy for betting based on odds movement, as 8 of the 10 time intervals yielded a profit for betting on favorites that had their odds rising within the observed time interval. The best returns were at the interval from 12 AM to the closing of the betting events, which yielded an ROI of 3.17% over 1564 bets that fit the criteria. Also every other interval starting at the opening of the events or at 12 AM of the game day and ending within the last hour before the start of the game, was profitable for favorites that had their odds rising during the interval. For this dataset, sophisticated staking strategies weren't really useful, as the best ROI's were achieved by simplistic staking strategies of either betting a fixed 1% of the starting bankroll or always betting 1% of the current bankroll. With these 1% investments, these strategies would've yielded a total bankroll growth of 49.57% and 55.35%, respectively, across the dataset.

8. Ideas on future research

The area of sports betting offers a plethora of opportunities for research and there are also several ways the research subject of this study can be utilized in future research. I will present some of these ideas in this final chapter.

First of all, research could be done about other sports and other bet types. Especially other US major sports leagues NFL, NBA and MLB would be great opportunities as well as the big European soccer leagues such as English Premier League, Spanish La Liga, Italian Serie A and German Bundesliga. For hockey, over/under bets could be another bet type to research but for other sports, at least NFL, NBA and soccer, handicap bets might be even the most suitable bet type to conduct this kind of research on.

The time points at which the odds data was recorded could also be a way to expand this study. Here, there were 5 predetermined time points where the odds data was collected for each game. For a more robust view on the development of the odds, the whole odds movement data of the events, if available, could be used for the analysis. This also links with the point of the previous paragraph because in for example NFL and soccer, where the amount of games is significantly lower than in NHL, the betting events for games are regularly open for at least a week before the start of the game, so focusing this intensively on the odds movement during game day wouldn't offer great insights about the overall movement of the odds.

This study also utilizes the odds of only one bookmaker, which in theory isn't the best strategy for profitable sports betting, although professional bettors for example in Finland are forced to use Pinnacle almost solely as their bookmaker of choice. This study could be expanded by collecting the odds movement data from also other bookmakers as well as utilizing the odds of other bookmakers for determining the odds at which the bets can be placed. The choice of bookmakers would have to be careful though from a practical point of view as the odds of many bookmakers are affected by the fact that winning bettors are eventually limited to very small stakes and the losses from abnormally high odds can be negated by limiting the players that regularly take advantage of those odds. This affects both, the information that can be obtained from odds movement and the extent to which the odds can be utilized in betting if they are highest in the market.

There are also other variables related to the betting market that could be incorporated in the analysis besides odds movement. Some examples of these variables are the proportion of bets placed or money wagered on each possible outcome of the event. At least for US major sports leagues these statistics are regularly offered even free of charge. Through the analysis of these variables, it could be possible to identify for example if the market is moving the line because the general public is heavy on one side of the possible outcomes or if there's a situation where general public is on one side but significant money is wagered on the other side. This kind of discrepancy between the amount of bets and money wagered could cause the odds to move down on the selection that was backed with a significant amount of money although the absolute number of bets placed is smaller.

This research also found a Favorite-Longshot Bias in the market, contrary to the previous literature about NHL odds. The efficiency of NHL betting markets could clearly be investigated further by utilizing a dataset more extensive than the one used in this study. It could be researched if the FLB is a recent development in the market, tied specifically to the odds of Pinnacle, or both. Also, the allocation of bookmaker margin between the favorite and underdog could be studied further as in the light of results of this study it would seem that, at least in Pinnacle NHL odds, the bookmaker margin is almost entirely allocated on the underdog side of the events.

9. References

Abinzano, I., Muga, L., Santamaria R. 2019. Hidden Power of Trading Activity: The FLB in Tennis Betting Exchanges. Journal of Sports Economics. Vol. 20. 261-285.

Alberola, J.M., Garcia-Fornes, A. 2013. Using a case-based reasoning approach for trading in sports betting markets. Applied Intelligence. Vol. 38. 465–477.

Algoet, P.H., Cover, T.M. 1988. Asymptotic optimality and asymptotic equipartition properties of log-optimum investment. Annals of Probability. Vol. 16. 876–898.

Ali, M.M. 1977. Probability and utility estimates for racetrack bettors. Journal of Political Economy. Vol. 85. 803–815.

Asch, P., Malkiel, B.G., Quandt, R.E. 1982. Racetrack betting and informed behavior. Journal of Financial Economics. Vol. 10. 187–194.

Asch, P., Malkiel, B.G., Quandt, R.E. 1984. Market efficiency in racetrack betting. Journal of
Business. Vol. 57. 165–175.

