JOHANNES KEPLER
UNIVERSITY LINZ

Author
Pöschl Philipp
k01323366

Submission
Department of Economics

Thesis Supervisor
Lackner, Mario (DR)

July 2019

HOW EFFICIENT IS THE EUROPEAN ONLINE SOCCER BETTING MARKET?

- AN ANALYSIS OF THE TOP 3 EUROPEAN LEAGUES OF THE 2007/8 TO THE 2017/8 SEASON

Master Thesis

to confer the academic degree of

Master of Science

in the Master Program

Management and Applied Economics

JOHANNES KEPLER
UNIVERSITY LINZ
Altenberger Str. 69
4040 Linz, Austria
www.jku.at
DVR 0093696

SWORN DECLARATION

I hereby declare under oath that the submitted Master Thesis has been written solely by me without any third-party assistance, information other than provided sources or aids have not been used and those used have been fully documented. Sources for literal, paraphrased and cited quotes have been accurately credited.

The submitted document here present is identical to the electronically submitted text document.

Saint-Cloud, 25th of July 2019

Pöschl Philipp

Table of Figures and Tables

Figures

Tables

JOHANNES KEPLER UNIVERSITY LINZ

Table of Contents

Introduction

'People make just as many mistakes when stakes go up, maybe more'
- Richard Thaler (2015)

While early economic theory based itself upon the assumption of rational, informed economic actors (e.g. the famous homo oeconomicus), gained insights in the field of Behavioral Economics (Akintoye, 2008) form a heavy challenge on the status quo. With the rationality of human behavior under question, the gap between economic theory and reality becomes more visible. Behavioral economist especially put the accuracy and validity of the Efficient Market Hypothesis (EMH) (Malkiel, 2005) into question, which is the foundation of Modern finance theory. EMH-supporters argue that financial markets are efficient to the extent that all relevant information is incorporated at any time in share prices and it is impossible to 'beat the market' (Fama, 1997). Yet, they fail to explain why bubbles and market crashes, such as the collapse of the subprime mortgage bubble in 2008, take place (Shiller, 2003).

As Behavioral economists believe (e.g. Thaler (2006) in 'Are markets efficient?' or Shiller (1998): 'Human Behavior and the Efficiency of the Financial System'), the EMH can be true if some people act irrationally in a random, non-systematic way resulting in 'noise'[1] which can lead to market anomalies. The debate between these two fronts has led to the question: How efficient are financial markets in regard of incorporating new information and clearing out market price misalignments?

The question of market efficiency is essential because at times markets fail the ability to self-regulate. This can furtherly lead to *Allocative Inefficiencies*. Particularly, allocative inefficiencies are caused by inefficiencies in the financial markets and lead to a misallocation of capital between companies or sectors. The result of it is a distorted supply of goods and services to society.
When market inefficiencies are identified, counteractive measures such as arbitrage or regulatory actions can be taken to regain market efficiency. To further explain, the term *Arbitrage* refers to opportunity-seeking market actors taking profit of arbitrage opportunities through which they correct existing market inefficiencies[2]. Although, in a real-life scenario market power is needed to take advantage of bigger arbitrage opportunities. This is based upon the idea that market misalignments can be persistent over years and put the arbitrageurs in a financially dangerous position (Crow, 2016). At times, regulatory authorities (as e.g. the European Commission or

[1] Noise=share price movements without the upcoming of new, significant information;
'noise traders' differ from the rational, informed market participant in the way that they fall for biases and heuristics;
For more look at: Ramiah, Xu and Moosa (2015): Neoclassic Finance, behavioural finance and noise traders: A review and assessment of the literature.
[2] Cf. https://www.investopedia.com/terms/a/arbitrage.asp (As of 08.05.2018)

ESMA) must step in and apply market regulations. The *Herbalife Inc*. case proves this point. Herbalife is a stock-listed enterprise producing and selling nutritional supplements that became a top trend in USA. Often judged for its retail politics, the company faced strong critics by Investor Bill Ackman, for being a pyramid-structured organization that takes advantage of young entrepreneurs (Pershing Square Capital Management, 2013). Although having significantly shorted the stock in 2012, the stock price finally fell in 2015, when the FTC filed a lawsuit against the firm.

Due to the high complexity of financial markets it is difficult to answer the question based on their efficiency. Academics of the two different fronts (EMH and BE) have conducted extensive research regarding the topic, coming up with contradictory evidence regarding the degree of efficiency of financial markets (Akintoye, 2008). To find out more about the biases and fallacies that misguide the conduct of some market participants, economists started to carry out researches in the setting of sports betting markets. Sports betting markets can be perceived as the ideal 'playground' due to their similarities to financial markets[3] (which facilitates inference). Furthermore, they display unique characteristics that simplify investigations as, for instance, the defined ending point of each asset at which its value becomes certain. *'As a consequence, the problems that arise in evaluating future dividends or fundamentals are mitigated.' (Law and Peel, 2001, p.327).*
Especially the Online Sports betting industry is of interest as it is subject to rapid growth. With the size of the online gambling market expected to reach a market volume of 59.79 Billion US-$ in 2020 (Gough, 2019).

As the most popular sport (soccer) in Europe, Online Soccer betting becomes an attractive field of research. With the rise of the internet and the increased availability and ease of access of information profitable betting opportunities/strategies should be taken advantage of ever faster. Meaning that Opportunities to 'beat the market' should disappear over time as informed bettors take advantage of them and force the counterparts to correct their shortcoming behavior. Furthermore, the increased availability of data should lead to increased competition of the bookies and be turning markets more efficient, with bets becoming closer to fair bets.
These market developments have taken my interest and lead to the formulation of my research questions as follows:

- How efficient is the European Online Soccer Betting Market?

[3] *'Betting markets are well suited for testing market efficiency, since they share many characteristics with financial markets, in particular large number of investors (betters), with readily available cheap sources of information, involved in the purchase of state-contigent assets (bets); see Sauer (1998); Thaler and Ziemba (1988) and Vaugahn William (1999)'*, Law and Peel(2001): Insider Trading, Herding Behaviour and Market Plungers in the British Horse -Race Betting Market; p.327.

- Has the European Online Soccer Betting market turned more efficient over the years?
- Is it possible for bettors to develop a profitable betting strategy based on ‘odds picking’?

As most research in the field of Sports Betting Markets is conducted concerning Information Market Efficiency, there is a gap in literature concerning Allocative Market Efficiency and its development over time. This thesis addresses this gap by studying the development of the Allocative Market Efficiency of the top three European Soccer Leagues over a 10-season period.

To answer the research questions, two different approaches are used. On the one hand a literature review is given to show the research status. On the other hand, an empirical analysis on the development of betting coefficients, overrounds, fair values and the Vigorish[4] will be used to examine the Online Soccer Betting Market Efficiency, by assuming that the price level and the risk free profit of the bookmakers are the sole necessary indicators for allocative efficiency of the market and its development over time. Therefore, this Thesis is structured into two Parts: A Theoretical Part consisting of a Literature Review which should give the reader all the basics to understand the topic and its problematic and an Empirical Analysis which goal it is to answer the research question.

The Literature Review is structured as follows: At first a definition of the term Betting is given as well as to what Betting markets are. This will be followed by a description of several significant researches in sports betting markets and their findings. This is followed by an introduction of the Efficient Market Hypothesis (EMH) and its implications for sports betting markets. This section clarifies the underlying theories on how financial and betting markets function and also outline their differences.

The Empirical Part is introduced with the Research Question and the Hypothesis, clarifying the goal of this paper. Then the institutional setting and the data is described. Finally, the methodology used for the empirical analysis is explained before the results and findings are interpreted

To draw to close, a Conclusion will be given with an outlook for the topic.

[4] https://en.wikipedia.org/wiki/Vigorish

1. Literature Review

In this chapter, the theoretical foundation for the empirical part is discussed. Betting markets and their forms will be described before the most important theoretical and empirical findings in sports betting markets and especially soccer betting markets will be outlined. Then the theoretical application of the EMH on sports betting markets will be debated.

1.1. Betting

Betting is as old as time. From betting on Gladiator fights in the ancient Rome, to bets on sports teams in modern sports arenas: the principles have stayed the same. A bet is as a contract over the outcome of a future event (Lottery Game), made directly between individuals, or over an intermediary (the bookmaker) who matches supply and demand. At the time the bet is placed, its realization is uncertain. That is why betting is also classified as a prediction game (Wolfers and Zitzewitz, 2004). In Economics a bet is seen as a lottery game **L** (y_1, y_2, π_1, π_2) because the future consumption/income (y_1, y_2) of an individual is dependent on the outcome of the bet with only its estimated outcome probability distribution (π_1, π_2) being known. The general assumption is that individuals which take part in a lottery game are therein trying to maximize their utility. Further assumptions regarding these individuals are that they *are rational* – choosing the best lottery available to maximize their utility (rationally comparing all its options and ranking them according to their preferences) – and that their preferences *fulfil the properties of the axioms of transitivity*, completeness and reflexivity (Varian, 2010).

Lottery games are distinguishable into *simple or compound* lottery games. A simple lottery has the characteristics that the possible realizations are known. Betting on e.g. the outcome of a single soccer game where the bettor can decide whether to bet on the win of team A, team B or a draw between the two teams. In this case the simple Lottery game could be described as **L**(X_A, X_B, X_D; π_A, π_B, π_D) with X_i describing the outcome i with its relative realization probability π_i. A compound lottery on the other hand is more complicated, as it *'allows the outcomes of a lottery themselves to be simple lotteries'* (Mas-Colell, Whinston, Green, 1995, p. 83).

Betting on more than one game at a time (known as combination-/multi bet) therefore is seen as a compound lottery game as each game is a simple lottery itself. The advantages are a higher possible gain, whereby more risk is taken. If we add to our previous example, a game between team E and team F in which the bettor bets on at the same time, the resulting compound lottery game could be described as **L** (X_A, X_B, X_D; π_A, π_B, π; X_E, X_F, X_D; π_E, π_F, $\pi_{D)}$ where the gambler must correctly guess the outcome of both simple lotteries to win. Regarding the probability distribution π_i of an outcome i with its possible realizations x_i it is to be mentioned that the true probability distribution is unknown until after the event. Therefore, the individuals use the information available to guess their best estimate of the true probability distributions. Depending on its estimated probability distributions and preferences, an individual will decide on their willingness-to-pay for a bet and therefore whether to gamble or not, maximizing its utility.

1.1.1. Betting Markets and their forms

In the following section, a definition of betting markets will be given, and the special characteristics of sports betting markets will be discussed. A market is generally defined as a place where buyers and sellers get together and interact. In the case of betting markets (which are a special form of prediction markets), individuals engage into contracts regarding the realization of future events (Wolfers and Zitzewitz, 2004). As a contract between individuals inherits several insecurities as e.g. the risk of default (inability of debtor to pay his contractual commitment) or settlement (Culp, 2015), a market maker steps-in to secure the liquidity and functioning of the market. The advantages of having an organised market are its *liquidity* (matching buy and sell orders faster), *security* (guaranteeing the compliance to contracts), and the protection of *anonymity of buyer and seller* (principle of novation) (Telser and Higinbotham, 1977; Nystedt, 2004). The explicit costs of trading in an exchange/ organised market are the 'mark-up' which the market maker charges for its services.

Although due to their similarities Sports betting markets are often compared to financial markets, they are a special kind of betting markets and exhibit several characteristics which are unique.
While in financial markets and some prediction markets the contractual obligations and settlements of debts take place upon realization of the event, the bettor must pay for his bet upon placement (to evade default) to the bookmaker (stake). In case of a successful bet, the bookmaker pays the bettor his prize money(=stake+yield-'overround'). In case of a loss, the bookmaker keeps the stake and 'overround' payed by the bettor (Forrest and Simmons, 2008). The 'overround' is the mark-up the bookmaker charges his clients for his services and is usually included into the price of a bet/betting odds (Leitner, Zeileis and Hornik, 2010). Furthermore, the bookmaker is not only the market-maker in sports betting markets but also acts as counterparty to every transaction (Gomber, Rohr and Schweikert, 2008) [can therefore be compared to the CCP in organised financial markets]. It is him who determines and posts the price of a bet prior to the event (Perison and Smith, 2009) – leaving the bettor with a 'take it or leave it' decision.

When setting the prices, the bookmaker relies on the 'expert knowledge' of its employees (Leitner, Zeileis, Hornik; 2010). When these experts set the odds for a game, they do not only rely on past performances, form and their predictions for the upcoming event, but also consider several other factors such as *bettor sentiment*. The setting of the right price is important to the bookmaker as he faces two major challenges. First, he has to find the right price to balance out demand for the bet on both sides (demand for realization A and B) in order to minimize his own risk (Levitt, 2004).

Second, if the bookmaker sets the odds wrongly, insiders or expert gamblers can take advantage of the market price misalignment, putting the bookmaker in a financially difficult situation. Therefore, price setting is crucial, and the bookmakers face high incentives to get the price right.
The most time the market-maker sets the prices in the form of odds (e.g. horse racing, soccer betting) or point spreads (American football). In the case of odds betting, the odds indicate the future payoffs in case of an accurate bet and are dependent on the implicit probability of the realization of event i. In the case of point spreads the bettor bets on the final point differences between the two teams (why it is also called Over-under betting) (Paul and Weinbach, 2004).
These specifics of the sports betting markets require bettors to have some market knowledge to be able to rationally assess their opportunity. Furthermore, the bettors have the possibility to bet on different realizations of a game (e.g. first goal scorer, first yellow card, ...) either as simple lottery, or compound lottery to increase the possible pay-off.

These special characteristics mixed with other important factors make sports betting markets a popular area to conduct research. One of the first economists to identify the setting of sports betting markets ideal for 'testing market efficiency and rationality' were Thaler and Ziemba in 1988. They argued that betting markets should be more efficient 'because the conditions (quick, repeated feedback) are those which usually facilitate learning' (See also Cain, Law and Peel, 2000). As Levitt (2004, p. 223) points out, the parallels between *'trading in financial markets and sports wagering' are quite a few. In both, 'investors with heterogeneous beliefs and information seek to profit through trading, as uncertainty is resolved over time'* (also see Law and Peel, 2001). On each bet, there is a trader on each side wanting to make financial gain with his transaction. Also Reade (2014) emphasized this point, describing that different to laboratory experiments, it is real human beings participating, who often are experienced in the matter and motivated to get their bet right as their equity is at risk. This helps to circumvent biases and fallacies which could arise in out of data collected in laboratory setting as it is the real world, therefore erasing the 'undesirable 'artificiality' of laboratory settings (e.g. Levitt and List, 2007). Another advantage is that the lifecycle of each bet is prior known to market participants who act accordingly[5]. In the setting of financial markets, the researcher has to determine the beginning and termination point on its own (not knowing the final/real value of a share), thereby involuntarily influencing the data and outcome. One of the consequences of this point is that problems in measuring the success of investment are 'mitigated' as outcomes on those games provide means for measuring the success of investment (Avery and Chevalier, 1999). As a result, sports betting markets with their special features provide *'an abundance of easily accessible data on characteristics and performance of individual personnel, making of professional team sports an ideal laboratory for*

[5] This specifity of the market also demonstrates a disadvantage as the emphasis on betting markets is more on short term investing. Not allowing much inference for short-term decision making of individuals.

the empirical scrutiny of economic theories and hypotheses that might prove difficult to test elsewhere' (Goddard, 2001, p.10).

However, there are also some disadvantages and limits in conducting research in the setting of sports betting markets. For instance, some critics argue that individuals engaging into betting tend to have different preferences than the 'standard citizen' and tend to be less risk averse (Golec and Tamarkin, 1995). Therefore, findings should be critically analyzed, and inference questioned. However, the argument of *'self-selection'* of participants into the betting game turning the market unreliable as a sample for research is not coherent since self-selection happens all the time and is also common in other industries. For example, most people working in financial markets but also in other industries as e.g. healthcare 'self-select' themselves into the industry by their interests but also by actively searching for a job in the field. The odds setting behavior of bookmakers is often put into question, with research findings suggesting that the odds setting is not only influenced by the outcome probabilities of each realization, but also depends e.g. on the number of supporters of each team (Forrest and Simmons, 2008), the likelihood of insiders taking part (Shin, 1991) or the bookmaker taking sides in order to increase profits etc. However, other researches neglect this argument. Choi and Hui, for example, found in their study 'The Role of Surprise: Understanding Overreaction and Underreaction to Unanticipated Events using In-Play Soccer Betting Market, (2012)' that *'…pre-match betting markets are fairly 'efficient,' in the sense that the probability inferred from the 'consensus forecast' in the pre-match betting markets are very close to the actual outcome probabilities'.*

To sum up, the characteristics of sports betting markets which facilitate conducting research, as Law and Peel (2011, p. 327) explained, are a '*large number of investors (betters)*', with '*cheap sources of information*', dealing in '*state-contingent assets (bets)*' with each bet '*having a well-defined termination point at which its value becomes certain*'. While disadvantages might exist, the advantages outweigh them.

1.1.2. Research Review and findings in sports betting markets

Research in sports betting markets leads back to the beginning of the 20th century. While first researches were conducted regarding horse betting markets and their efficiency, (see Griffith, 1949) attention has been spread over the following decades regarding the topics of interest.
In the following chapter the most important research findings, will be discussed to outlay the state of research.

Horse betting markets

With the finding that bettors in horse betting markets exhibit the tendency to bet on longshots, even though their underlying winning probability is lower than the favorites Griffith/McGlothin not only challenged the *theory of homo oeconomicus* in economics, but also were the first to find the famous favorite longshot bias (Griffith 1949 and McGlothlin 1956). Inspiring their peers with their findings to conduct further research on what is today one of the most written-on subjects in sports economics. In their 1988 paper, Thaler and Ziemba investigated Anomalies found in the horse racing betting market and the lottery market, trying to explain the irrational behavior displayed by bettors. They found that people tend to fall for the favorite longshot bias (FLB) for various reasons, including miscalculated winning probabilities for longshots and overweighing of winning probabilities in the utility calculation of the bet. (See Prospect Theory, Kahneman and Tversky, 2013). Additionally, they suggest that gamblers (could) fall into an 'illusion of control' (Langer, 1975). Overweighing winning probabilities because they think that they control the outcome of an event rather than it being determined by chance. Crafts (1985) and Schnytzer and Shilony (1995) also investigated the phenomenon of the FLB on the market, trying to justify the favorite longshot bias by arguing that insiders place their bets early at more 'favorable odds', causing the odds to fall. This idea, the effect of insiders on the betting market and their influence on the odds setting by bookmakers was further developed by Shin. In his 1991 paper he investigated the effect of insider traders on the betting system. Under the assumption that insiders know the true winning probability of the racing horses ('asymmetric information') he develops a theoretical model for bookmakers on how to best set odds when facing a group of insiders with superior information in order to avoid financial losses. Overall, he states that the bookie should '*avoid setting very long odds*' as insiders would take profit of the price setter. Further Shin stated that the uncertainty is higher for 'low probability horses' meaning that the '*potential insider information declines in magnitude as the expected probability of winning increases'*.

The theory of Shin is backed up by several empirical investigations. In 2009, Peirson and John for example tried to empirically test the hypothesis made by Shin. Investigating the 'uncertainty faced by bookmakers' in setting their odds, they find that the 'potential presence of insiders' motivates bookies to increase the overall price level for bets (by giving lower odds), confirming Shin's model.

Furthermore, they find that the 'insider effect falls with increasing price level' and suggest that insiders could be the reason the FLB exists. However, they found that the effect of insiders on the market is no different to the effect of the activity of expert gamblers. However, their study also suggests that there is little difference between the effect of insider- or expert gambler activity on the market. The negative effect of insider trading on the market is that the other gamblers face higher prices, as bookmakers increase the price level to counteract possible losses – turning the market less efficient (Crafts,1985); (Paton, Vaughan Williams and Fraser, 1999). However, Peirson and John make certain assumptions in their study which could explain their result. They assume that *'betting takes place at one instant of time', that the 'amount gambled on each horse is fixed and not related on the offered odds'* and that bookmakers are not able to readjust their odds to balance their books. In real life however, bookmakers do readjust their odds[6] (Forrest and Simmons, 2008) when new information is coming up and odds are influencing the decision of uninformed bettors whether to invest or not (Levitt, 2004), giving the bookie the chance to balance his books. Also Law and Peel (2001) focused on the effect of insider trading, investigating the UK Horse-Racing betting market. In their empirical study they used the Shin measure to distinguish the activity of insiders from herd behavior. They found that when the Shin measure is high and odds are shortened, it is often due to insider activity. When the Shin measure declines over the auction period it is a sign for herd behavior. Their findings suggest that herd behavior is quite common in the British horse-race betting market. Furthermore, they confirm the outcome of the study by Crafts (1985) that insiders '*tend to bet early in the game'*.

Smith and Vaughan Williams focused in their 2010 study on the odds bias in UK betting markets. Their findings proved a favorite longshot bias (FLB) in the UK horse racing market. However, they found in their empirical study that the impact of the FLB decreased overtime. According to the authors, this phenomenon could be due to '*competitive pressure on (the) Bookmaker industry (leading to) a reduction of odds-biases in bookmaker odds'* but could not be furtherly proven.

Football betting

While the research in the horse betting market can be seen as the pioneer work, research interest spread over time, taking over different markets but also different topics of interest. The classical American sports – American football, Baseball and Basketball –find themselves as popular research subjects due to their popularity and ease of availability of data.

In his paper, Levitt (2004, p.223), discusses the difference between financial- and gambling markets, analyzing the way market makers are organizing the market. He finds that even tough

[6] Nota bene: Argument valid for horse betting markets where bookies constantly update the odds; not true for soccer betting where the bookie publishes the odds one week before the event and than does not change them

gambling markets display similarities to financial markets, bookmakers in the American football betting market organize the market differently. While in financial markets the market maker matches supply and demand and earns by charging commissions, bookmakers tend to *'take large positions with respect to the outcome of a game…'* rather than choosing odds to level out bets on both sides. Furthermore, evidence is found that bookmakers are better in predicting game outcomes than bettors, maximizing their profits by taking advantage of their superior skills (Levitt, 2004).
In addition, Levitt finds that in American football betting the market makers tend to follow the odds given by experts of Las Vegas casinos. Avery and Chevalier (1999) on the other hand used the football market to investigate the effect of investor sentiment on (bet) prices. As in American Football betting is done on point spreads (Jaffe and Winkler, 1976), Avery and Chevalier (1999, p.494) tried to find if the 'price line' moved in a systematic way due to the effects of several biases as *'(1) so-called expert opinions (which are actually uninformative), (2) a hot-hand bias, and (3) a bias toward prestigious teams.'* Their data suggests that these biases exert an effect on bettors, and this leads to a – to a certain degree – predictable betting line movement. They also find that a betting strategy trying to profit from this systematic bias would be 'borderline profitable', at least for their sample. Their findings prove that sometimes market inefficiencies persist in the market as it is sometimes difficult to take advantage of them. Golec and Tamarking (1992) tested the efficiency of the American football betting market (NFL), finding that the NFL exhibits some inefficiencies like e.g. an underestimation of the 'home field advantage', meaning that bettors underestimate the effect while focusing too much on hot hands, winning streaks or the favorite.

To sum up, the findings of the horse betting market and the betting market for American sports suggests that through statistical analysis on e.g. betting odds and game statistics sometimes inefficiencies can be found. However, it sometimes remains difficult to formulate profitable betting strategies based on the found inefficiencies.

Soccer Betting

With the beginning of the 20th century, economists discovered the market for themselves to test out various hypothesis and economical principles. While in the beginning research focused on the labor market for sports players and the nature of production in sports (Rottenberg, 1956) or the utility function/maximization of soccer clubs and even leagues (Neale, 1964), the focus altered towards spectator attendance and demand in the 1970s and -80s before occupying itself with questions regarding betting strategies and market efficiency (Goddard, 2001).

The question of market efficiency is one of the most popular questions in economics as only if a market is efficient, it assures the perfect functioning of the law of demand and supply. Pope and Peel (1989) were one of the first ones interested with market efficiency of odds setting on game outcomes in the football betting market. In their study they examined the 1981/2 season and found that the odds setting system '*seemed to be efficient since they found no evidence for a profitable betting strategy within the market*' (Cain, Law and Peel, 2000, p. 26). Building up on their empirical study, Cain, Law and Peel (2000) conducted a research regarding the market efficiency and the favorite-longshot bias in the UK football betting market. Their results suggest that the market exhibits a favorite longshot bias (FLB) as found in previous studies within the horseracing betting market. However, they try to rationalize the FLB found in their study by building up the hypothesis that the bookmaker, facing a certain percentage of insiders, needs to make a certain profit with outsiders – whose preferences '*are evenly distributed all over the race*' in order to pay out the insiders. In their paper '*Modelling Association Football Scores and Inefficiencies in the Football Betting Market*', Dixon and Coles (1997) came up with the theory that, as the bookmakers odds in football betting are set one week before the event, a statistical model which calculates the underlying winning probabilities more efficiently could be the base for a successful betting strategy. To test their hypothesis, they defined a set of variables which should determine the underlying winning probability of a team. Once their winning probability is calculated, they set it in relation with the bookmaker's and if the coefficient passed a certain threshold, they took the bet. They proved that for high levels this strategy can be successful. However, tests were made after the events already took place putting into question the accuracy of the model.

Motivated by the theory of Kruypers (2000) that teams with a large fan base will receive less favorable odds set by the bookmaker resulting out of the law of supply and demand (Levitt, 2004), Forrest and Simmons (2008) conducted a study on the 'Sentiment in the betting market on Spanish football'. Opposite to common thought of most economists their findings suggest that more favorable odds are offered to clubs with a larger fanbase. The authors furtherly suggest that the different outcomes of the theoretical and empirical models may be caused by the fact that theoretical models take the bettors behavior as given due to their preference (non-financial motivation to wager) for 'their team'. In real life however clients are likely to wager with a rival bookmaker, looking for better odds. A study by Franck, Verbeek and Nüesch (2010) supports the findings of Forrest and Simmons. When researching on arbitrage opportunities within the online betting industry they found that around 0.8% of the games in the Top-5 European Leagues offer Arbitrage opportunities from the 2004/5 to the 2011/12 season. Their explanation to the findings is that betting websites might sometimes set inefficient odds in order to attract customers as there is a certain lock-in effect between bettors and 'their' bookie. Further studies include the question regarding differences in efficiency within the European soccer market. Bruce and Johnson (2012, p.5) e.g. found competitiveness to be a strong influence '*in explaining cross-league efficiency differences*'.

As previous study by Coleman (2011) and Chan (2003) proved over- and underreaction to unanticipated unforeseen events, Choi and Hui (2012) tried to come up with a hypothesis explaining the opposed findings. Using the '*In-Play Soccer Betting Market*' to understand the phenomenon of '*overreaction and underreaction to unanticipated events*'. They tested their theory that people 'generally underreact to new events due to conservatism' and suggest that when unanticipated events occur and the level of surprise is very high, people tend to overreact. Their findings support their hypothesis, with findings suggesting that people generally underreact to goals occurring in the match, except when they are highly surprising.

Other empirical studies focused on the factor of chance inside football. Quitzau and Vöpel (2009) for example empirically investigated the effect of the factor chance on game outcomes in the Bundesliga and the Premier League and tried to use their findings to come up with conclusions about the balance of forces within the leagues. With chance defined as events which cannot be predicted prior to a game, the researchers find that the more balanced two teams are, the more likely it is the factor chance deciding the game outcome. They also find that the factor of chance is more game-deciding within the Premier League, even though within the Bundesliga teams are more balanced. This could be because the topflight in the Premier League is more balanced than in the German League.

Motivated by a study of Gil and Levitt (2007) in which the authors analyzed the reaction of the In-trade betting exchange participants to goals scored in the FIFA World cup 2002 (using live-data), [finding that the betting exchange seems to be information inefficient7], Croxon and Reade (2013) investigated the market efficiency of soccer betting markets regarding their ability to incorporate new information into prices. Opposite to most studies, they used live betting data (source: Betfair), allowing them to analyse the direct '*reaction of prices to goals scored*'. Their conclusion is that the Betfair betting exchange trends to act as if is information efficient, with robustness checks supporting their findings. In their 2008 paper '*Information Salience, Investor Sentiment, And Stock Returns: The Case of British Soccer Betting*', Palomino, Reeneboog and Zhang investigated the reaction of stock listed soccer clubs share prices to the arrival of new information. They found that while there was no such reaction to the publication of bookmaker odds, prices tend to react to the publication game results. The authors suggest that this phenomenon could be due to media coverage. Furthermore, they found proof for investor sentiment within the market; with investor overreaction induced by highly anticipated wins leading to abnormal high average returns days after the result. In the contrary case (if they favorite would loose), no significant reaction could be found. However, their study proved the influence of investor mood in decision making and in the formation of over- and underreaction to new, upcoming information (Choi and Hui, 2012). As

[7] That means that they say that the reaction regarding a goal scored is a bit slowly regarding the formation of new prices/bets incorporating the new information (=a goal)

shown in previous studies, investors tend to overreact when an event validates their expectations and underreact to adverse events. With only extremely surprising events being able to change their expectation formation (Pezzo, 2003).

Another discussed subject in sports economics is the prediction ability/accuracy of bookmakers, betting exchanges or different forecasting models. In the light of making profits by estimating the true outcome probabilities more accurately (taking advantage of irrationalities of bookmakers' odds setting behavior), researchers got interested just how efficient the odds setting process by bookmakers is. The research results are not very conclusive as some studies prove superior skill of bookmakers in forecasting outcomes (e.g. Forrest, Goddard and Simmons, 2005) with e.g. Choi and Hui (2012) finding that pre-match betting odds seem 'very close to the actual outcome probabilities. Other researchers found the existence of biases in the odds setting process (e.g. Levitt, 2004). The inefficiencies found however are only marginal and therefore cannot be used as basis for developing a systematic profitable betting strategy (e.g. Dixon and Pope, 2004). As example, a study by Cain, Law and Peel (2000) proved the existence of a favorite longshot bias in the UK Football betting market, with average betting returns on the longshot being inferior to the average betting returns of the favorite teams.

Smith, Paton and Vaughan Williams (2009) based themselves on the findings of Forrest, Goddard and Simmons (2005), Song, Boulier and Stekler (2003) and others who found that bookmaker odds offer a good prediction for game outcomes when formulating their hypothesis whether bookmakers '*display superior skills to bettors in predicting the outcome of sporting events*'(Smith, Paton and Vaughan Williams, 2009, p.2). In their 2009 paper '*Do bookmakers possess superior skills to bettors in predicting outcomes?*' they are using data from betting exchange odds and compare it to classic bookmaker odds. Their findings suggest that, even when cleaning the bookmaker odds with the Shin adjustments, betting exchange odds possess '*a marginal superiority*' in predicting game outcomes. They try to justify their finding by a FLB existing in bookmaker odds through which bookies try to 'insure' themselves of insider trading. Also, Franck, Verbeek and Nüesch (2010) compared the prediction accuracy of bookmakers and betting exchanges, confirming the findings of Smith, Paton and Vaughan Williams. Leitner, Zeileis and Hornik (2010) on the other hand investigated the forecasting ability of bookmakers compared to other - ability ratings based - models to predict the outcome of sports tournaments. Analysing the EURO 2008 soccer tournament. More explicitly, the authors assume that 'winning probabilities and underlying abilities – can be derived from both types of information – ability ratings and bookmaker odds. Based on this, they develop a bookmaker odds-based consensus model which they compare with models based on e.g. the Elo rating of teams or on the FIFA/Coca Cola World rating of teams. In their study, the model based on bookmakers' odds outperforms the other models, suggesting superior ability of bookmakers to predict outcomes. They found that the bookmaker consensus model is especially arrant when compared with models solely based on '*team/player abilities from past performances*', confirming the findings of Boulier and Stekler (2003), Spann and Skiera (2009) and others that fixed odds by bookmakers are an '*efficient forecasting instrument for the outcome of single matches*'(Leitner, Zeileis and Hornik, 2010, p.2).

1.2. The Efficient Market Hypothesis (EMH)

The research review showed that no matter how diverse the different topics of interest might be, they all question or base themselves on the Efficient Market Hypothesis (EMH). The theory of the EMH was originally formulated by Eugene Fama (1970) who analyzed the formation of prices on stock markets and their respective incorporation in financial markets. In his paper, Fama defines an efficient market as follows '*A market in which prices always 'fully reflect' available information'* (Quitzau, 2005, p. 8).

Based on this definition, Fama formulated three conditions[8] a market should fulfil to assure its efficiency. These are as follows:

- All market participants have free access to all available information
- All market participants agree to the implications/meaning of the actual available information in regard of the actual and future price formation (prediction of the Expected Present Value)
- There are no transaction costs on financial markets

As a result, the market price of a financial bet is the best possible price prediction. The searching of individuals for profit opportunities is leading to noise, which in the short run can alter the market price – however, the market price will always return to its fundamental value (Quitzau and Vöpel, 2009) and is only changed when new information becomes available. This leads to negative (negative news→ price sinks) or positive shocks. These shocks follow a random walk and therefore are unpredictable (Shiller, 2003). As a result, it is impossible to beat the market, meaning that it is not possible to gain above the market average returns without taking on extra risk in return for a higher interest rate (Akintoye, 2008; Malkiel, 2005). Therefore, markets are only inefficient if some market participants are able to systematically beat the market on a constant basis, earning higher average revenues from trading/betting than the market average returns. Because buy-or sell decisions are formed under uncertainty, information efficiency is important for the proper functioning of the markets (Quitzau and Vöpel, 2009). Fama formulated three different levels of information efficiency of financial markets. These are as follows:

[8] As in reality all three points are hardly fulfilled simultaneously, the criteria's are to be seen as favourable but not necessary for market efficiency.

- **Weak-from efficiency**: All publicly available (historical) information, of which the marginal acquisition costs are almost free, are fully incorporated in the market prices
- **Semi-strong Information Efficiency**: Additionally, to all publicly information, also all decentral-published information is incorporated in the market prices. Their marginal procurement costs are positive
- **Strong Information Efficiency**: All available information – including monopolised information like 'insider knowledge'– is reflected by the market prices. (Quitzau, Vöpel ,2008)

Following the definition of the (information) market efficiency by Fama (1970) several empirical studies regarding the efficiency of the American financial markets were conducted. The results suggest that the hypothesis of weak-form and semi-strong efficiency can be confirmed, while results are not significant enough to confirm strong-form information efficiency (e.g. Fama, 1991; Fuhrmann, 1988; etc.)

1.2.1. EMH for sports betting markets

With the increasing betting volume and importance of sports betting markets, the EMH got adapted (and applied) for sports betting markets.
Thaler and Ziemba (1998, p. 163) were one of the firsts to adapt the EMH to sports betting markets, defining two possible scenarios:

'Market efficiency condition 1 (weak). No bets should have positive expected values.

Market efficiency condition 2 (strong). All bets should have expected values equal (1-t) times the amount bet.'

Hence, according to their condition for weak market efficiency no betting strategy based on betting odds should exhibit positive returns. According to condition 2, markets are strong efficient if there is no betting strategy that *'would improve on the (negative) expected return from betting randomly'*.

Vaughan Williams (2005) in his book '*Information efficiency in financial and betting markets*' was taking the definition of Thaler and Ziemba one step further and was defining three levels of market efficiency for betting markets based on the theory by Fama (1970):

- **Weak-from efficiency**: All historical information (on betting odds) is incorporated into bet prices. Therefore, it is impossible to develop a profitable betting strategy - based on the analysis of the historical betting odds (=technical Analysis) - which guarantees higher returns than the negative average Vigorish. Mathematically weak-from efficiency can be described as follows:

$$\pi\left(\sigma_i^T\right) > -V\left(\sigma_i^T\right);\ given\ that\ \nexists\ \sigma^T \in \Sigma^T$$

with Σ^T being the set of all possible betting strategies, σ^T one respective betting strategy with its potential profit $\pi\left(\sigma_i^T\right)$ and $V\left(\sigma_i^T\right)$ the respective average Vigorish by the bookmaker

- **Semi-strong efficiency**: Additional to the conditions for weak-form efficiency, all relevant public information needs to be incorporated into the bet prices. Therefore, not only a technical- but also a fundamental Analysis could not lead to returns higher than the negative average Vigorish. Mathematically, the set of possible betting strategies is enlarged by all betting strategies based on a fundamental analysis. Therefore, semi-strong efficiency can be described as follows:
$$\pi\left(\sigma_i^{T,F}\right) > -V\left(\sigma_i^{T,F}\right);\ given\ that\ \nexists\ \sigma^{T,F} \in \Sigma^{T,F}$$
- **Strong efficiency**: Additional to the conditions for semi-strong efficiency, also insider information needs to be incorporated into bet prices. Hence, it is not possible for anyone to formulate any profitable betting strategy σ_i as the market fully reflects all information in the market prices. Consequently, strong market efficiency is described as follows:
$$\pi\left(\sigma_i\right) > -V\left(\sigma_i\right);\ given\ that\ \nexists\ \sigma \in \Sigma$$

The definition by Vaughan Williams can be seen as the basis of several empirical investigations regarding the market efficiency of sports betting markets.
With most empirical studies regarding market efficiency of betting markets conclude that the markets actually fulfil the conditions for weak-and semi-strong market efficiency (Quitzau and Vöpel, 2009).