Bernoulli, D. 1738. Specimen theoriae novae de mensura sortis. Commentarii Academiae Scientiarum Imperialis Petropolitanae 5, 175–192. Translated by Sommer, L.: Bernoulli, D. 1954. Exposition of a new theory on the measurement of risk. Econometrica. Vol. 22. 23–36.

Bitcoin Chaser. 2018. The Most Popular Sports to Bet on Today. [online document]. [Accessed 17.7.2019]. Available: https://bitcoinchaser.com/bitcoin-sportsbook/most-popular-sports-to-bet-on-today

Borghesi, R. 2012. The implications of a reverse favourite-longshot bias in a prediction market. Journal of Prediction Markets. Vol. 6. 12–21.

Boulier, B.L., Stekler, H.O. 2003. Predicting the outcomes of National Football League games. International Journal of Forecasting. Vol. 19. 257–270.

Breiman, L. 1961. Optimal gambling systems for favorable games. In: Neyman, J. (Ed.). Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume I: Contributions to the Theory of Statistics. Berkeley, California, USA. University of California Press. 65–78.

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

Buchdal, J. 2003. Fixed Odds Sports Betting: the Essential Guide. London, Great Britain. High Stakes Publishing. 224 p. ISBN 978-1843440192.

Busche, K., Hall, C.D. 1988. An exception to the risk preference anomaly. Journal of Business. Vol. 61. 337–346.

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

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. Vol. 55. 263–273.

Church-Sanders, R. 2011. Online Sports Betting: a Market Assessment and Outlook. London, Great Britain. iGaming BusinessLondon.

Croxson, K., Reade, J.J. 2011. Exchange vs. dealers: a high-frequency analysis of in-play betting prices. Department of Economics Discussion Paper 11–19. University of Birmingham. Birmingham.

Croxson, K., Reade, J.J. 2014. Information and efficiency: goal arrival in soccer betting. Economic Journal. Vol. 124. 62–91.

Dare, W.H., MacDonald, S.S., 1996. A generalized model for testing the home and favorite team advantage in point spread markets. Journal of Financial Economics. Vol. 40. 295–318.

Deschamps, B., Gergaud, O. 2007. Efficiency in betting markets: evidence from English football. Journal of Prediction Markets. Vol. 1. 61–73.

Direr, A. 2013. Are betting markets efficient? Evidence from European football championships. Applied Economics. Vol. 45. 343–356.

Domeneghetti, R. 2014. From the Back Page to the Front Room: Football's Journey Through the English Media. Huddersfield, Great Britain. Ockley Books. 380 p. ISBN 978-0957141056.

Dowie, J. 1976. On the efficiency and equity of betting markets. Economica. Vol. 43. 139–150.

Draftkings. 2019. Some of the most common hockey betting types. [online document]. [Accessed 14.7.2019]. Available: https://sportsbook.draftkings.com/how-to-bet/hockey-betting-guide

Finnigan, M., Nordsted, P. 2010. The Premier Football Betting Handbook 2010/11. Hampshire, Great Britain. Harriman House. 412 p. ISBN 978-1906659820.

Forrest, D., Simmons, R. 2000. Forecasting sport: the behaviour and performance of football tipsters. International Journal of Forecasting. Vol. 16. 317–331.

Franck, E., Verbeek, E., Nüesch, S. 2010. Prediction accuracy of different market structures—bookmakers versus a betting exchange. International Journal of Forecasting. Vol. 26. 448–459.

Franck, E., Verbeek, E., Nüesch, S. 2013. Inter-market arbitrage in betting. Economica. Vol. 80. 300–325.

Fulton L.V., Mendez, F.A., Bastian, N.D., Muzaffer Musal, R. 2012. Confusion Between Odds and Probability, a Pandemic? Journal of Statistics Education. Vol. 20. Issue 3.

Gandar, J.M., Zuber, R.A., Johnson, R.S., Dare, W. 2002. Re-examining the betting market on Major League Baseball games: is there a reverse favourite-longshot bias? Applied Economics. Vol. 34. 1309–1317.

Gandar, J.M., Zuber, R.A., Johnson, R.S. 2004. A reexamination of the efficiency of the betting market on National Hockey League games. Journal of Sports Economics. Vol. 5. 152–168.

Gil, R., Levitt, S.D. 2007. Testing the efficiency of markets in the 2002 World Cup. Journal of Prediction Markets. Vol. 1. 255–270.

Golec, J., Tamarkin, M. 1991. The degree of inefficiency in the football betting market. Journal of Financial Economics 30. 311–323.