2. Research Questions, (Hypothesis) and Assumptions

With the rise of the internet and the ease of information access, betting markets should have turned more efficient over the years (2007-2018) the advances in technology information gets easier to access, gather and analyze. This increased competitive pressure on bookmakers consequently should lead to odds being closer to fair values. The result of this should be observable by a diminishing vigorish throughout the online soccer betting industry over time, resulting in a more (pareto) efficient market. Under this context, taking the price level of a bet ceteris paribus as the sole indicator for allocative market efficiency, the first research question can be formulated as follows:

- How efficient is the European online soccer betting market?
    - Has the European online soccer betting market become more efficient over time?

This question is of major importance as allocative efficiency is key in maximizing utility among individuals engaging in betting markets (and therefore maximizing a societies welfare). According to economic theory, competition within an industry should lead to lower profit margins of the sellers (which should tend towards 0 economic profit over time in a perfectly competitive market) while maximizing the total benefit. But to measure the level of utility of a market, and therefore of its individual participants which interact through the market, is quite complex as heterogenous utility functions and preferences have to be taken into account (not to forget the production functions of the individual bookmakers which act as the 'suppliers' and also market makers in the sports betting markets).

To circumvent this issue, the assumption that the price level alone for taking a bet is the sole indicator for allocative market efficiency along with the profit-percentage of the suppliers is made. As the price of taking a bet is reflected by the bookmakers betting odds, the analysis will be focused on the development of the '*overround'* and the bookmakers profit margin – the '*vigorish'*. It is assumed that the German, Spanish and English League are a representative sample for the whole European Online Soccer Betting Market. This is rationalized by the fact that these leagues are not only the most competitive but also most followed league tournaments in European Soccer in the period from the 2007/8 to the 2017/8 season. Additionally, it is taken as given that the bookmaker odds within the sample are representative for the entire market and that there should be no differences in bookmaker odds on game outcomes depending on where the bet is taken (consequently, no country differences in bookmaker betting odds offered).

While the first research question explores the development of allocative market efficiency over time, the second research question is focused towards the topic of (Information) market efficiency and its development over time. According to its definition by Eugene Fama, it should not be possible for individuals to develop a profitable betting strategy as it is *'impossible to beat the market'*. However, the increased competitive pressure on bookmakers should lead to more opportunities for bettors. As bets tend to be closer to 'fair bets' the eventual pay-out for bettors is higher, increasing the chance of positive returns. Consequently, a betting strategy based on 'fair odds' and 'odds picking/cross-betting' as one of the indicators of when to take a bet promises to exhibit positive (average) returns. Based on this hypothesis, the second research question results:

- Is it possible for bettors to develop a profitable betting strategy based on 'odds picking'?

In this paper, '*odds picking*' is defined as the process an individual bettor does before engaging into betting. A rational individual compares all available bet prices on the market and consequently chooses the '*best offer*' – the highest betting odd (offering the highest potential return). While we refer to this process as '*odds picking*', other studies may refer to this as '*cross betting*'. Further, for the analysis of possible profitable betting strategies it is assumed that the bettor has no other 'fees' to pay except the '*vigorish'* that the market maker imposes on the bettor by incorporating an evenly distributed '*overround'* into the betting odds.
The '*fair odds indicator*' (difference between the overround and one) will be used as an indicator for when the individual should engage into taking a bet. If this value is close enough to zero it indicates a 'fair bet'. Consequently, the individual bettor in this paper evaluates the game as a potential profit opportunity and calculates his probability on each game outcome – using the bookmaker consensus model - and compares the results to the bookmakers offer in order to decide whether to engage into the bet. Should the '*fair odds indicator*' be negative, it is a sign that there is an arbitrage opportunity in the market.
This thesis touches both forms of market efficiency – pareto and information market efficiency.

The first research question examines the development of the pareto efficiency of the European soccer betting market by focusing on the development of 'fair values' and the bookmakers' vigorish percentage over time. It is assumed that the price level and the supplier's profit is the sole indicator for market efficiency. According to theory – higher competition should lead to higher market efficiency. Consequently, the price level for betting is expected to decrease over time and the mark-up should tend towards 0. The estimations are based on the data on bookmakers' odds. The methodology and computations to achieve is further outlined in the next chapter. Secondly, the market efficiency of the European online soccer betting market and its development will be explored. In this case, a betting strategy based on the bookmaker consensus model – which is based on the bookmakers' odds - will be used to conduct the enquiry. Both betting strategies are further outlined in the next chapter.

3. Data

For the empirical analysis of betting odds, the data on respective game outcomes from the 2007/8 to the 2017/8 season is used to test the hypothesis. The data sums up to 11,726 single observations, specifically 380 single observations per season per league for the English 'Premier League' and Spanish 'La Liga' (20 teams compete for the league trophy), and 306 games for the German Bundesliga (18 teams compete for the league trophy). These three leagues have been chosen as they are the most competitive and most followed leagues in European soccer (classified as the three top European leagues according to the UEFA ranking[9]). As the aim of this thesis is to understand the development of market efficiency in the European online soccer betting market, these three leagues are key to understand the main developments in the market as these leagues promise to have the most potential followers/bettors on each game.

The data studied was collected by football-data.co.uk, a UK-based website that collects the data on soccer betting odds and game statistics and makes it publicly available on their website. To assure the comparability of the data, the website engages into collecting the data on the betting odds in two instants of time: For in-week games each Tuesday afternoon and for weekend games each Friday afternoon. After the games take place, the results and statistical data of the game are added, and the data is freely provided on the website in the format of an excel spreadsheet.

Before describing the data and the used variables in detail, the '*Institutional setting*' will shortly be outlined to explain how the competition is organized.

3.1. Institutional Setting

The league competition (e.g. Deutsche Bundesliga) is a tournament which takes place every year with the aim to find out which is the best team. It is organized by a country's national soccer association (e.g. '*Deutscher Fußball Bund*' [DFB] in Germany; '*The Football Association*' [FA] in England and '*Real Federación Española de Fútbol*' [RFEF] in Spain). The league system usually has several divisions according to the skill level of the soccer clubs. At the end of a season – depending on the system design of the league – the best clubs of lower division get promoted while the worst performing teams of the higher divisions get relegated to keep the number of teams per division constant over the seasons. This promotes an incentive to perform, assuring the attractiveness of the competition through new clubs, raising the uncertainty of outcomes and thereby assuring the competitivity of the tournament. Additionally, for the highest national league, the best teams not only play for prize money but also have the chance to compete internationally

[9] https://en.wikipedia.org/wiki/UEFA_coefficient#Top_leagues_by_period; https://www.uefa.com/memberassociations/uefarankings/country/#/yr/2019 (As of 21/07/2019)

in the following season. By taking part in e.g. the *UEFA Champions League*- or the *UEFA Europa League* tournament teams can not only compete against the best European teams (from other Leagues) but also play for more prize money and fame. The horizon over which the competition takes place is split into seasons. In one full season, every team play against each other twice – once at ‘home’ (in their own stadium) and once ‘away’ (in the stadium of their opponents)[10]. A soccer game is regularly 90 minutes long whereby there are two halves of each 45 minutes with a pause of 15 minutes in between - excluding stoppage time which the referee adds to each half depending on game interruptions[11]. Additionally, if a player is playing unfair and committing a foul, the referee can stop the game and give the opponent team a freekick, a penalty (depending if the foul was committed in the penalty area or on the rest of the field) or even a yellow or red card. When getting two yellow or one red card a player is suspended and his team must finish the game without him, which is an inconvenience for his team since the team has one player less on the pitch. Also, depending on the seriousness or bad intention of the foul committed - the player can be suspended for several games (which is determined by the sports commission of a league after the game took place).

The target of each team is to score as many goals as possible in a game while conceding as few as possible (A goal is when a team shoots the ball into the ‘goal’ of the opponent team). Depending on the outcome of a game a team is acquiring points. For a win, the winning team is acquiring 3 points, for a draw each team acquires 1 point and for a loss a team does not get any points. The teams get listed in a table after each round, sorted by the points acquired. The team with the most points is on top of the table and the team with the least points is on the bottom of the table. Consequently, it is possible at every instance, to know the rank of all teams participating in the competition. At the end of the season the team which has acquired the most points win the league trophy. If two teams end up with the same number of points at the end of the season, the goal difference (=goals scored goals conceded) decides which team is higher ranked in the table.

3.2. Data description

The raw data used for the empirical analysis of this thesis contains information on betting odds and games results. Betting odds are determined and published before the game by the bookmakers, reflecting their implicit outcome estimation of each outcome possibility (home win, away win, draw) of a game. Game results on the other hand are published after a game and can be used as an indicator of the performance of a team in a single game. They build the basis for

[10] While this is true for the English Premier League, the Spanish La Liga and the Deutsche Bundesliga there might be different designs of the league.
[11] Game interruptions may be caused by e.g. the injury of a player; players fouling their opponents, et cetera

game/team statistics and - together with past performances - allow to deduct the form of a team, a team's strengths and weaknesses. Therefore, they are important in predicting the possible game outcomes.

3.2.1. Betting odds

The betting odds are an essential part of the empirical analysis. They are the basis of any successful betting strategy and their respective tests. Thereby, three different forms of betting odds will be used:

- **Individual Bookmakers Betting odds**

The individual bookmakers betting odds do not only represent the respective outcome expectation of a bookmaker, but also allow to calculate for indicators as e.g. fair values or the vigorish. In the empirical analysis, the individual betting odds will be considered as exogenous variables. As such, the individual bookmakers' estimations are key in identifying eventual profit opportunities and inefficiencies in the market.

- **Average Betting Odds**

The average betting odd ($\hat{x}$) is the sum of the principal individual bookmakers' odds (x_i) on a game outcome, divided by the number of the bookmakers (n). As an indicator, it represents the average expected (implicit) outcome probability that an event x_i is taking place. It shows the average return the bettor gets when betting on a game outcome. Furthermore, it is used as an indicator to see the difference between the offered betting odds by the bookmakers and thereby serves as basis to identify the most attractive odds for a game as well as possible outliers.

$$\hat{x} = \sum_{i=1}^{n} \frac{x_i}{n} \ \ with\, x_i \geq 0; \ \ i = 1, 2, \dots n$$

- **Highest odds on a game outcome**

The highest odd offered for a game outcome (x*) is the maximum value of the set of odds (O) offered by the bookmakers. This value is picked from the betting odds (x_i) of the individual bookmakers in the dataset. In an open market where cross-betting is permitted, it can be assumed that a rational bettor will always look for the highest odd offered for the game outcome he bets on in order to maximize his potential return.

$$x^* \in O\ (x_i) and\ x^* \geq x_i\ for\ all\ x_i \in O\ (x_i)$$

It should be mentioned that odds setting is a delicate task for bookmakers, as setting the wrong odds on a game could lead from financial losses up to bankruptcy. Therefore, individual bookmakers do not only rely on all public information but also on the expert knowledge of their employees to predict the outcome of a game (Leitner, Zeileis and Hornik, 2010).

Several scientific studies prove that bookmakers seem to possess superior skills in predicting game outcome compared to bettors (Levitt, 2004), which is key in assuring their success. While Levitt could prove that in NFL betting bookmakers sometimes take on positions in order to maximize their potential profit, studies by Choi and Hui (2012) as well as Smith et al (2009), found that pre-match odds of soccer games reflect true outcome odds accurately. Consequently, in this thesis it is assumed that bookmakers try to balance out their books, making profits only by charging the bettors a mark-up (the 'vigorish') for their services. Furthermore, it is assumed that bookmakers will try to give their best estimate on the possible outcome probabilities for each match, given all available information. This particularity makes it possible to not only to derive the implicit ('true') outcome probabilities from the bookmaker's odds, but also to find the underlying abilities of singular sport teams. While ability ratings like e.g. the Coca Cola Elo rating for the World Cup gives a score for the underlying ability of a team, bookmakers odds already incorporate this information (Leitner, Zeileis and Hornik, 2010).

Additionally, based on the findings of Forrest and Simmons (2000), Choi and Hui (2012), Smith et al (2010) and Leitner, Zeileis and Hornik (2010), it is assumed that the bookmaker consensus model is the best predictor of the actual outcome probabilities. This is because the model is based on individual bookmakers' odds – which are already incorporating all available information and the expert's know-how.

3.2.2. Game statistics

Game statistics summarize key events that took place within a game and – contrary to betting odds – are available only after a game took place. They are the basis for team/player statistics which allow to draw conclusions about e.g. the playing style of a team, its strengths and weaknesses and their recent form.[12] Due to this property, bettors often consider game statistics as key in their decision-making process as it allows them to deduct a team's form and other key variables as e.g. past results between the opponents which they consider when predicting future outcomes. According to the definition of the market hypothesis this data (game statistics and betting odds as both are freely available and 'historic', hence the cost of acquiring them should

[12] With the advances of technology data on individual player performances becomes more valuable and is used e.g. by professional sport teams to e.g. rate the performance of a player or see where rooms for improvements are on individual positions within a team's starting line-up. However, it can be assumed that bookmakers as well as professional bettors will use this new source of information to improve the accuracy of their game predictions.

tend towards 0) should already be included in the betting odds and therefore not allow for a profitable betting strategy. However, bettors will still try to find betting opportunities or place their bet if they thinking that they have more accurate or better information than the bookmaker.

3.2.3. Variables

In the thesis, the data on the betting odds is of higher importance than the data on game statistics as – according to the EMH – the data on games statistics should already be incorporated into the betting odds (Fama, 1970). As game statistics are freely available after each match, they should be incorporated into the market prices – hence the betting odds -even under the assumption of weak market efficiency. This assumption is reasonable as the bookmakers betting odds are their best guess on the respective game outcomes (see Choi and Hui, 2012).

The first examination of the data shows that on average betting odds on home team winning (2.665±1.956) were lower than on a draw (3.965±1.345) or on the away team winning (4.733±4.156).

The respective standard deviation of these values is very high and demonstrates that betting odds are very game specific and vary depending on the home team-opponent constellation. The odds in the data imply that home teams are more likely to be estimated to win a game, followed by a draw with away team wins being the least likely. This is consistent throughout the different leagues. The maximal odds show also that on some games there are clear favorites and underdogs, with the maximum odds found in the data being 41.263 on an away team win in the Spanish league.
Further, -on average the home team is scoring 1.598 (±1.349) goals per Home game, with the away team scoring 1.181 (±1.165) goals on average. Within the data 46.83% of the times the Home Team won a match, 28,74% the away team won and 24,42% of the games ended in a tie.
In addition, the biggest and most successful clubs of each league tend to stay in the highest division which can be seen when tabulating the frequency of the home and away teams, with clubs like e.g. Atletico- & Real Madrid and Barcelona having featured in each season from 2007/8 to 2017/8 (N.B. in the same period, the league champion in the 'La Liga' was always one of these three clubs). This finding is also valid for the other leagues studied where bigger and more established clubs tend to stay within the league and tend to be more successful than smaller clubs. This could be due to the fact that bigger clubs have a higher budget and therefore can afford better players, trainers and training facilities while smaller clubs tend to lose their best players after each season.

The dataset includes observations on 11 726 games. When checking the completeness of the dataset it is found that some bookmaker observations of betting odds offered on the games are incomplete over the seasons/leagues. Consequently, to assure the comparability of the data as well as its comparability over time only bookmakers with more than 10 000 observations are relevant for the empirical analysis. The following bookmakers fulfilled this criterion: B365, Bet and Win, Interwetten, Lad-Brokers, SB, William Hill and Victor Chandler. As a result, the average overround and Vigorish for the industry are calculated based on these individual bookmakers to see how the margin in the industry have developed. This sample is held as being big enough for making an inference as they hold most of the market share in the European Online Soccer Betting Market.

A detailed description of the betting odds and game statistics is available in the appendix.

4. Empirical Analysis

The empirical analysis is focused on two different dimensions of efficiency: allocative market efficiency and information market efficiency (as defined by Eugene Fama, 1970). First, the development of the allocative efficiency of the online sports betting market will be analyzed in order to respond to the first research question:

'How efficient is the European online soccer betting market?'

This analysis is based on the development of the vigorish (risk-free profit margin charged by bookmaker for taking a bet) and the bookmakers average overround in the industry over the last 10 seasons. According to the hypothesis, as the competitivity on the supply side increased, profit margins were supposed to decrease over the same period. As the vigorish (also 'vig') is the risk-free (profit) margin that a bookie charges for every bet taken, increased competition should lead to a decrease in the vigorish over time and since bookies tend to 'balance out' their books the ´vig` can be seen as an equivalent to the economic profit made by the bookmaker. To assure the accuracy of the estimation, the results will be controlled by verifying the development of betting odds versus fair values over the same period which should lead to similar results.

The betting odds serve as basis for this calculation. In football, betting decimal odds exhibit an important property, they represent the bookmakers estimated probability (=predictions) that an event x_i will take place. Therefore, they represent implicit outcome probabilities.

We can calculate the bookmakers estimated probability that event x_H (home team wins) takes place as follows:

$$P_H = \frac{1}{Decimal\ Odds\ of\ Outcome\ H} * 100$$

In soccer betting, we have three possible game outcomes x_i (where i=H, A, D) – A home win (H), an away win (A) or a draw (D). The sum of the probabilities of the outcomes is always 1. This is what we call 'a fair bet' or 'fair odds'. Mathematically, this can be denoted as following:

$$1 = \frac{1}{Decimal\ Odds\ of\ Outcome\ H} + \frac{1}{Decimal\ Odds\ of\ Outcome\ A} + \frac{1}{Decimal\ Odds\ of\ Outcome\ D}$$

The bookmaker adds a certain 'mark-up' on every bet, so that the sum of the probabilities is higher than 1. The difference between 'fair odds' and the actual bookmakers betting odds is called the 'overround' (If the sum would be lower than 1 this would indicate an arbitrage opportunity in the market) and can be calculated as follows:

$$o = \left(\frac{1}{Decimal\ odds\ of\ outcome\ H} + \frac{1}{Decimal\ odds\ of\ outcome\ A} + \frac{1}{Decimal\ odds\ of\ outcome\ D} \right) - 1$$

Based on this we can derive the vigorish (bookmaker's risk-free profit percentage on the total stake made on the event) by using the following formula:

$$v = \frac{o}{(1 + o)}$$

And vice versa (again) the overround if only the vigorish would be known:

$$o = \frac{v}{(1 - v)}$$

With 'v' being the 'vigorish' and 'o' being the 'overround' (commission, percentage that the event book is above 100).

Consequently, by integrating above formulas the vigorish percentage (the risk-free profit a bookmaker makes on event 'i' (if he has balanced out the books) regardless of the outcome) can be calculated as follows:

$$v = 100 * \frac{\left(\frac{1}{H} + \frac{1}{A} + \frac{1}{D}\right) - 1}{\frac{1}{H} + \frac{1}{A} + \frac{1}{D}}$$

In theory, increased competition in the sports betting industry should lead to a lower vigorish and a lower overround throughout the industry. The analysis will be made on industry level using a.) the average odds on each outcome and b.) the highest odds on each outcome offered. Using this information, the question of the development of the efficiency over time is answered.

This is followed by a comparison of the development of bookmaker odds to fair odds as robustness check. For this the fair values of each game are calculated and compared to the actual odds offered by the bookmakers on individual outcomes. If the market became more competitive, the vigorish is supposed to be decreasing but also bookmaker odds are supposed to be closer to fair odds (hence, tend towards 1 in aggregate). Also, the development of the overround over time (which should be equivalent to the development of the 'vig') will be analyzed.

The following calculation – based on the formula by Leitner, Zeileis and Hornik (2010) - will be used to calculate the fair odds 'odds$_i$' on outcome I of event x_i, based on the bookmakers odds 'rawodds$_i$' and the respective bookmakers overround 'o' for the respective event:

$$odds_i = (rawodds_i) * (1 + o)$$

By computing this calculation, it is assumed that the overround is evenly distributed over all possible outcomes of event x_i. To fulfill the property of being fair odds, the sum of all outcome probabilities on event x_i must be 1. This property is fulfilled when using the calculation present in the research findings. Furthermore, based on the 'fair odds' it is now possible to calculate the real estimated outcome probabilities using the bookmaker consensus model. This feature is of importance as the overround would otherwise be polluting the estimation model.

After this first analysis, the second research question will be analyzed:

'Is it possible for bettors to develop a profitable betting strategy based on 'odds picking/cross-betting'?'

While for this analysis the difference between fair odds and bookmaker odds serves as identification strategy of when to possibly take a bet, the bookmaker consensus model by Leitner, Zieleis and Hornik (2010) employed for the forecasting of game outcomes regarding the EURO 2008 will be used to identify a.) whether or not take a bet and b.) on what outcome to bet. The choice for the bookmaker consensus model is justified with the findings that bookmakers display superior prediction abilities regarding the outcome of events (e.g. Choi and Hui, 2012 among others). The estimation model calculates the probability of each event by aggregating the single forecast $p_{i,b}$ values across bookmakers for each underlying game outcome of event x_i. Below the model as by Leitner et al (2010, p.477):

$$Logit(p_{i,b}) = logit(p_i) + \epsilon_{i,b}$$ [13]

'where pi is the latent winning probability for team I and $e_{i;b}$ is the deviation of bookmaker b for team i.'

Hence, the bookmaker consensus model can be mathematically formulated as follows:

13 Leitner, Zieleis and Hornik (2010)

$$logit\ (p_i) = \frac{1}{n}\sum_{b=1}^{n} logit(p_{i,b})$$ [14]

In the model, it is assumed that the sum of the bookmaker's deviation for the single outcomes tends towards zero. If the prediction by the model surpasses a certain difference with an individual bookmaker's prediction, the bet will be taken. For this, the single outcome probabilities of events are calculated and compared to the individual bookmaker's prediction- whereby we especially focus on the 'best odds offered on a game outcome' as the bettor in this analysis is rationally comparing the prices before engaging into betting. Another check will be made using the average odds offered on each game outcome and the respective return of utilizing this betting strategy. Whereby the second strategy is focused on the overround charged on a bet with the target to take only bets which are just close enough to fair odds. If the hypothesis is correct, there is supposed to be an increasing number of betting opportunities over seasons.

In the following section the research findings – based on the formulated research questions - will be outlined.

4.1. Research findings

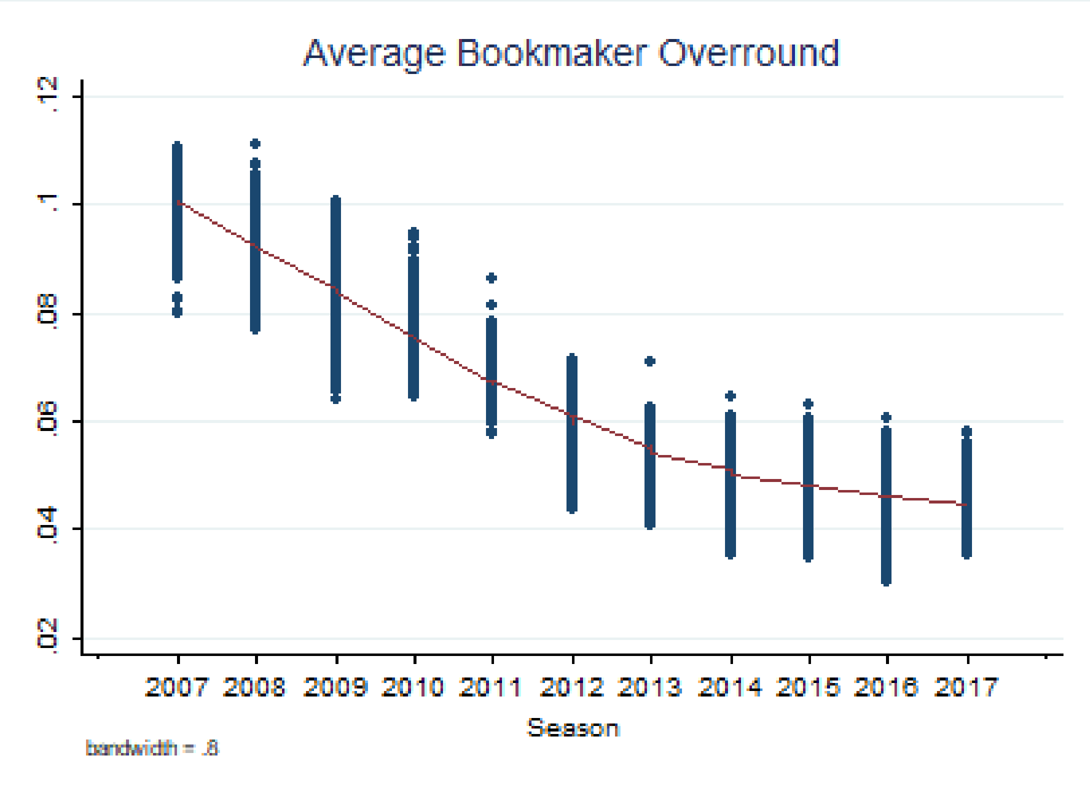
This chapter outlines the research findings, sorted by research question.

'How efficient is the European Online Soccer Betting Market?'

When examining the data for allocative efficiency and its development over time based on the development of the price for taking a bet (using the overround), it is found that the European online soccer betting market has turned more efficient over time. While the average overround charged by the bookmaker for the period of the 2007/8 to the 2017/8 season was 6,5% ± 1,94% for the total dataset, the average overround charged is decreasing over time. For the individual betting markets, the price level for betting on the Spanish market (6.77% ± 1.98%) was the highest, followed by the German Bundesliga (6.75% ± 1.95%) with taking a bet being clearly cheaper for the English Premier League (6.02% ± 1.8%). When observed individually, all three markets exhibit the same trends over the respective seasons. Using the Lowess regression, it can be graphically demonstrated how the average price for taking a bet fell over the studied time period (Figure 1):

14 With n= number of bookmaker; p_i= the predicted probability that outcome I will be realized; i=(H,A,D); b=bookmaker$_n$

Figure 1: Average Bookmaker Overround

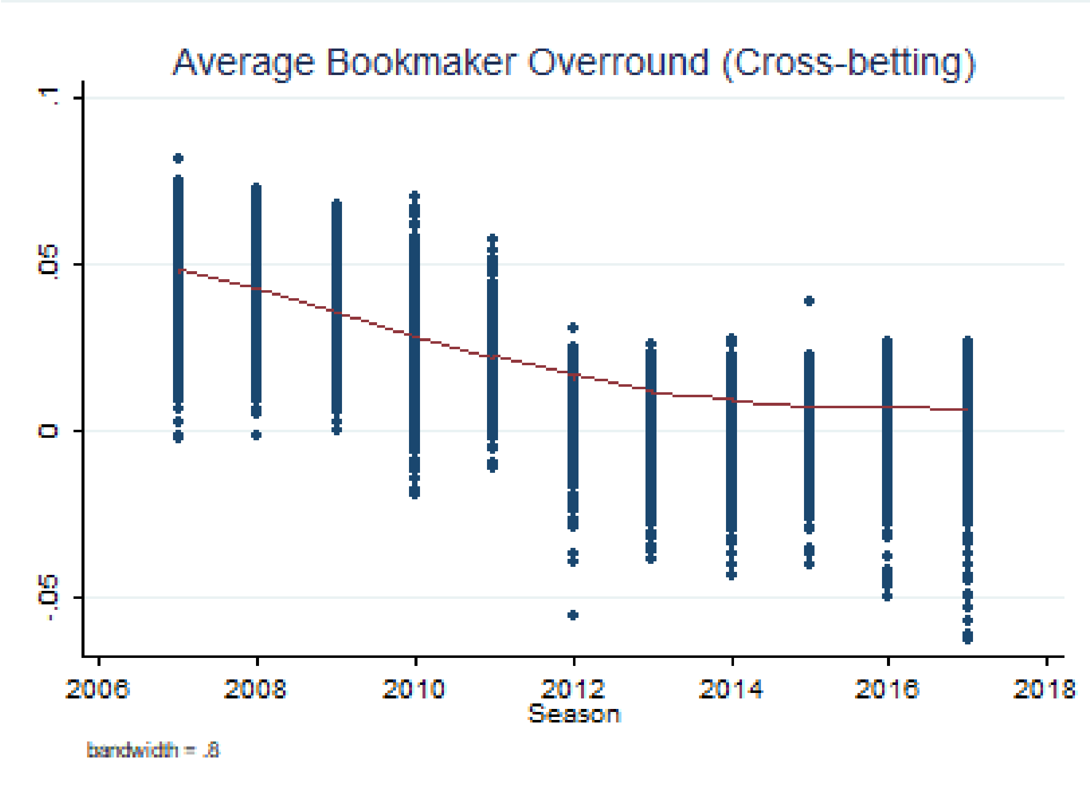


N=11,697

While in the 2007/8 season the average charges for taking a bet was 9.93% (±0.5%), the average price for taking a bet fell to 5.41% (±0.46%) in the 2012/3 season and further decreased to 4.54% (±0.41%) to the 2017/8 season. These findings are also consistent when checking through the different leagues (Figure A.1), whereby bets in the Spanish league were slightly more expensive.

However, if cross-betting between our six eligible/representative bookmakers is permitted, the resulting over-round is lower with the average overround being 2.05% (±0.19%) and the value falling throughout the seasons as Figure 2 shows:

Figure 2: Average Bookmaker Overround (Cross-betting)

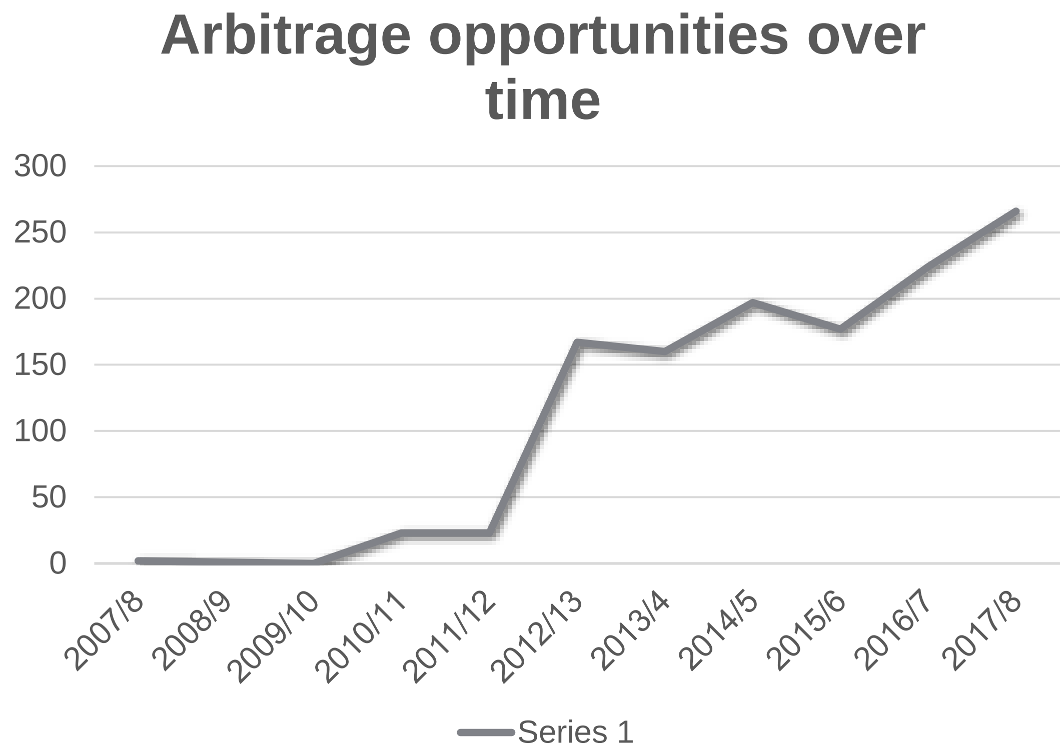


N=11,726

This proves that cross-betting permits the bettor to access more favorable prices with the average overround approaching ‘0’, meaning that bets become closer to fair bets over the seasons. It is also visible in the graph that some games exhibit a negative overround (see Figure 2). This is an indicator that that there is an arbitrage opportunity in the market for bettors. Hence, bettors do have a risk-free profit opportunity in case they engage into cross-betting, under the condition that they distribute their investment over all possible game outcome of event x_i so that regardless of the game outcome they make a profit (equal to the negative overround) Therefore, the findings support the results of Forrest and Simmons (2008), that there is the possibility of risk-free profit on some games in case of cross-betting.

In the data set, in 10.55% of the games arbitrage is possible when engaging into cross-betting. While the number of arbitrage opportunities has significantly increased since the 2009/10 season (Figure 3):

Figure 3: Arbitrage opportunities over time

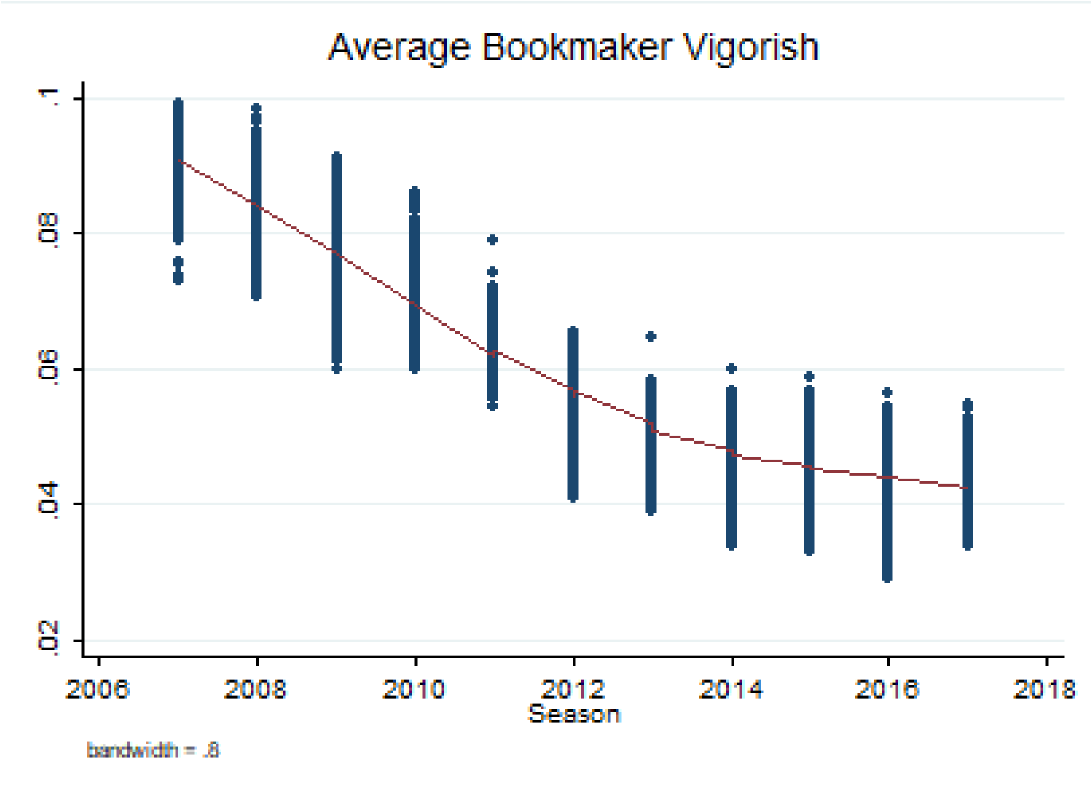


N=1,238

While in the 2007/8 season there were only two games where arbitrage was possible, the value increased to 167 games in the 2012/3 season and up to 266 games in the 2017/8 season. This trend can be explained by increased competition in the industry, leading to lower betting prices as bookmakers compete for customers. While the individual bookmakers' price level (overround) always remains positive throughout the seasons on the single bets, cross betting allows arbitrage within the industry. While it appears logical that the increased competition should lead to falling bet prices, the fact that arbitrage opportunities multiplied over the same period is not logic on first sight. According to economic theory – when arbitrage opportunities appear in the market, individuals will take advantage of it, forcing the market to correct the anomaly. However, in this case the number of arbitrage opportunities has increased in every of the studied markets. This suggests on the one hand that a locked-in effect of bettors with their bookie could be existing, meaning that once a customer has set up his betting account with a bookmaker, he is likely to stay with the bookmaker of their choice as Franck, Verbeek and Nüesch (2012) suggest. This locked-in effect might be due to convenience (do not having to compare betting odds, having all the money in one account and not having to transfer between several accounts with different bookmakers), trust or other reasons.