Gramm, M., Ziemba, W.T. 2008. The dosage breeding theory for horse racing predictions. In: Hausch, D.B., Ziemba, W.T. (Ed.). Handbook of Sports and Lottery Markets. Amsterdam, Netherlands. North-Holland. 307–340. ISBN 978-0444507440.

Gramm, M., McKinney, C.N. 2009. The effect of late money on betting market efficiency. Applied Economics Letters. Vol. 16. 369-372.

Gramm, M., McKinney, C.N., Parker, R.E. 2016. Late money and betting market efficiency: Evidence from Australia. The Journal of Gambling Business and Economics. 2016. Vol. 10. 11- 21.

Gray, P.K., Gray, S.F. 1997. Testing market efficiency: evidence from the NFL sports betting market. Journal of Finance. Vol. 52. 1725–1737.

Griffith, R.M. 1949. Odds adjustments by American horse-race bettors. American Journal of Psychology. Vol. 62. 290–294.

Hausch, D.B., Ziemba, W.T., Rubinstein, M. 1981. Efficiency of the market for racetrack betting. Management Science. Vol. 27. 1435–1452.

Hvattum, L.M., Arntzen, H. 2010. Using ELO ratings for match result prediction in association football. International Journal of Forecasting. Vol. 26. 460–470.

Jullien, B., Salanié, B. 2000. Estimating preferences under risk: the case of racetrack bettors. Journal of Political Economy. Vol. 108. 503–530.

Kelly, J.L. 1956. A new interpretation of information rate. Bell System Technical Journal. Vol. 35. 917–926.

Koning, R. 2012. Regression tests and the efficiency of fixed odds betting markets. International Journal of Sport Finance. Vol. 7. 262–274.

Kotlyar, V., Smyrnova, O. 2012. Betting market: analysis of arbitrage situations. Cybernetics and Systems Analysis. Vol. 48. 912–920.

Laffey, D. 2005. Entrepreneurship and innovation in the UK betting industry: the rise of person-to-person betting. European Management Journal. Vol. 23. 351–359.

Levitt, S.D. 2004. Why are gambling markets organised so differently from financial markets? Economic Journal. Vol. 114. 223–246.

MacLean, L.C., Ziemba, W.T., Blazenko, G. 1992. Growth versus security in dynamic investment analysis. Management Science. Vol. 38. 1562–1585.

MacLean, L.C., Ziemba, W.T. 2006. Capital growth: theory and practice. In: Zenios, S.A., Ziemba, W.T. (Ed.). Handbook of Asset and Liability Management, Vol. 1: Theory and Methodology. Amsterdam, Netherlands. North-Holland. 429–473. ISBN 978-0444508751.

MacLean, L.C., Thorp, E.O., Ziemba, W.T. 2010. Long-term capital growth: the good and bad properties of the Kelly and fractional Kelly capital growth criteria. Quantitative Finance. Vol. 10. 681–687.

Nyberg, H. 2014. A multinomial logit-based statistical test of association football betting market efficiency. HECER Discussion Paper No. 380. Helsinki Center of Economic Research. Helsinki.

Odachowski, K., Grekow, J. 2012. Using Bookmaker Odds to Predict the Final Result of Football Matches. In: Grana, M., Toro, C., Howlett, R.J., Lakhmi, C.J. (Ed.). Knowledge Engineering, Machine Learning and Lattice Computing with Applications. 16th International Conference on Knowledge-Based and Intelligent Information and Engineering Systems. San Sebastian, Spain, September 10-12, 2012. Springer. 196-205. ISBN 9783642373435.

Oikonomidis, A., Johnson, J. 2011. Who can beat the odds? The case of football betting reviewed. In: Vaughan Williams, L. (Ed.). Prediction Markets: Theory and Applications. Abingdon, Great Britain. Routledge. 204–220. Routledge International Studies in Money and Banking (Book 66). ISBN 978-0415572866.

Opta. 2019. Advanced metrics. [online document]. [Accessed 23.7.2019]. Available: https://www.optasports.com/services/analytics/advanced-metrics/

Paton, D., Vaughan Williams, L. 1998. Do betting costs explain betting biases? Applied Economics Letters. Vol. 5. 333–335.

Paul, R.J., Weinbach, A.P. 2005. Market efficiency and NCAA college basketball gambling. Journal of Economics and Finance. Vol. 29. 403–408.

Paul R.J., Weinbach A.P. 2007. Does Sportsbook.com set pointspreads to maximize profits? Tests of the Levitt model of sportsbook behavior. Journal of Prediction Markets. Vol. 1. 209–218.

Paul R.J., Weinbach A.P. 2008. Price setting in the NBA gambling market: tests of the Levitt model of sportsbook behavior. International Journal of Sports Finance. Vol. 3. 2–18.