While the betting odds and the overround indicate the price level bettors face when wanting to take a bet, the vigorish symbolizes the bookmaker's risk-free profit (which can be compared to a supplier's economics profit) in case he balances out his books. According to theory - when a market turns more (allocative) efficient, the sellers (supply side) economic profit should tend towards zero. Therefore, if the market turned more efficient over the studied periods, the vigorish should have decreased over the seasons. The research findings support this hypothesis. The Average Bookmaker Vigorish has indeed decreased over time (Figure 4). While the average bookmaker risk-free profit per game in the 2007/8 season was 9.00% (± 0.43%) it decreased the following seasons to 5.08% (± 0.41%) in the 2012/3 campaign and went further down to 4.3% (± 0.38%) to the 2017/8 season. Below graph shows this development in more detail:

Figure 4: Average Bookmaker Vigorish

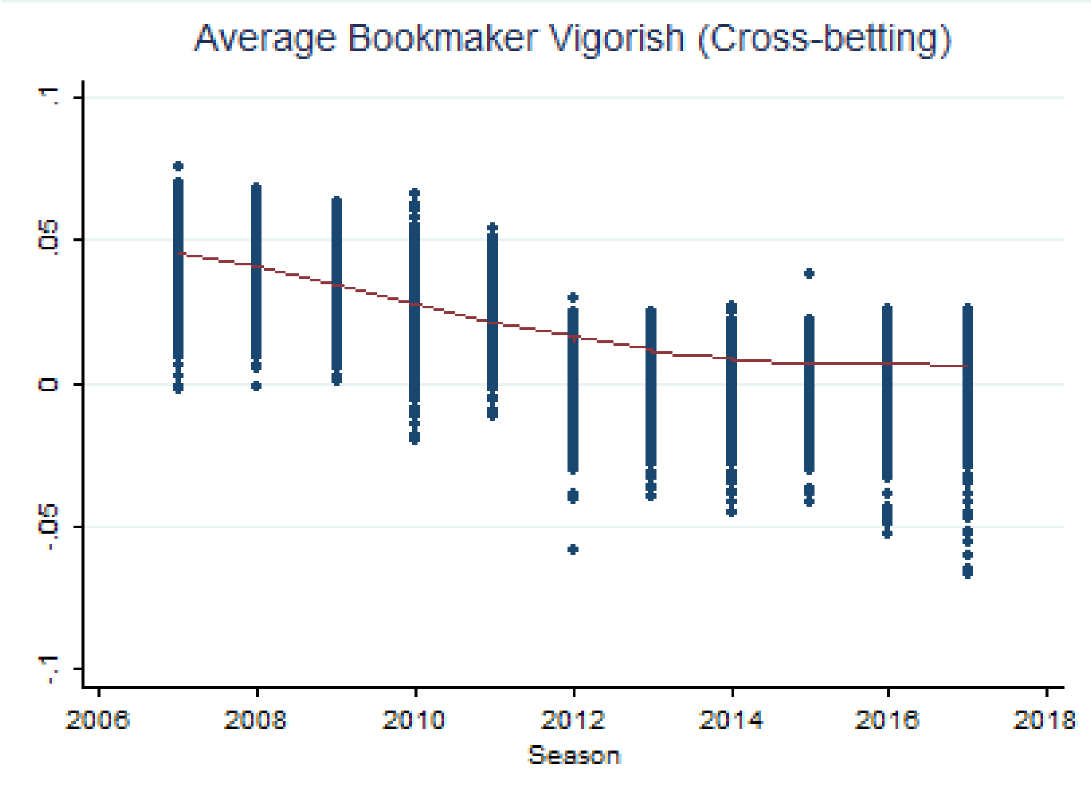


N=11,697

As can be seen on average the vigorish may have decreased but remains positive. This means that the vigorish in no instance is or has been negative for one single bookmaker (Analysis by League in Appendix; see Figure A.2). Further, there are no big differences in the vigorish charged between whether a team plays at home or away, which supports the assumption that bookmakers balance out their books.

However, when expanding the analysis and if the individual bettors are engaging into cross-betting among our six bookmakers (Figure 5), the average vigorish tends closer towards zero, while being negative for several bets taken. This negative vigorish is the result of the arbitrage opportunities found when cross-betting and strengthens the assumption that a lock-in effect should be existing in the industry as otherwise the anomaly should have been resolved.

Figure 5: Average Bookmaker Vigorish (Cross-betting)



N=11,726

In combination, the results of the development of the vigorish and the overround show that the allocative efficiency in the European online soccer betting market has increased over the seasons. While the overround serves as indicator for the price level of taking a bet for bettors (consumers), the vigorish demonstrates the economic profit of the individual bookmakers who act as suppliers and market-makers on the soccer betting markets.

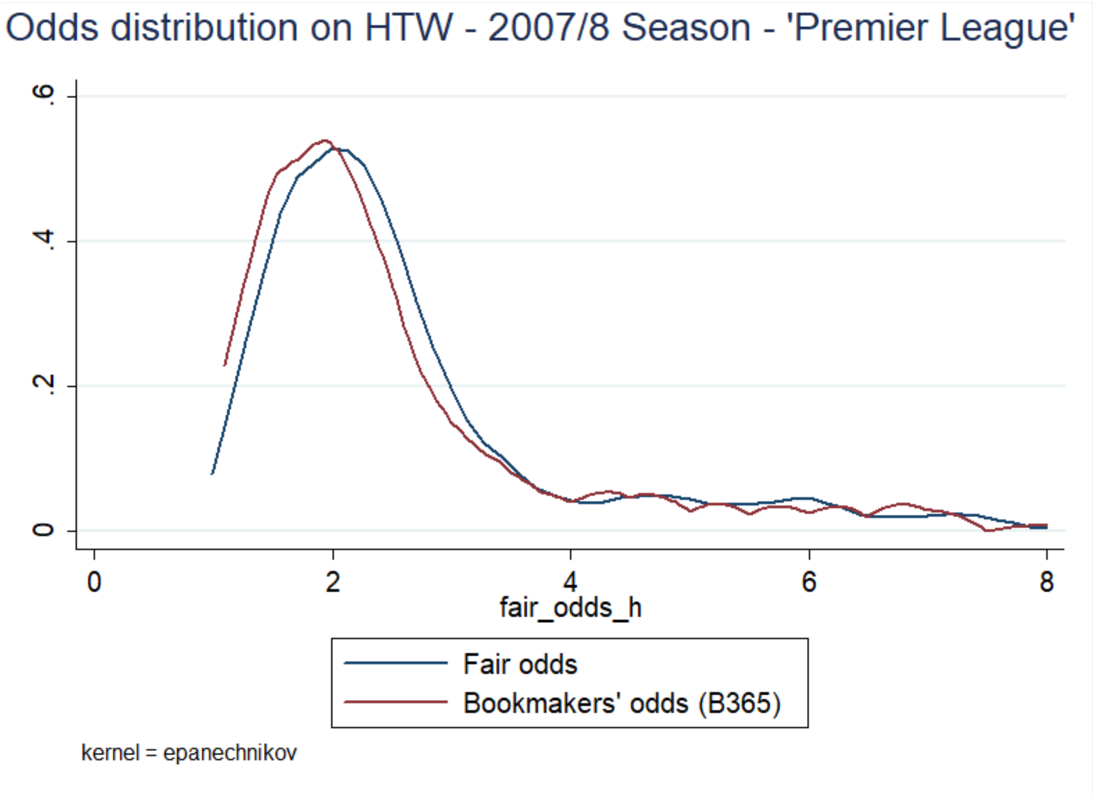
According with economic theory, the increased competitive pressure within the industry – induced by e.g. technological developments (M-commerce, increased Internet availability, Social Media, etc.) – has reduced the price level on the market, by lowering the suppliers economic profit which over time tends closer towards 0.

To verify the validity of the findings, whether the bookmaker's odds approached fair odds values. Theoretically – when the market becomes more allocative efficient, bookmaker odds on each outcome should approximate the actual fair odds.

The research findings show that bookmaker odds on outcome 'I' of event x_i approximated fair odds. This result was found by examining the correlation between the individual bookmakers' odds with the average fair odds on the different outcome probabilities. Over the seasons, the individual bookmaker odds converge towards the value of fair odds.

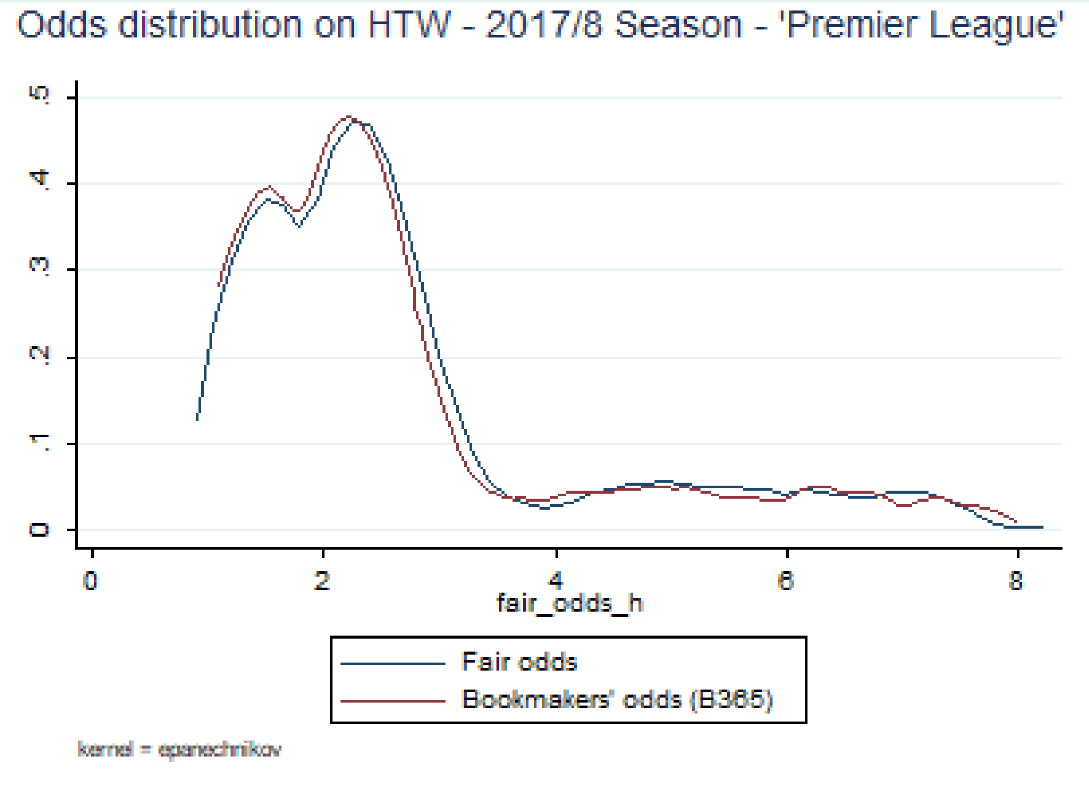
This can also be observed using the kernel density function. When comparing the individual bookmaker betting odds offered on outcome i of event x_i with the average fair odds values it is found that bookmakers odds approximate toward the real outcome odds/probabilities over the seasons (this is proven by the comparison of the results of the kernel density function of the Premier league for each outcome probability for the 2007, 2012 and 2017 season by comparing fair odds vs odds by B365 – see Figure 6, 7, 8 and 9).

Figure 6: Odds distribution on HTW - 2007/8 Season - 'Premier League'



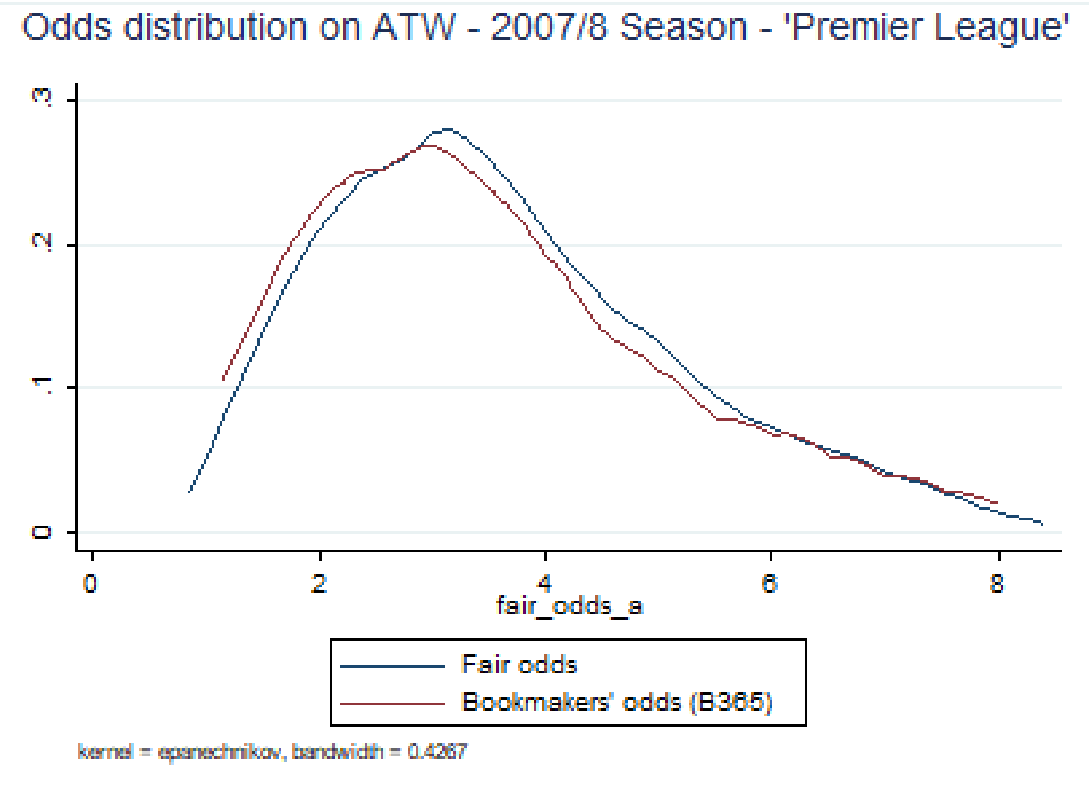
N=370 (for Fair odds & Bookmaker’s odds (B365))

Figure 7: Odds distribution on HTW - 2017/8 Season - 'Premier League'



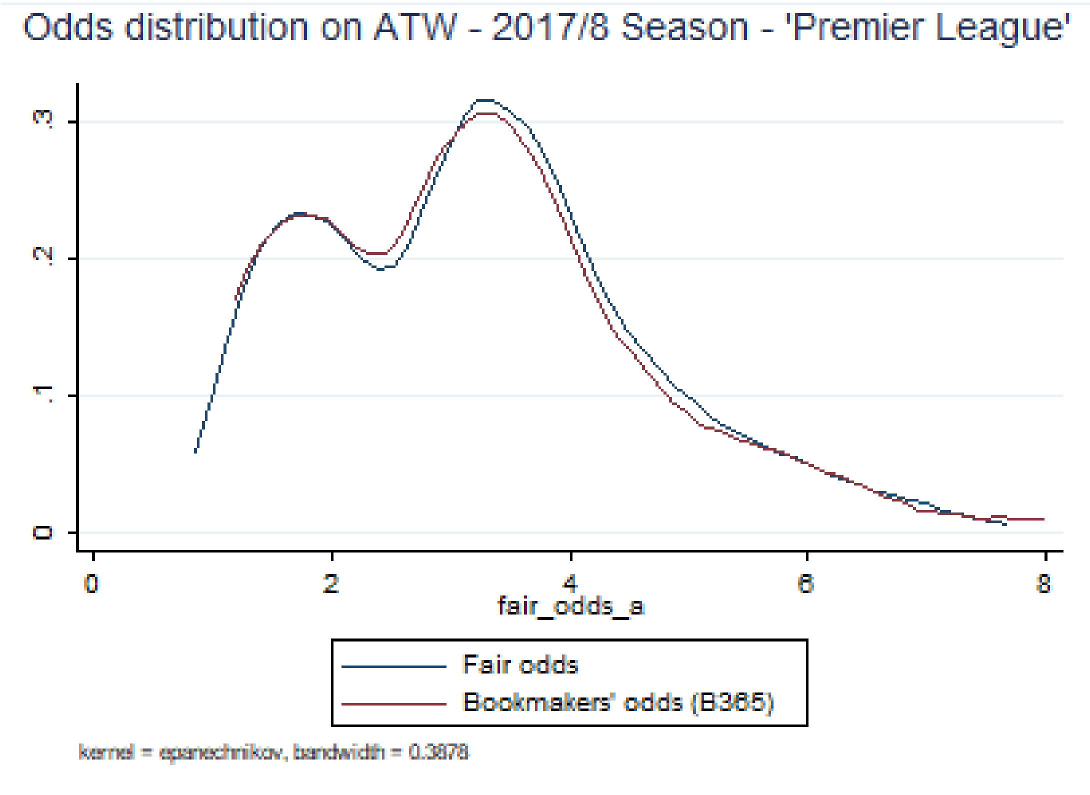
N=357 (for Fair odds & Bookmaker's odds (B365))

Figure 8: Odds distribution on ATW - 2007/8 Season - 'Premier League'



N=323 (Fair odds) & 326 (Bookmakers' odds (B365))

Figure 9: Odds distribution on ATW - 2017/8 Season - 'Premier League'



N=301 (Fair odds) & 303 (Bookmakers' odds (B365))

To sum up, the European online soccer betting market has become more efficient from the 2007/8 season to the 2017/8 season in terms of allocative efficiency. Although arbitrage opportunities in the market have increased (only when engaging into cross-betting), this can be rationalised with the increased competition between bookmakers trying to attract new customers. Further, the fact that it is necessary for the bettor to set up an account with each bookmaker and to pay in money via credit card before engaging into betting / different user interfaces (e.g. Betting app) seems to lead to a certain 'Lock-in' effect in the industry, leading to bettors settling with 'their bookmaker' rather than staying on the lookout for profit opportunities by engaging into cross-betting.