Paul, R.J., Weinbach, A.P. 2012. Sportsbook pricing and the behavioral biases of bettors in the NHL. Journal of Economics and Finance. Vol. 36, 123–135.

Pope, P.F., Peel, D.A. 1989. Information, prices and efficiency in a fixed-odds betting market. Economica. Vol. 56. 323–341.

Sauer, R.D., Brajer, V., Ferris, S.P., Marr, M.W. 1988. Hold your bets: another look at the efficiency of the gambling market for National Football League games. Journal of Political Economy. Vol. 96. 206–213.

Smith, M.A., Paton, D., Vaughan Williams, L. 2006. Market efficiency in person-to-person betting. Economica. Vol. 73. 673–689.

Smith, M.A., Paton, D., Vaughan Williams, L. 2009. Do bookmakers possess superior skills to bettors in predicting outcomes? Journal of Economic Behavior and Organization. Vol. 71. 539–549.

Snowberg, E., Wolfers, J. 2010. Explaining the favorite-long shot bias: is it risk-love or misperceptions? Journal of Political Economy. Vol. 118. 723–746.

Snyder, W.W. 1978. Horse racing: testing the efficient markets model. Journal of Finance. Vol. 33. 1109–1118.

Sportsbook Review. 2019. Sportsbook Rating Guide. [online document]. [Accessed 11.11.2019]. Available: https://www.sportsbookreview.com/betting-sites/#no-filters

Tax, N., Joustra, Y. 2015. Predicting The Dutch Football Competition Using Public Data: A Machine Learning Approach. Research Paper. University of Twente. Department of Electrical Engineering, Mathematics and Computer Science. Enschede. DOI 10.13140/RG.2.1.1383.4729.

Tiitu, T. 2016. Abnormal returns in an efficient market? Statistical and economic weak form efficiency of online sports betting in European soccer. Master's thesis. Aalto University School of Business. Department of Finance. Helsinki. 122 p.

Thorp, E.O. 2006. The Kelly criterion in blackjack, sports betting and the stock market. In: Zenios, S.A., Ziemba, W.T. (Ed.). Handbook of Asset and Liability Management, Vol. 1:

Theory and Methodology. Amsterdam, Netherlands. North-Holland. 385-428. ISBN 978-0444508751.

Vaughan Williams, L., Paton, D. 1997. Why is there a favourite-longshot bias in British racetrack betting markets? Economic Journal. Vol. 107. 150–158.

Veikkaus. (2018). Vuosiraportti 2018. [online document]. [Accessed 10.7.2019]. Available: https://cms.veikkaus.fi/site/binaries/content/assets/dokumentit/vuosikertomus/2018/veikkaus_vuosiraportti_2018.pdf

Ville, J. 1939. Étude critique de la notion de collectif. Monographies des Probabilités. Vol. 11. 824-825.

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. Vol. 28. 426–444.

Woodland, L.M., Woodland, B.M. 1994. Market efficiency and the favorite-longshot bias: the baseball betting market. Journal of Finance. Vol. 49. 269–379.

Woodland, L.M., Woodland, B.M. 2001. Market efficiency and profitable wagering in the National Hockey League: can bettors score on longshots? Southern Economic Journal. Vol. 67. 983–995.

Woodland, L.M., Woodland, B.M. 2003. The reverse favourite-longshot bias and market efficiency in Major League Baseball: an update. Bulletin of Economic Research. Vol. 55. 113–123.

Zuber, R.A., Gandar, J.M., Bowers, B.D. 1985. Beating the spread: testing the efficiency of the gambling market for National Football League games. Journal of Political Economy. Vol. 93. 800–806.

10. Appendices

10.1. Appendix A: Results of linear regressions with 40 odds groups

Table A.1
Linear regression sorted by open with 40 groups

This table presents the results of the standard linear regression model, as defined in Eq. (14), for 40 equally sized odds groups. The regressions have been run separately for each time point. The groups are sorted by implied probability, as defined in Eq. (7), at the opening of the betting events, so that the group size is 100 for 28 and 99 for 12 groups. The p-values of the estimated coefficients are reported in parentheses. The sixth and seventh rows give the F-statistics for the joint test that α = 0 and β = 1. ***, ** and * indicate significance at the 1%, 5% and 10% level respectively. The best values of β and R^2 are bolded.