'Is it possible for bettors to develop a profitable betting strategy based on 'odds picking/cross-betting'?'

The second question of the empirical analysis concerns the possibility of developing a profitable betting strategy which allows above average-returns. According to the (information) market efficiency hypothesis by Eugene Fama, it shall not be possible for bettors to produce above average-returns on a constant basis since it should be 'impossible to beat the market'. This hypothesis by Fama is based upon the assumption that the market incorporates all available information to estimate and set the prices of a security/bet. Consequently, it should not be possible for an individual bettor to 'outperform' the market since all available information is already incorporated into the market prices.

However, due to the increased competition between the bookmakers and the fact that bookmaker odds tend to be closer to fair odds, the opportunity for positive returns for bettors could be increasing. To test this hypothesis, two betting strategies are developed with the aim to achieve positive returns, using the bookmaker consensus model as basis.

The first betting strategy is based on the estimated probability of event x_i by the bookmaker in comparison to the 'objective/real outcome estimation', calculated using the bookmaker consensus model for each outcome of event x_i after 'cleaning' the bookmaker odds. By 'cleaning' it is meant that the bookmakers underlying 'fair odds' for the outcomes I of event x_i have been calculated. This has been done under the assumption that the overround is evenly distributed among all outcome probabilities. Following this manipulation, a possible bet must meet two conditions to be considered a betting opportunity:

- The difference between the bookmakers' raw probability forecast on outcome x_i and the estimated outcome probabilities must be negative. When this value is negative, it means that the individual bookmaker is underestimating the probability of event x_i happening. This is an indicator that there might be a profit opportunity in the market.
- Secondly, the real estimated outcome probability of event x_i must be higher equal to 55%. This restriction is made under the light that each bettor only has restricted resources and consequently must make decisions regarding his budget. Further, a rational bettor would not engage in speculation and consequently not take the risk.

The results of the analysis are sorted by game outcome, as bettors have the possibility to bet on a home team win, away team win or draw and each outcome probability promises to have different outcome probabilities.

The second betting strategy is built up on the overround charged by the bookmakers for event x_i. For this, the following two assumptions are made:

- As a rational bettor is knowing the market prices, he tries to maximise his eventual pay-off by seeking for betting odds which are as close as possible to fair values. Therefore, it is assumed that a rational bettor will not engage into betting if the overround on event x_i is higher than 1%.

- Secondly, due to his restricted resources, the bettor will not engage into bets which's estimated outcome probabilities are lower than 0.65 or 65%.

In this case as well as in the prior, it is assumed that cross-betting is possible and that a rational individual will engage into 'odds picking' in order to boost his potential earnings (Therefore, choose the best odds allowing the highest return on the market). Also, the possible return of engaging into betting using the average market odds will be evaluated.

The result of the betting strategy 1 for the whole dataset, as well as per country is shown by Table 1:

Table 1: Returns of Betting Strategy 1

Betting strategy 1	Obs.	Return (Average odds)	Return (Maximal odds)
Home Team-win	74	12.57%	19.60%
German League	29	31.94%	40.06%
English League	22	-14.27%	-9.23%
Spanish League	23	13.84%	21.39%
Away Team-win	46	4.74%	11.15%
German League	24	20.90%	28.88%
English League	14	-26.71%	-22.85%
Spanish League	8	11.24%	17.50%

N='Obs.'

The results indicate that especially bets on home team wins using this strategy promise positive returns. The underestimation of the home team winning by bookmakers supports the theory of a FLB existing within the European soccer betting market, as found priory by Oikonomidis, Bruce and Johnson (2012).

When checking for cross-country differences, it is found that while this betting strategy returns positive returns for the German and the Spanish league, its returns are negative for the English league on home team win as away team win. The negative returns in the English betting market are victim to the increased decisiveness of the factor chance on game outcomes. As proven by the researches of Quitzau (2005), Quitzau and Vöpel (2009) the unpredictability of games in the English league is higher than in other European top leagues, due to the tight concurrency within the top-flight. While this might lead to a more interesting league/games, it also leads to more unpredictability of outcomes since games are more often decided by the factor of chance than in the other leagues. While this lowers the chance for a general model to produce positive returns, it might allow insiders or expert gamblers to achieve positive returns. Further, the results suggest that 'odds-picking' allow the bettor to maximise his returns, while reducing his losses as in the case on bets on the English market.

While the second betting strategy covers far more games than the first, its possible returns are lower (Table 2):

Table 2: Returns of Betting Strategy 2

Betting strategy 2	Obs	Return (Average odds)	Return (Maximal odds)
Home Team-win	583	0.02%	2.67%
German League	118	-5.05%	-2.37%
English League	259	0.85%	2.37%
Spanish League	206	1.86%	5.95%
Away Team-win	190	-5.75%	-0.52%
German League	64	3.28%	3.79%
English League	54	-5.49%	14.24%
Spanish League	72	-14.06%	-15.41%
Draw	7		-0.44%

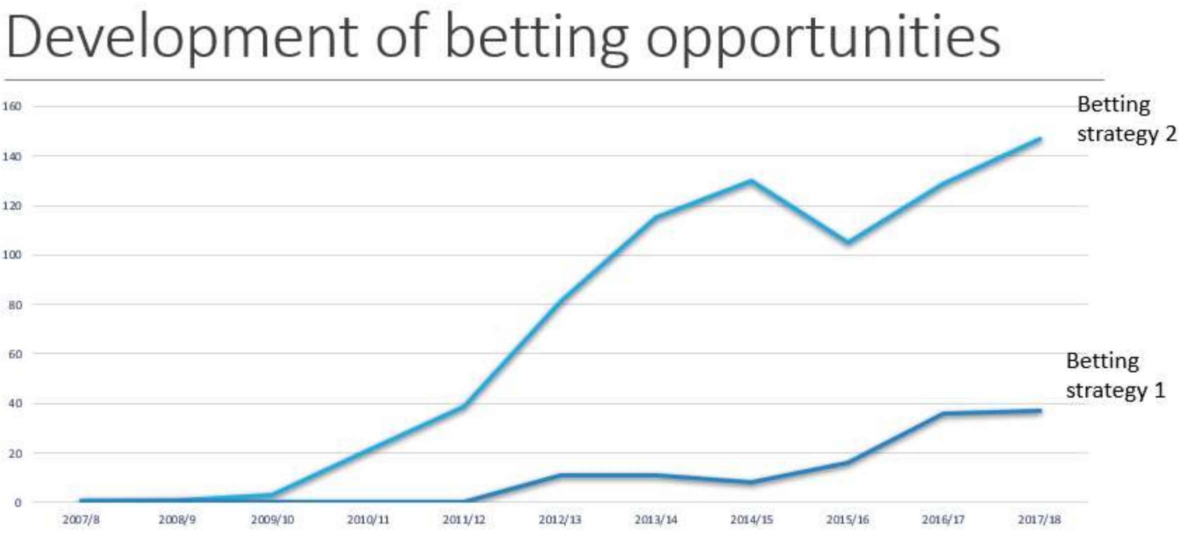
N='Obs.'

As can be seen, betting strategy 2 is only borderline successful, only promising an average return when engaging into 'odds picking' on home-team wins. The difference between the returns of the average betting odds and the maximal odds when betting on an away team win in the English market are worth to be mentioned. This indicate that in England– opposite to the other markets – away teams are likely to be underrated. Another cross-country difference can be observed when comparing returns on the German- to the English- and Spanish league. While the returns when betting on Home team win are negative for the German league, and the returns are positive when betting on away team wins, the opposite is true for the English- and Spanish league. This shows that while in the German league the home team advantage is overrated, bookmakers tend to underestimate the home team advantage for the English- and the Spanish league.

Additionally, the difference between the return when betting on average odds and maximal odds, using ‘odds picking’, confirms that ‘odds picking’ can boost a bettor’s returns. The property of bettors achieving higher returns while taking the same bet(s) allows them to lower their risk as well as maximise their profits.

Another point of interest is the development of the number of possible betting opportunities. In theory, induced by the increased competition among bookmakers – betting opportunities for bettors should have been increasing over the seasons. Figure 10 captures the development of the number of betting opportunities:

Figure 10: Development of betting opportunities



N=893

As can be seen in Figure 10, the number of betting opportunities has increased significantly in the same years’ bookmaker odds approached fair odds.

Because betting strategy 1 is depending on the difference between the bookmakers’ odds and the real underlying outcome probabilities, its development can be interpreted as more betting opportunities arising since bookmaker odds became closer to the real underlying outcome probabilities. Betting strategy 2 on the other side demonstrates that the bookmakers overround got closer to fair odds, reflected in increased opportunity for bettors.

To sum up, the graph demonstrates that the increased competition on supply side has led to an increase in betting opportunities for bettors.

Further, the analysis proved the positive implications of odds picking/cross-betting on an individual bettors performance while the first betting strategy exhibited positive returns in aggregate and demonstrated that the information market efficiency of the European online soccer betting market might be put into question. However, the low number of betting opportunities in this case (120 observations; see Figure 10) only allows to deduct a trend but does not allow any inference. Conducting this analysis with a larger dataset and more recent data might support the findings, based on the increasing number of opportunities the bettor can encounter using betting strategy 1.

5. Conclusion

The research findings provide an argument that the European online soccer betting markets allocative efficiency has increased over the seasons. This is observed by the development of the overround and vigorish over the studied seasons. While the overround is representative for the overall price level for taking a bet the vigorish represents the economic profit of the suppliers. According to economic theory, increased competition on the supply side has led to a decreasing vigorish resulting in a decreasing overround over the seasons. With bookmakers betting odds approaching fair values, it can be argued that the competition has led to higher allocative efficiency of the European online soccer betting market. A side effect of the increased competition on the supply side is the increasing number of arbitrage opportunities on all studied markets from the 2007/8 to the 2017/8 season when engaging into 'odds picking'/cross-betting. This finding can be seen as the result of bookmakers competing to attract new customers. This is consistent with the findings of Forrest and Simmons (2008) who suggest that the increased competition on the supply side might lead to possible risk-free profit in the case of cross-betting. Further, the increasing number of arbitrage opportunities over time indicate a possible lock-in effect of bettors with their bookmaker, as priory found by Franck, Verbeek and Nüesch (2010).

The second research question was focused on the development of a profitable betting strategy, based on the bookmaker consensus model by Leitner, Zeileis and Hornik (2010). The results indicate an increasing number of betting opportunities for bettors, promoted by the competition to attract new customers by the bookmakers.

Further, the findings support the theory of an existing positive Favourite Longshot Bias by Cain, Law and Peel (2000) which seems to exist in the German- and Spanish League. The results also support the theory of Quitzau and Vöpel (2009), that the factor of chance is more game-decisive in the English Premier League than in the other European Top Leagues.

To conclude, betting strategy 1 proved to be profitable (Table 1), with the average return of 12.57% when betting on a home team win when using average odds and 19.6% when engaging into 'odds picking'. The average return using average odds are 4,74% when betting on an away team win and 11.15% when using the method of 'odds picking'. However, since the number of betting opportunities using betting strategy 1 was rather low, further research will be necessary using a bigger sample size. Betting strategy 2 proved to be borderline successful, with the return on average odds on a home team win being 0.02% and maximal odds 2,67%, compared to return on average odds on an away team win of -5,75% and -0.52% when engaging into 'odds picking'.

The findings prove that 'odds picking' could be the base for developing a profitable betting strategy as it allows the bettor to boost his potential return while minimizing his potential loss.

Bibliography

Akintoye, Ishola Rufus (2008): Efficient Market Hypothesis and Behavioural Finance: A Review of Literature. European Journal of Social Sciences, Vol. 7 (2)

Ali, Mukhtar M. (1977): Probability and utility estimates for racetrack bettors. Journal of Political Economy, Vol. 85 (August), p. 803-815

Arnott, Robert D; Hsu, Jason;, Kalesnik, Vitali and Tindall,Phil (2013): The Surprising Alpha from Malkiel's Monkey and Upside-Down Strategies. The Journal of Portfolio Management, Volume 39, Number 4 – Summer 2013, p. 91-106

Avery, Christopher; Chevalier, Judith(1999): Identifying Investor Sentiment from Price Paths: The Case of Football Betting. The Journal of Business, Vol. 72, No.4 (October 1999), p. 493-521

Baum, Christopher F. (2011): Introduction to Stata. Faculty Micro Resource Center, Boston College

Boulier, Bryan L. and Stekler, H.O. (2003): Predicting the outcomes of National Football League games. International Journal of Forecasting, Vol. 19 (2), p. 257 - 270

CAIN, Michael; LAW, David; PEEL, David (2000): The favourite-longshot bias and market efficiency in UK Football betting, in Scottish Journal of Political Economy, Vol. 47, No.1, February 2000

Chan, Wesley S. (2003): Stock Price Reaction to News and No-News: Drift and Reversal after Headlines. Journal of Financial Economics, Vol. 70, p. 223 - 360

Chevallier, Judith and Ellison, Glenn (1999): Ae Some Mutual Fund Managers Better Than Others? Cross-Sectional Patterns in Behavior and Performance. The Journal of Finance, Vol. 54 (3), p. 875 - 899

Choi, Darwin & Hui, Sam K. (2012) : The Role of Surprise: Understanding Overreaction and Underreaction to Unanticipated Events using In-Play Soccer Betting Market

Coleman, Les (2011): An Exploratory Analysis of Factors Influencing Initial Market Response and Media Reports following Shock Corporate Events. The Financiel Review, Vol. 46, p. 313 - 336

Crafts, N. F. R. (1985): Some evidence of insider knowledge in horse race betting in Britain. Economica, Vol. 52, p. 295 - 304

Crow, David (2016): Bill Ackman admits fault in Valeant investment. Financial Times. URL: https://www.ft.com/content/d2ffb65a-0cbe-11e6-b41f-0beb7e589515 (Date of access: 21.06.2019)

Croxon, Karen and Reade, J. James (2013): Information and Efficiency: Goal Arrival in Soccer Betting. The Economic Journal, Vol. 124, p. 62 - 91

Culp, Christopher L. (2015): OTC-Cleared Derivatives: Benefits, Costs, and Implications of the 'Dodd-Frank Wall Street Reform and Consumer Protection Act'. Journal of Applied Finance, Vol.20(2)

Dixon, Mark and Pope, Peter (2004): The value of statistical forecasts in the UK association football betting market. International Journal of Forecasting, Vol. 20, p. 697 - 711

Dixon, Mark J. and Coles, Stuart G. (1997): Modelling Association Football Scores and Inefficiencies in the Football Betting Market. Applied Statistics, Volume 46, Issue 2 (1997), p.265-280

Dobson, Stephen and Goddard, John (2001): The Economics of Football. Cambridge University Press

Fama, Eugene F. (1970): Efficient Capital Markets: A Review of Theory and Empirical Work. Journal of Finance, Vol. 25, p. 383 - 417

Fama, Eugene F. (1991): Efficient Capital Markets: II. Journal of Finance, Vol. 46, p. 1575 - 1617

Fama, Eugene F. (1997): Market efficiency, long-term returns, and behavioral finance. Journal of Financial Economics, Vol. 49, p. 283 - 306

Forrest, David and Simmons, Robert (2000): Forecasting sport: The behaviour and performance of football tipsters. International Journal of Forecasting, Vol. 16, p. 317 - 331

Forrest, David and Simmons, Robert (2008): Sentiment in the betting market on Spanish football, Applied Economics, 40:1, 119-126

Forrest, David; Goddard, John and Simmons, Robert (2005): Odds-setters as forecasters: The case of English football. International Journal of Forecasting, Vol. 21, p. 551 - 564

Franck, Egon; Verbeek, Erwin and Nüesch, Stephan (2010): Prediction accuracy of different market structures – bookmakers versus a betting exchange. International Journal of Forecasting, Vol. 26, p. 448 - 459

Fuhrmann, W. (1988): Theorie effizienter Finanzmärkte.

Gil, Ricard and Levitt, Stephen D. (2007): Testing the Efficiency of markets in the 2002 World Cup. The Journal of Prediction Markets, Vol. 1 (3), p. 255 - 270

Goddard, John and Asmikapolous, Ioannis (2004): Modelling football match results and the efficiency of fixed-odds betting. Journal of Forecasting, Vol. 23, p.51 - 66

Goddard, John and Dobson, Stephen (2001): The Economics of Football. Cambridge University Press

Golec, Joseph and Tamarkin, Maurry (1995): Do bettors prefer long shots because they are risk-lovers, or are they just overconfident? Journal of Risk and Uncertainty, Vol. 11 (1), p. 51 - 64

Golec, Joseph and Tamarking, Maurry (1992): The degree of inefficiency in the football betting market. Journal of Financial Economics, Vol. 30 (1991), p. 311 – 323

Gomber, Peter; Rohr, Peter and Schweikert, Uwe (2008): Sports betting as a new asset class – current market organization and options for development. Financial Markets and Portfolio Management, Vol. 22(2), p. 169 - 192

Gorgieva-Trajkovska, Olivera and Adzipetrov, Robert (2016): Moral Hazard as One of the Main Reasons for the Financial Crisis of 2007. Journal of Economics, Vol 4 (1)

Gough, Christine (2019): Sports Betting and Gambling Market/Industry – Statistics&Facts. Statista.

Griffith, Richard M. (1949): Odds adjustments by American horse-race bettors. The American Journal of Psychology, Vol. 62.2, p.290 - 294

H.L. Wilgus, Purchase of Shares of Corporation by a Director from a Shareholder, 8 MICH. L. REV. 267, 297 (1910)

Haddock, David D. and Macey, Jonathan R. (1987): Regulation on Demand: A Private Interest Model, with an Application to Insider Trading Regulation. The Journal of Law and Economics, Vol. 30(2), p. 311 - 352

Hanson, Robin; Oprea, Ryan; Porter, David (2004) : Information aggregation and manipulation in an experimental market. Journal of Economic Behaviour&Organization Vol.60 (2006), 449-459

https://www.investopedia.com/terms/a/arbitrage.asp (As of 08.05.2018)

Jaffe, Jeffrey F. and Winkler, Robert L. (1976): Optimal Speculation against an Efficient Market. The Journal of Finance, Vol. 31(1), p. 49 - 61

Kahneman, Daniel and Tversky, Amos (2013): Prospect Theory: An Analysis of Decision Under Risk. Handbook of the fundamentals of Financial Decision making (vol. 4), Chapter 6.

Kapeller, Jakob (2008): Das Menschenbild moderner Ökonomie. Trauner Verlag, Linz.

Katzke, Nico and Bothma, Christiaan (2014): The Monkey Prevail: Questioning the Efficiency of Cap-Weighted Indexes

Kruypers, Tim (2000): Information and efficiency: an empirical study of a fixed odds betting market. Applied Economics, Vol. 32, p. 1353 - 1363

Langer, Ellen J. and Roth, Jane (1975): Heads I win, tails it's chance: The illusion of control as a function of the sequences of outcomes in a purely chance task. Journal of Personality and Social Psychology, Vol. 32(6), p. 951 - 955

LAW, David and PEEL David A. (2001): Insider Trading, Herding Behaviour and Market Plungers in the British Horse-Race Betting Market. In: Economica 69 (2002) p.327-338

Leitner, Christoph; Zeileis, Achim and Hornik, Kurt (2010): Forecasting sports tournaments by ratings of (prob)abilities: A comparison of the EURO 2008. International Journal of Forecasting 26, p. 471-481.

Leitner, Christoph; Zeileis, Achim; Hornik, Kurt (2010): Forecasting sports tournaments by ratings of (prob)abilities: A comparison for the EURO 2008. International Journal of Forecasting 26, 2010, p.471-481

Levitt, Steven D. and List, John A. (2007): What do Laboratory Experiments Measuring Social Preferences Reveal about the Real World? Journal of Economic Perspectives, Vol. 21, p. 153 - 174

Levitt, Steven D: (2004): Why are gambling markets organised so differently from financial markets? The economic journal, 114 (April) 223-246.

Malkiel (1999): Superior performance of index funds in comparison to actively managed mutual fund

Malkiel, Burton G. (2005): Reflections on the Efficient Market Hypothesis: 30 Years Later

Mas-Colell, Andreu; Whinston, Michael D. and Green, Jerry R. (1995): Microeconomic Theory.
Wolfers, Justin and Zitzewitz, Eric (2004): Prediction Markets. Journal of Economic Perspectives, Vol. 18 (2), p. 107 - 126

McGlothlin, William H. (1956): Stability of choices among uncertain alternatives. The American Journal of Psychology, Vol. 69.4, p. 604 - 615

Neale, Walter C. (1964): The peculiar economics of professional sports. The Quarterly Journal of Economics, Vol. 78 (1), p. 1 - 14

Nystedt, Jens (2004): Derivative Market Competition: OTC Markets Versus Organized Derivative Exchanges. IMF Working Paper No. WP/04/61

Oikonomidis, Anastasios; Bruce, Alistair C.; Johnson, Johnnie E.V. (2012): Does transparency imply efficiency? The case of the European soccer betting market.

Palomino, F. A., Renneboog, L. D. R., & Zhang, C. (2008). Information Salience, Investor Sentiment, and Stock Returns: The Case of British Soccer Betting. (Center Discussion Paper; Vol. 2008-99). Tilburg: Finance.

Paton, David; Vaughan Williams, Leighton and Fraser, Stuart (1999): Regulating Insider Trading in Betting Markets. Bulletin of Economic Research, Vol. 51 (3), p. 237 - 241

Paul, Rodney J. and Weinbach, Andrew P. (2004): Efficient Markets, fair bets, and profitability in NBA total 1995-96 to 2001-02. The Quarterly Review of Economics and Finance, Vol. 44(4), p. 624 - 632

Peirson, John; Smith, Michael A. (2009): Expert Analysis and Insider Information in Horse betting: Regulating informed market behaviour

Pershing Square Capital Management, L.P. (2013): https://assets.factsaboutherbalife.com/content/uploads/2013/01/09145402/Who-wants-to-be-a-Millionaire.pdf

Pezzo, Mark (2003): Surprise, defence, or making sense: What removes hindsight bias? Journal of Memory, Vol. 11, p. 421 - 441

Plott, Charles R. and Sunder, Shyam (1988): Rational Expectations and the Aggregation of Diverse Information in Laboratory Security Markets. Econometrica, Vol. 56 (5), p. 1085 - 1118

Pope, Peter F. and Peel, David A. (1989): Information, prices and efficiency in fixed-odds betting market. Economica, Vol. 56, p. 323 - 341

Pope, Peter F. and Peel, David A. (1990): Insider Trading: Some Evidence on Market Efficiency and directors` share dealings in Great Britain. Journal of Business Finance & Accounting, Vol. 17 (3), p. 359 - 380

Quitzau, Jörn (2005) : Faktor Zufall als Spielverderber: zur Prognostizierbarkeit von Fussballergebnissen – Wettmärkte als effizienter Informationslieferant, Research Notes, No. 18

Quitzau, Jörn; Vöpel, Henning (2009): Der Faktor Zufall im Fußball – Eine empirische Untersuchung für die Saison 2007/08. In: HWWI Research; Paper 1-22 des HWWI Kompetenzbereiches Wirtschaftliche Trends

Reade, Karen (2014): Information and predictability: bookmakers, prediction markets and tipsters as forecasters. Journal of Prediction Markets, Vol. 8, p. 43 - 76

Robinson, Sarah (2001): Play Money? Not to these news Investors. In: New York Times.

Rottenberg, Simon (1956): The baseball players' labor market, Journal of Political Economy, Vol. 64(3)

Schnytzer, Adi and Shilony, Yuval (1995): Inside Information in a Betting Market. The Economical Journal, Vol. 105, Issue 431, p. 963 - 971

Servan-Schreiber, Emile; Wolfers, Justin; Pennock, David M.; Galebach, Brian (2004): Prediction Markets: Does Money Matter? In Electronic Markets Vol. 14 No 3

Sharma, Prateek and Paul, Samit (2015),'Testing the skill of mutual fund managers: evidence from India', Managerial Finance, Vol. 41 Iss 8 pp. 806 – 824

Shiller, Robert J. (1998): Human Behavior and the Efficiency of the Financial System. NBER Working Paper No. 6375

Shiller, Robert J. (2003): From Efficient Markets Theory to Behavioral Finance. Journal of Economic Perspectives, Vol. 17 (1), p. 83 - 104

Shin, Hyun Song (1991): Optimal Betting Odds against Insider Traders. In: The Economic Journal, Volume 101, Issue 408, 1179-1185

Shin, Hyun Song (1992): Prices of state contingent claims with insider traders, and the favourite longshot bias. Economic Journal, Vol. 102 (March), p. 426 – 435

Shin, Hyun Song (1993): Measuring the incidence of insider trading in a market of state-contingent claims. Economic Journal, Vol. 102 (September), p. 1141 – 1153

Sloane, Peter J. (1971): The economics of professional football: the football club as a utility maximiser. Scottish Journal of Political Economy, Vol. 18 (2), p. 121 – 146

Smith, Michael A. and Vaughan Williams, Leighton (2010): Forecasting horse race outcomes: New evidence on odds bias in UK betting markets. International Journal of Forecasting 26, 2010, p. 543-550

Smith, Michael A.; Paton, David and Vaughan Williams, Leighton (2009): Do bookmakers possess superior skill to bettors in predicting outcomes? Journal of Economic Behaviour & Organization, Vol. 71 (2), p. 539 - 549

Snigaroff, Robert (2012): The Myth of the Monkey. The journal of Portfolio-Management, Volume 39 Number 1, 2012, p.136 – 142

Song, ChiUng; Boulier, Bryan L. and Stekler, Hermann O. (2007): The comparative accuracy of judgmental and model forecasts of American football games. International Journal of Forecasting, Vol. 23, p. 405 - 413

Spann, Martin and Skiera, Bernd (2009): Sports forecasting: A comparison of the forecast accuracy of prediction markets, betting odds and tipsters. Journal of Forecasting, Vol. 28, p. 55 - 72

Telser, Lester G.; Higinbotham, Harlow N. (1977): Organized Futures Markets: Costs and Benefits. Journal of Political Economy, Vol. 85, No. 5

Thaler and Ziemba (1988): Anomalies – Parimutuel Betting Markets: Racetracks and Lotteries. Journal of Economic Perspective; Volume 2, Number 2 – Spring 1988; p.161-174

Thaler, Richard J. (2015): Misbehaving: The Making of Behavioural Economics. Norton & Company New York

Thaler, Richard J. (2006) in: 'Are markets efficient?', URL: https://youtu.be/bM9bYOBuKF4 (As of 31/07/2019)

Varian, Hal (2010): Intermediate Microeconomics. W.W. Norton & Company, New York

Vaugahn Williams, Leighton (1999): Information efficiency in betting markets. Bulletin of Economic Research, Vol. 51 (1), p. 1 - 30

Vaughan Williams, Leighton (2005): Information efficiency in financial and betting markets. Cambridge University press

Vlastikis, Nikolaos, Dotsis, George and Markellos, Raphael N. (2009): How efficient is the European Football Betting Market? Evidence from Arbitrage and Trading Strategies. Journal of Forecasting, Volume 28, p. 426 - 444

Wolfers, Justin and Zitzewitz, Eric (2004): Prediction Markets. Journal of Economic Perspectives, Vol. 18, p. 107 – 126

Appendix

Betting Odds

The below variables – sorted by bookmaker - represent the odds offered by bookmaker x_i on game y in a given season z on:

1. Win of the Home team (H),
2. Draw between the teams (D) or
3. Win of the Away team (A).

In the first line the meaning of the different variables for bookmaker x_i is outlined while the other follow the same scheme and only the name of the respective bookmaker x_i is mentioned:

- Odds on game x in season y from bookmaker Bet365 on:
  1. B365H- Bet365 odds on a win of the Home side
  2. B365D- Bet365 odds on a win of the Away side
  3. B365A- Bet365 odds on draw
- Bet&Win
  1. BWH
  2. BWD
  3. BWA
- Blue Square
  1. BSH
  2. BSD
  3. BSA
- Gamebrookers
  1. GBH
  2. GBD
  3. GBA
- Interwetten
  1. IWH
  2. IWD
  3. IWA
- Ladbrokes
  1. LBH
  2. LBD
  3. LBA
- Pinnacle Sports Betting
  1. PSH

2. PSD
3. PSA

- Sports betting
  1. SBH
  2. SBD
  3. SBA
- Stan James
  1. SJH
  2. SJD
  3. SJA
- William Hill
  1. WHH
  2. WHD
  3. WHA
- Victor Chandler
  1. VCH
  2. VCD
  3. VCA

Additionally, the dataset includes a Maximum and Average odd offered on a game y of bookmakers x_i. This data is drawn from Bet brain -a betting odds comparing web page. In this case, the variable Bb1X2 stands for the number of bookmakers used to calculate these respective values. In the dataset the variables are named as below:

- Highest odd offered on outcome z in game x:
  1. BbMxH - Highest odds offered on Home win
  2. BbMxD - Highest odds offered on a Draw between the two sides
  3. BbMxA - Highest odds offered on Away win
- Average odds offered on outcome z in game x:
  1. BbAvH- Highest odds offered on Home win
  2. BbAvD- Highest odds offered on a Draw between the two sides
  3. BbAvA- Highest odds offered on Away win

Last, the Pinnacle Closing odds – the last odds before a match starts – are given by the below variables:

- PSCH – Pinnacle Closing odds on Home win

- PSCD – Pinnacle Closing odds on Draw
- PSCA - Pinnacle Closing odds on Away win

Game Statistics

Below follows a description of the game statistic variables available in the dataset and their meaning:

- *Season - Season in which games were played*
  - In the national leagues the horizon over which the league tournament takes place is split into seasons. At the end of each season it is known which team won the league, scored the most goals etc. Further, at the end of each season the composition of teams within the league changes (due to promotions and relegations) and clubs have the possibility to engage into player transfers (buy/sell). Consequently, the empirical observation of this thesis will be split into seasons. The research question will be investigated for each single season. Further, the results of the single seasons will be compared with each other.[15]
- *Spieltag - Game day*
  - The game day indicates at which instance in time the relevant game takes place. At the beginning of the season, there is not much data available on historic results of the respective teams. Consequently, it can be assumed that uncertainty for the outcome of a game is higher at the beginning of a season than at the middle or towards the end.
- *Div - Division/League*
  - Indicates in which Division/League the game is taking place. In our sample, this value can be either 'PL' for the English *'Premier League'*, 'SL' for the Spanish *'La Liga'* or 'DB' for the German *'Deutsche Bundesliga'*
- *Date - Date when the game was played*
  - Usually the games of a game day are distributed over the course of on one to two days. These days are usually Saturday and Sunday (as on these days most attendees are off work and fan attendance is higher). Games are usually distributed over these days as there is TV coverage for most of them. However, fixtures can also occasionally take place on week days.
- HomeTeam - Name of the Team playing at Home

  Team which is playing in its stadium. In sport it is assumed that the home team has a certain 'Home Team Advantage' when playing at their home stadium: As their

[15] Due to the data being panel data, and the measurement unit being seasons (instead of years).

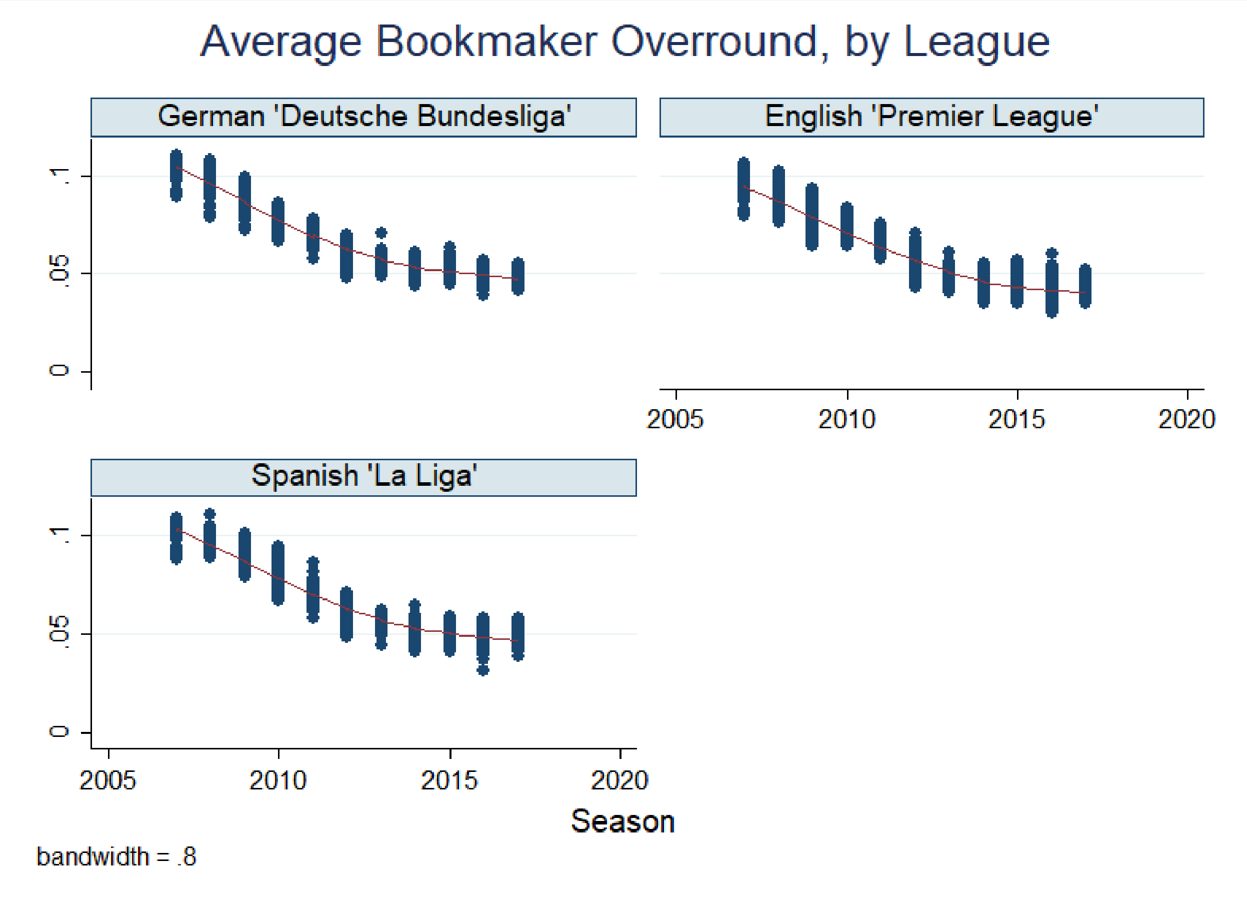
supporters are in place and cheering them, they know the field, etc. On the other hand, this factor is very often overestimated as research by e.g. Vlastikis, Dotsis and Markellos (2009) shows.