	Sorted by Open				
	Open	**12:00 AM**	**1 hour**	**15 mins**	**Close**
α	-0.0811*	-0.1004**	-0.0946*	-0.0920	-0.0915
	(0.0860)	(0.0408)	(0.0512)	(0.0585)	(0.0580)
β	**1.1435*** **	1.1811***	1.1706***	1.1660***	1.1653***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
R^2	0.8335	**0.8337**	0.8335	0.8314	0.8333
F	190.29***	190.48***	190.25***	187.38***	189.89***
(p-value)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
n	40	40	40	40	40

Table A.2
Linear regression sorted by 12 AM with 40 groups

This table presents the results of the standard linear regression model, as defined in Eq. (14), for 40 equally sized odds groups. The regressions have been run separately for each time point. The groups are sorted by implied probability, as defined in Eq. (7), at 12 AM of the game day, so that the group size is 100 for 28 and 99 for 12 groups. The p-values of the estimated coefficients are reported in parentheses. The sixth and seventh rows give the F-statistics for the joint test that α = 0 and β = 1. ***, ** and * indicate significance at the 1%, 5% and 10% level respectively. The best values of β and R^2 are bolded.

	Sorted by 12am				
	Open	**12:00 AM**	**1 hour**	**15 mins**	**Close**
α	-0.1313**	-0.1067**	-0.1029**	-0.1003*	-0.1015**
	(0.0150)	(0.0390)	(0.0438)	(0.0503)	(0.0449)
β	1.2346***	1.1926***	1.1855***	**1.1811*** **	1.1835***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
R^2	0.8230	0.8219	0.8236	0.8206	**0.8252**
F	176.65***	175.42***	177.38***	173.87***	179.40***
(p-value)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
n	40	40	40	40	40

Table A.3
Linear regression sorted by 1 hour with 40 groups

This table presents the results of the standard linear regression model, as defined in Eq. (14), for 40 equally sized odds groups. The regressions have been run separately for each time point. The groups are sorted by implied probability, as defined in Eq. (7), at 1 hour before the scheduled start of the game, so that the group size is 100 for 28 and 99 for 12 groups. The p-values of the estimated coefficients are reported in parentheses. The sixth and seventh rows give the F-statistics for the joint test that α = 0 and β = 1. ***, ** and * indicate significance at the 1%, 5% and 10% level respectively. The best values of β and R^2 are bolded.

	Sorted by 1 hour				
	Open	**12:00 AM**	**1 hour**	**15 mins**	**Close**
α	-0.1397**	-0.1160**	-0.0863*	-0.0864*	-0.0863*
	(0.0106)	(0.0274)	(0.0856)	(0.0844)	(0.0856)
β	1.2499***	1.2096***	**1.1553*** **	1.1558***	1.1557***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
R^2	**0.8242**	0.8221	0.8188	0.8196	0.8188
F	178.13***	175.62***	171.70***	172.67***	171.68***
(p-value)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
n	40	40	40	40	40

Table A.4
Linear regression sorted by 15 minutes with 40 groups

This table presents the results of the standard linear regression model, as defined in Eq. (14), for 40 equally sized odds groups. The regressions have been run separately for each time point. The groups are sorted by implied probability, as defined in Eq. (7), at 15 minutes before the scheduled start of the game, so that the group size is 100 for 28 and 99 for 12 groups. The p-values of the estimated coefficients are reported in parentheses. The sixth and seventh rows give the F-statistics for the joint test that α = 0 and β = 1. ***, ** and * indicate significance at the 1%, 5% and 10% level respectively. The best values of β and R^2 are bolded.

	Sorted by 15 minutes				
	Open	**12:00 AM**	**1 hour**	**15 mins**	**Close**
α	-0.1440**	-0.1190**	-0.0924*	-0.0865*	-0.0874*
	(0.0132)	(0.0344)	(0.0808)	(0.0969)	(0.0935)
β	1.2577***	1.2151***	1.1664***	**1.1558*** **	1.1577***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
R^2	0.8069	0.8020	0.8059	0.8072	**0.8076**
F	158.80***	153.97***	157.73***	159.13***	159.51***
(p-value)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
n	40	40	40	40	40

Table A.5
Linear regression sorted by Close with 40 groups

This table presents the results of the standard linear regression model, as defined in Eq. (14), for 40 equally sized odds groups. The regressions have been run separately for each time point. The groups are sorted by implied probability, as defined in Eq. (7), at the closing of the betting events, so that the group size is 100 for 28 and 99 for 12 groups. The p-values of the estimated coefficients are reported in parentheses. The sixth and seventh rows give the F-statistics for the joint test that α = 0 and β = 1. ***, ** and * indicate significance at the 1%, 5% and 10% level respectively. The best values of β and R^2 are bolded.

	Sorted by Close				
	Open	**12:00 AM**	**1 hour**	**15 mins**	**Close**
α	-0.1416***	-0.1196*	-0.0934*	-0.0879*	-0.0816*
	(0.0083)	(0.0195)	(0.0517)	(0.0629)	(0.0809)
β	1.2533***	1.2161***	1.1682***	1.1585***	**1.1472*** **
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
R^2	0.8314	0.8323	0.8363	**0.8376**	0.8370
F	187.39***	188.54***	194.09***	195.98***	195.15***
(p-value)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
n	40	40	40	40	40

10.2. Appendix B: The results of linear regressions with 80 odds groups

Table B.1
Linear regression sorted by open with 80 groups

This table presents the results of the standard linear regression model, as defined in Eq. (14), for 80 equally sized odds groups. The regressions have been run separately for each time point. The groups are sorted by implied probability, as defined in Eq. (7), at the opening of the betting events, so that the group size is 50 for 68 and 49 for 12 groups. The p-values of the estimated coefficients are reported in parentheses. The sixth and seventh rows give the F-statistics for the joint test that α = 0 and β = 1. ***, ** and * indicate significance at the 1%, 5% and 10% level respectively. The best values of β and R^2 are bolded.

	Sorted by open				
	Open	**12:00 AM**	**1 hour**	**15 mins**	**Close**
α	-0.0779*	-0.0936*	-0.0829*	-0.0795	-0.0772
	(0.0970)	(0.0558)	(0.0912)	(0.1057)	(0.1161)
β	**1.1375*** **	1.1689***	1.1492***	1.1433***	1.1392***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
R^2	**0.7042**	0.6978	0.6884	0.6852	0.6835
F	185.67***	180.07***	172.30***	169.74***	168.41***
(p-value)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
n	80	80	80	80	80

Table B.2
Linear regression sorted by 12 AM with 80 groups

This table presents the results of the standard linear regression model, as defined in Eq. (14), for 80 equally sized odds groups. The regressions have been run separately for each time point. The groups are sorted by implied probability, as defined in Eq. (7), at 12 AM of the game day, so that the group size is 50 for 68 and 49 for 12 groups. The p-values of the estimated coefficients are reported in parentheses. The sixth and seventh rows give the F-statistics for the joint test that α = 0 and β = 1. ***, ** and * indicate significance at the 1%, 5% and 10% level respectively. The best values of β and R^2 are bolded.

	Sorted by 12am				
	Open	**12:00 AM**	**1 hour**	**15 mins**	**Close**
α	-0.1217**	-0.1019*	-0.1003*	-0.0990*	-0.1002*
	(0.0322)	(0.0604)	(0.0610)	(0.0639)	(0.0592)
β	1.2171***	1.1840***	1.1808***	**1.1787*** **	1.1811***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
R^2	0.6527	0.6584	0.6635	0.6635	**0.6670**
F	146.60***	150.32***	153.81***	153.78***	156.23***
(p-value)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
n	80	80	80	80	80

Table B.3
Linear regression sorted by 1 hour with 80 groups

This table presents the results of the standard linear regression model, as defined in Eq. (14), for 80 equally sized odds groups. The regressions have been run separately for each time point. The groups are sorted by implied probability, as defined in Eq. (7), at 1 hour before the scheduled start of the game, so that the group size is 50 for 68 and 49 for 12 groups. The p-values of the estimated coefficients are reported in parentheses. The sixth and seventh rows give the F-statistics for the joint test that $\alpha = 0$ and $\beta = 1$. ***, ** and * indicate significance at the 1%, 5% and 10% level respectively. The best values of β and R^2 are bolded.

	Sorted by 1 hour				
	Open	**12:00 AM**	**1 hour**	**15 mins**	**Close**
α	-0.1354**	-0.1137**	-0.0851	-0.0852	-0.0852
	(0.0214)	(0.0444)	(0.1155)	(0.1145)	(0.1146)
β	1.2420***	1.2054***	**1.1531*** **	1.1535***	1.1537***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
R^2	0.6467	**0.6483**	0.6470	0.6477	0.6475
F	142.77***	143.78***	142.95***	143.43***	143.26***
(p-value)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
n	80	80	80	80	80

Table B.4
Linear regression sorted by 15 minutes with 80 groups

This table presents the results of the standard linear regression model, as defined in Eq. (14), for 80 equally sized odds groups. The regressions have been run separately for each time point. The groups are sorted by implied probability, as defined in Eq. (7), at 15 minutes before the scheduled start of the game, so that the group size is 50 for 68 and 49 for 12 groups. The p-values of the estimated coefficients are reported in parentheses. The sixth and seventh rows give the F-statistics for the joint test that $\alpha = 0$ and $\beta = 1$. ***, ** and * indicate significance at the 1%, 5% and 10% level respectively. The best values of β and R^2 are bolded.

	Sorted by 15 mins				
	Open	**12:00 AM**	**1 hour**	**15 mins**	**Close**
α	-0.1447**	-0.1172**	-0.0905	-0.0855	-0.0871
	(0.0173)	(0.0477)	(0.1082)	(0.1237)	(0.1165)
β	1.2589***	1.2118***	1.1629***	**1.1540*** **	1.1571***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
R^2	0.6385	0.6298	0.6322	0.6349	**0.6366**
F	137.75***	132.67***	134.06***	135.65***	136.63***
(p-value)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
n	80	80	80	80	80

Table B.5
Linear regression sorted by Close with 80 groups

This table presents the results of the standard linear regression model, as defined in Eq. (14), for 80 equally sized odds groups. The regressions have been run separately for each time point. The groups are sorted by implied probability, as defined in Eq. (7), at closing of the betting events, so that the group size is 50 for 68 and 49 for 12 groups. The p-values of the estimated coefficients are reported in parentheses. The sixth and seventh rows give the F-statistics for the joint test that α = 0 and β = 1. ***, ** and * indicate significance at the 1%, 5% and 10% level respectively. The best values of β and R^2 are bolded.

	Sorted by Close				
	Open	**12:00 AM**	**1 hour**	**15 mins**	**Close**
α	-0.1412**	-0.1161**	-0.0904*	-0.0839*	-0.0777
	(0.0101)	(0.0293)	(0.0737)	(0.0941)	(0.1173)
β	1.2527***	1.2099***	1.1629***	1.1513***	**1.1400*****
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
R^2	**0.6835**	0.6778	0.6816	0.6804	0.6804
F	168.48***	164.09***	166.95***	166.06***	166.04***
(p-value)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
n	80	80	80	80	80

10.3. Appendix C: Return and staking figures of profitable intervals

Table C.1
Returns of rising favorites in 12 AM to 15 mins interval with different staking plans

This table presents the results of different staking plans for betting on favorites with rising odds in the interval from 12 AM to 15 minutes before the scheduled start of the game. Fixed staking 1% refers to staking 1% of the original bankroll for each bet and win 1% means betting to win 1% of the original bankroll, regardless of the previous results. Percentage staking refers to betting 1% of the current bankroll for every bet. Kelly/x's refer to the Kelly criterion calculated as in Eq. (12) using the average return of the interval as the perceived edge. X is the divisor for the fractional Kelly. The best value of each column is bolded.

Staking strategy	ROI	Average stake	Ending bankroll	Lowest bankroll	Highest bankroll
Fixed staking 1%	**1.12 %**	1.00 %	**1.1715**	0.8689	1.2787
Win 1%	0.73 %	1.47 %	1.1642	0.8608	**1.2890**
Proportional staking	0.78 %	1.04 %	1.1240	0.8662	1.2528
Kelly/1	0.18 %	**1.57 %**	1.0427	0.7275	1.2033
Kelly/2	0.46 %	0.82 %	1.0581	0.8636	1.1356
Kelly/3	0.56 %	0.55 %	1.0466	0.9094	1.0969
Kelly/4	0.60 %	0.41 %	1.0378	0.9322	1.0749
Kelly/5	0.63 %	0.33 %	1.0316	**0.9458**	1.0609
Kelly/10	0.68 %	0.16 %	1.0171	0.9729	1.0314

Table C.2
Returns of rising favorites in 12 AM to 1 hour interval with different staking plans

This table presents the results of different staking plans for betting on favorites with rising odds in the interval from 12 AM to 1 hour before the scheduled start of the game. Fixed staking 1% refers to staking 1% of the original bankroll for each bet, and win 1% means betting to win 1% of the original bankroll, regardless of the previous results. Percentage staking refers to betting 1% of the current bankroll for every bet. Kelly/x's refer to the Kelly criterion calculated as in Eq. (12) using the average return of the interval as the perceived edge. X is the divisor for the fractional Kelly. The best value of each column is bolded.

Staking strategy	ROI	Average stake	Ending bankroll	Lowest bankroll	Highest bankroll
Fixed staking 1%	**1.22 %**	1.00 %	1.1788	0.8300	1.2338
Win 1%	1.05 %	1.46 %	**1.2261**	0.6911	**1.3149**
Proportional staking	0.95 %	0.97 %	1.1349	0.8285	1.2005
Kelly/1	0.53 %	**1.57 %**	1.1224	0.6488	1.2538
Kelly/2	0.82 %	0.85 %	1.1028	0.8169	1.1648
Kelly/3	0.90 %	0.58 %	1.0769	0.8766	1.1168
Kelly/4	0.94 %	0.44 %	1.0607	0.9070	1.0899
Kelly/5	0.97 %	0.35 %	1.0499	0.9254	1.0730
Kelly/10	1.01 %	0.18 %	1.0263	**0.9625**	0.9625

Table C.3
Returns of rising favorites in Open to Close interval with different staking plans

This table presents the results of different staking plans for betting on favorites with rising odds in the interval from the opening to the closing of the betting events. Fixed staking 1% refers to staking 1% of the original bankroll for each bet, and win 1% means betting to win 1% of the original bankroll, regardless of the previous results. Percentage staking refers to betting 1% of the current bankroll for every bet. Kelly/x's refer to the Kelly criterion calculated as in Eq. (12) using the average return of the interval as the perceived edge. X is the divisor for the fractional Kelly. The best value of each column is bolded.

Staking strategy	ROI	Average stake	Ending bankroll	Lowest bankroll	Highest bankroll
Fixed staking 1%	0.97 %	1.00 %	1.1600	0.9747	1.3038
Win 1%	**1.50 %**	1.46 %	**1.3613**	0.9639	1.5254
Proportional staking	0.60 %	1.08 %	1.1069	0.9598	1.2923
Kelly/1	-0.07 %	**4.63 %**	0.9442	0.5466	**2.0296**
Kelly/2	0.71 %	2.69 %	1.3148	0.8497	1.8122
Kelly/3	0.98 %	1.76 %	1.2829	0.9238	1.5674
Kelly/4	1.11 %	1.29 %	1.2359	0.9526	1.4289
Kelly/5	1.19 %	1.02 %	1.1988	0.9670	1.3432
Kelly/10	1.35 %	0.49 %	1.1081	**0.9860**	1.1700

Table C.4
Returns of rising favorites in Open to 15 mins interval with different staking plans

This table presents the results of different staking plans for betting on favorites with rising odds in the interval from opening of the betting events to 15 minutes before scheduled start of the game. Fixed staking 1% refers to staking 1% of the original bankroll for each bet, and win 1% means betting to win 1% of the original bankroll, regardless of the previous results. Percentage staking refers to betting 1% of the current bankroll for every bet. Kelly/x's refer to the Kelly criterion calculated as in Eq. (12) using the average return of the interval as the perceived edge. X is the divisor for the fractional Kelly. The best value of each column is bolded.

Staking strategy	ROI	Average stake	Ending bankroll	Lowest bankroll	Highest bankroll
Fixed staking 1%	1.55 %	1.00 %	1.2576	0.8626	1.3787
Win 1%	**1.64 %**	1.46 %	**1.4002**	0.7858	1.6011
Proportional staking	1.21 %	1.09 %	1.2195	0.8537	1.3916
Kelly/1	0.05 %	**4.80 %**	1.0415	0.3322	**2.5492**
Kelly/2	0.85 %	2.74 %	1.3897	0.6411	2.0370
Kelly/3	1.12 %	1.78 %	1.3330	0.7612	1.6956
Kelly/4	1.26 %	1.30 %	1.2725	0.8221	1.5161
Kelly/5	1.34 %	1.02 %	1.2275	0.8585	1.4085
Kelly/10	1.49 %	0.49 %	1.1215	**0.9305**	1.1983

Table C.5
Returns of rising favorites in Open to 1 hour interval with different staking plans

This table presents the results of different staking plans for betting on favorites with rising odds in the interval from opening of the betting events to 1 hour before scheduled start of the game. Fixed staking 1% refers to staking 1% of the original bankroll for each bet and win 1% means betting to win 1% of the original bankroll, regardless of the previous results. Percentage staking refers to betting 1% of the current bankroll for every bet. Kelly/x's refer to the Kelly criterion calculated as in Eq. (12) using the average return of the interval as the perceived edge. X is the divisor for the fractional Kelly. The best value of each column is bolded.

Staking strategy	ROI	Average stake	Ending bankroll	Lowest bankroll	Highest bankroll
Fixed staking 1%	0.57 %	1.00 %	1.0949	0.7179	1.2216
Win 1%	**0.69 %**	1.47 %	**1.1696**	0.5768	**1.3793**
Proportional staking	0.23 %	0.96 %	1.0364	0.7385	1.1796
Kelly/1	-1.16 %	**2.61 %**	0.4974	0.1691	1.1031
Kelly/2	-0.11 %	2.00 %	0.9625	0.4590	1.3669
Kelly/3	0.18 %	1.44 %	1.0439	0.6096	1.3135
Kelly/4	0.32 %	1.11 %	1.0596	0.6961	1.2569
Kelly/5	0.40 %	0.90 %	1.0603	0.7517	1.2146
Kelly/10	0.55 %	0.46 %	1.0424	**0.8707**	1.1149