- AwayTeam - Name of the Away Team
- FTHG - Full Time Home-Team Goals
  - Goals scored by the Home team in a certain game x at Game day y facing opponent z.
  - Together with FTAG, this data allows to see how many goals a team scores per match in average. Further, it allows us to see e.g. whether a Team scores more goals when playing at home than when playing away and therefore if the 'Home Effect' is relevant for a respective team in a certain season.
- FTAG - Full Time Away-Team Goals
  - Goals scored by the Away team in a certain game x at Game day y facing opponent z.
- FTR - Full Time Result
  - The outcome of a game. Realization can be either a Home Win, Away Win or a draw between the teams.
- HTHG - Half Time Home-Team Goals
  - Goals scored at Half time by the Home team in a certain game x at Game day y facing opponent z.
- HTAG - Half Time Away-Team Goals
  - Goals scored at Half time by the Home team in a certain game x at Game day y facing opponent z.
- HTR - Half Time Result
  - The score between Team A and B in a certain game x at Game day y at half time
- HS - Home Team Shots
  - Shots by players of the Home Team within game x
- AS - Away Team Shots
  - Shots by players of the Away Team within game x
- HST - Home Team Shots on target
  - Shots by players of the Home Team which were on goal of the opponents within game x
  - Can be used as indicator as of e.g. how effective a Team is in scoring goals (e.g. Shots-Goal Ratio)
- AST - Away Team Shots on target
  - Shots by players of the Away Team on goal of the opponents within game x
- HF - Home Team Fouls

- Fouls committed by the Home Team
- AF - Away Team Fouls
  - Fouls committed by the Away Team
- HC - Home Team Corners
- AC - Away Team Corner
- HY - Home Team Yellow Cards
  - Sum of Yellow cards received in a match. Indicates playing style/aggressiveness of a side, but also is an indicator for the severeness of fouls committed by a team.
- AY - Away Team Yellow Cards
- HR - Home Team Red Cards
  - Sum or Red cards received in a match. A red card for a player means the exclusion of the game and therefore is a big disadvantage. Many red cards stand for an unfair playing team, or a team who is not able to solve critical situations without a foul.
- AR - Away Team Red Cards

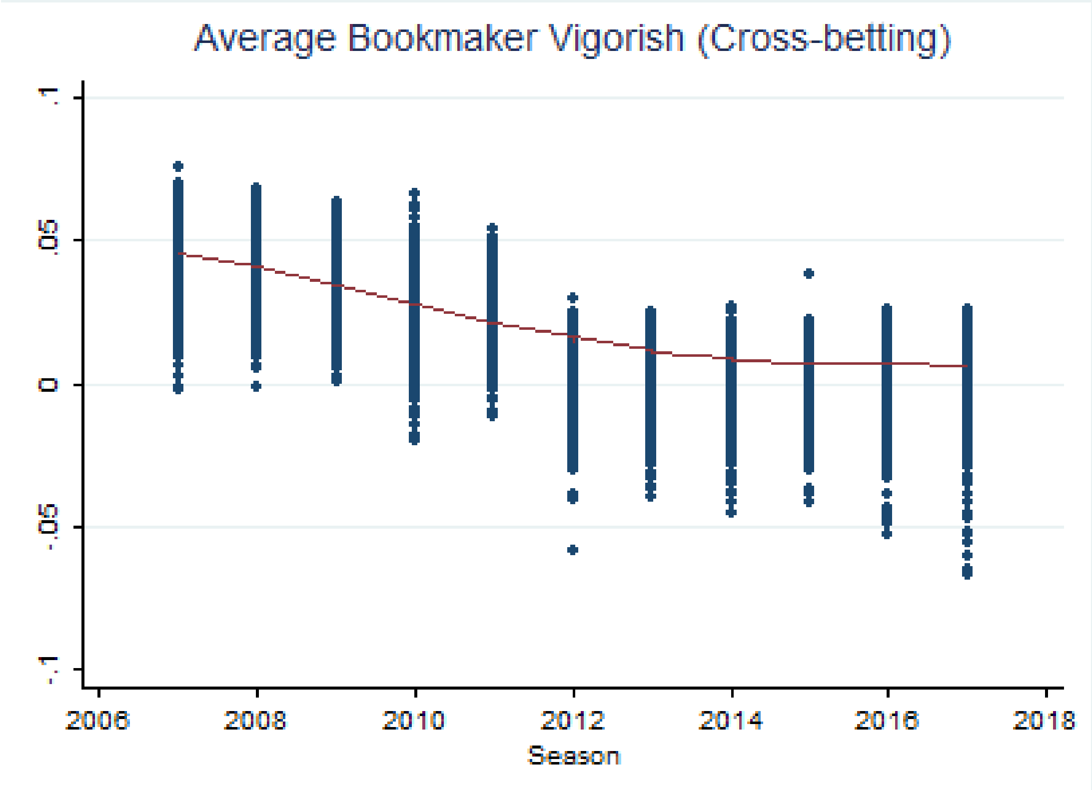
Empirical results/ Figures

Figure A.11: Average Bookmaker Overround, by League



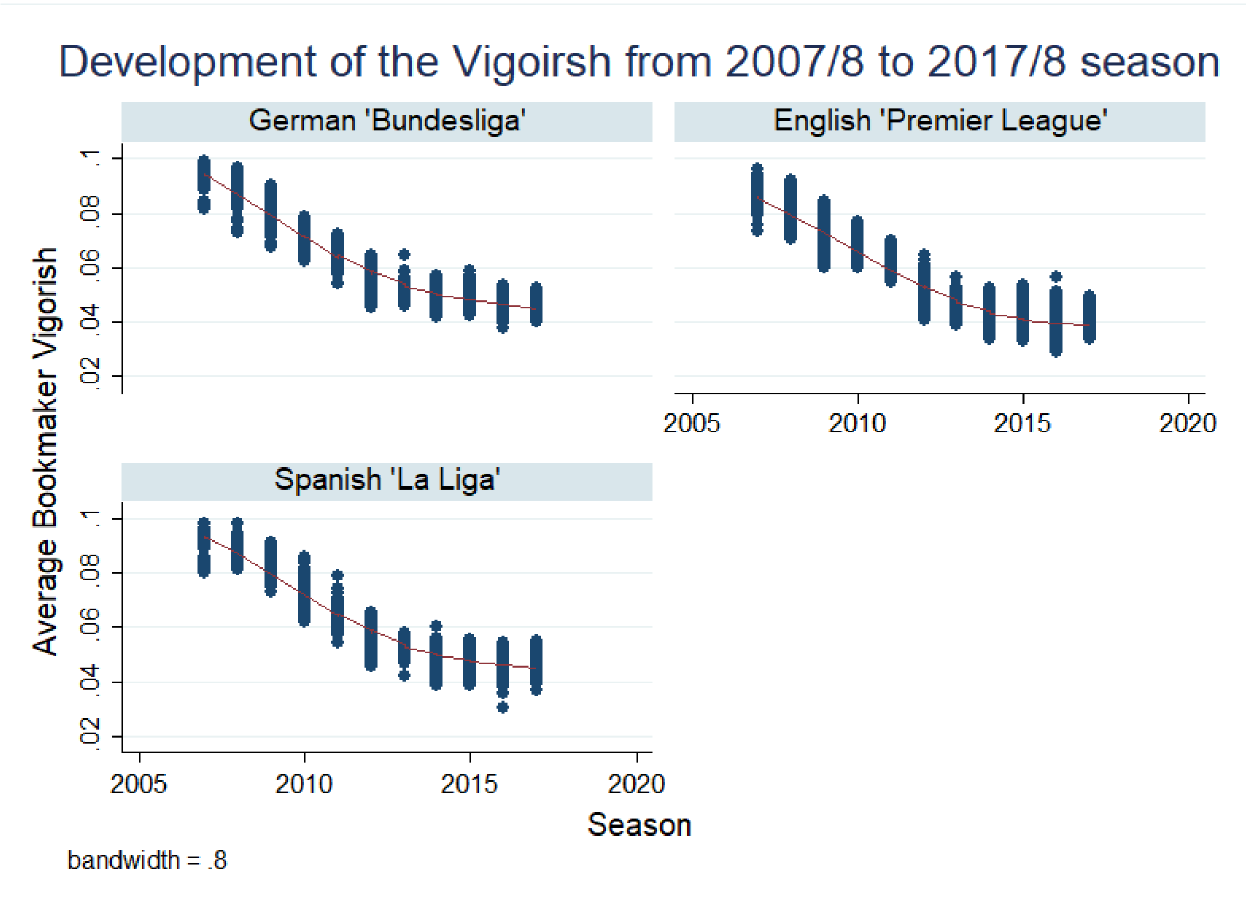
N=11,697

Figure A.12: Average Bookmaker Vigorish (Cross-betting)



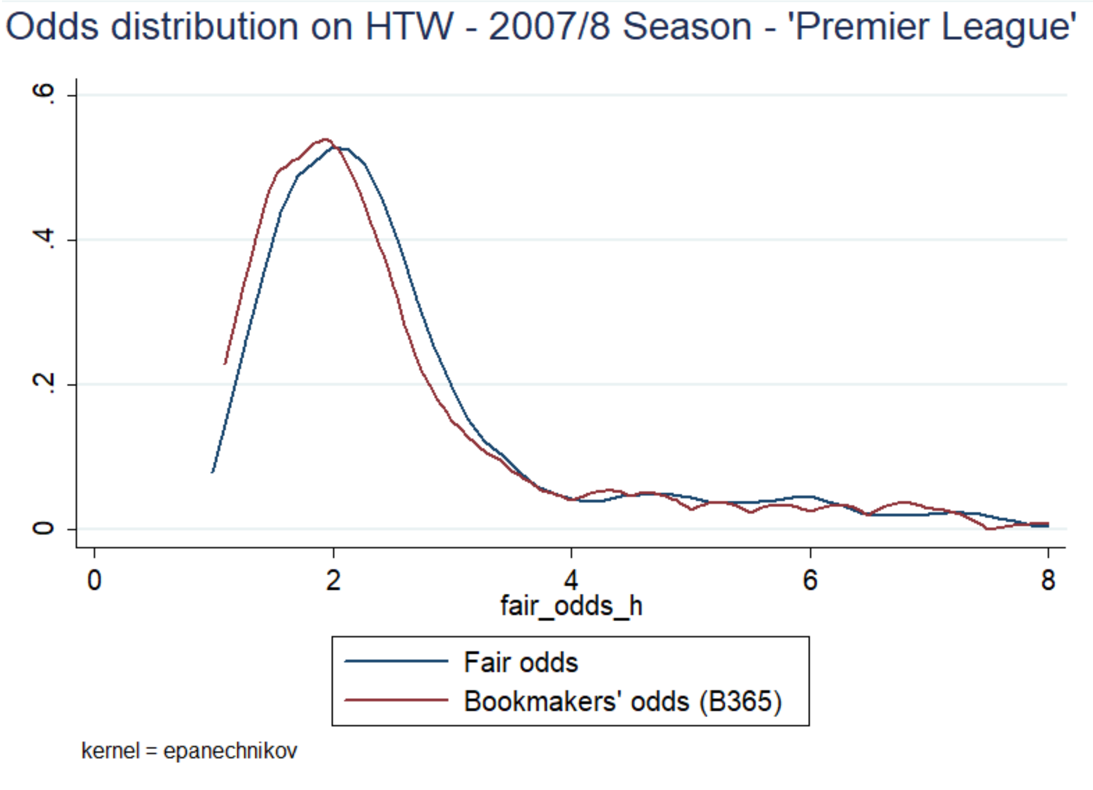
N=11,726

Figure A.13: Development of the Vigorish from 2007/8 to 2017/8 season, by League



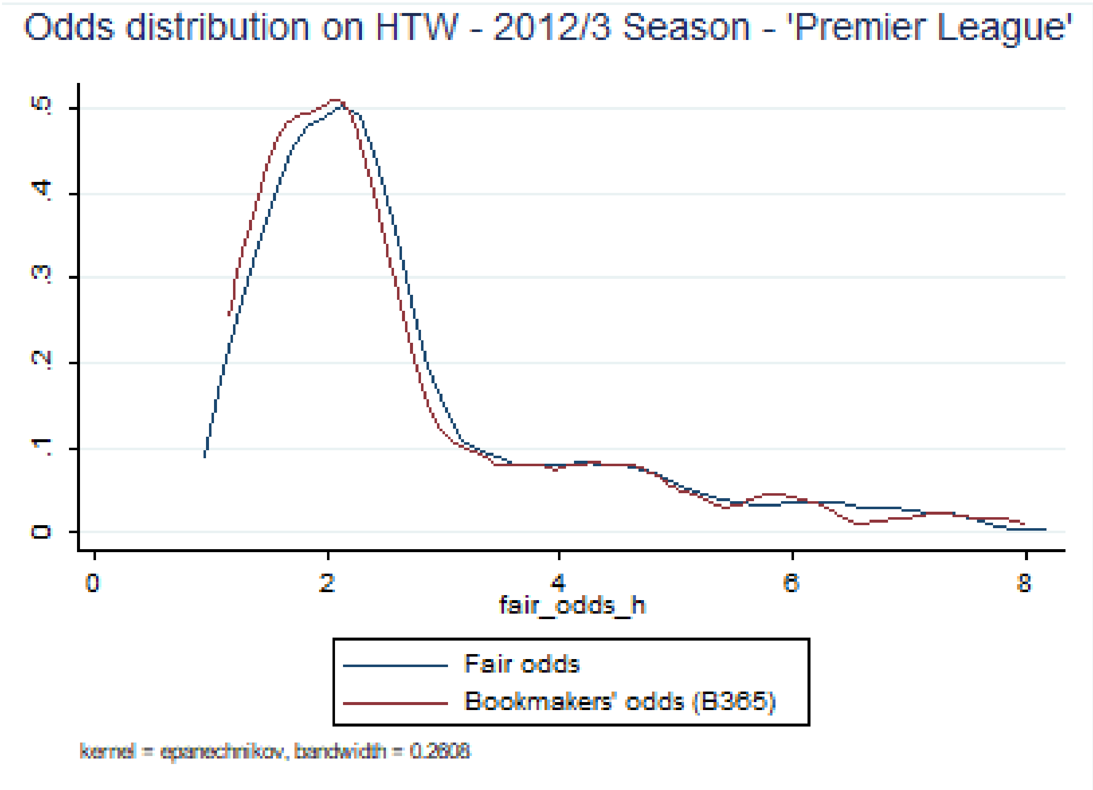
N=11,697

Figure A.14: Kernel Density function on Odds distribution on HTW - 2007/8 Season - 'Premier League'



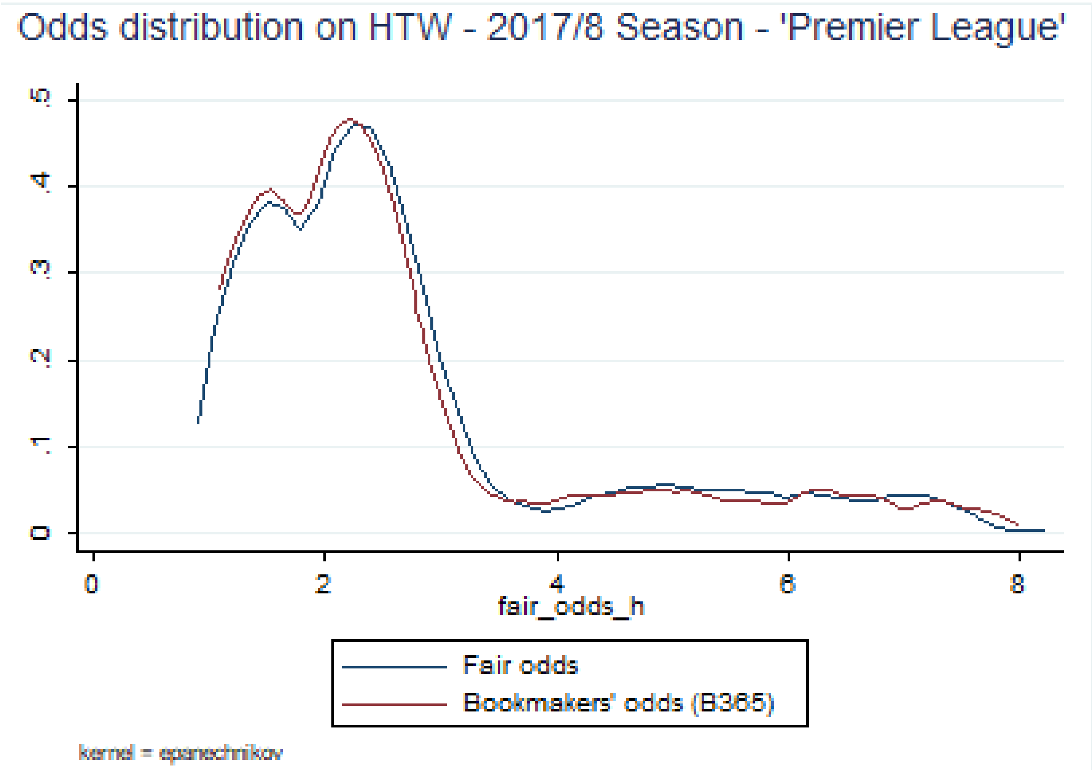
N=370 (For each Fair odds and Bookmakers' odds (B365))

Figure 15: Kernel Density function on Odds distribution on HTW - 2012/3 Season - 'Premier League'



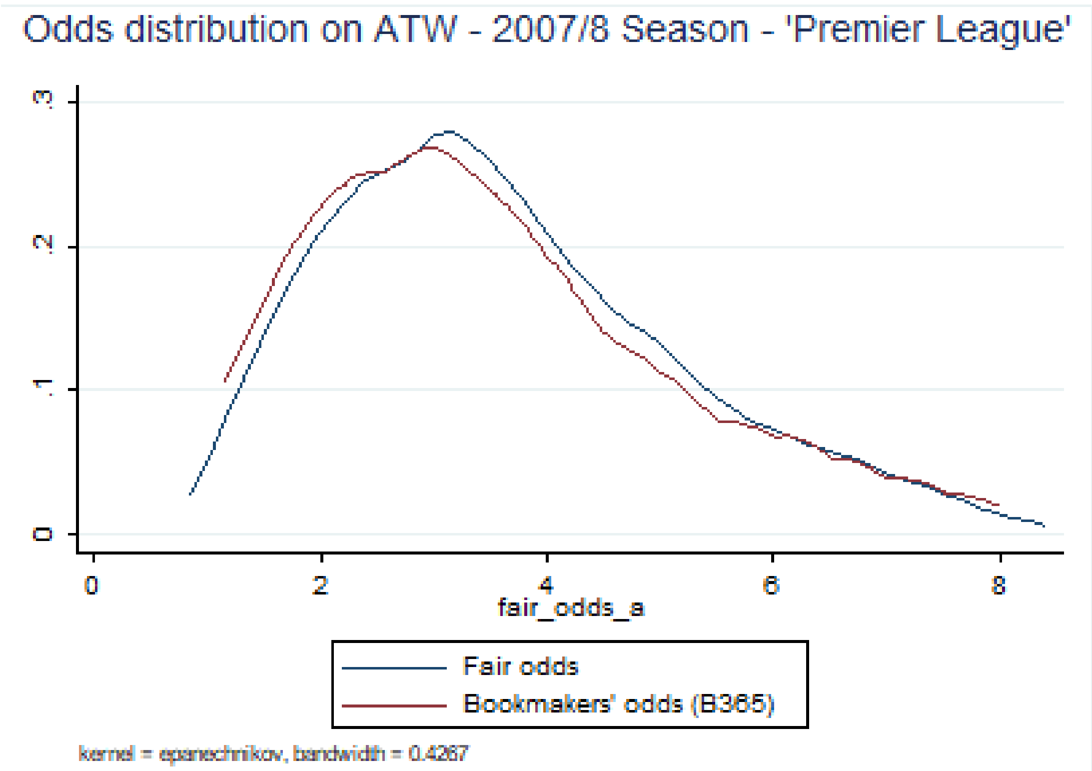
N=378 (Fair odds) and 380 (B365)

Figure 16: Kernel Density function on Odds distribution on HTW - 2017/8 Season - 'Premier League'



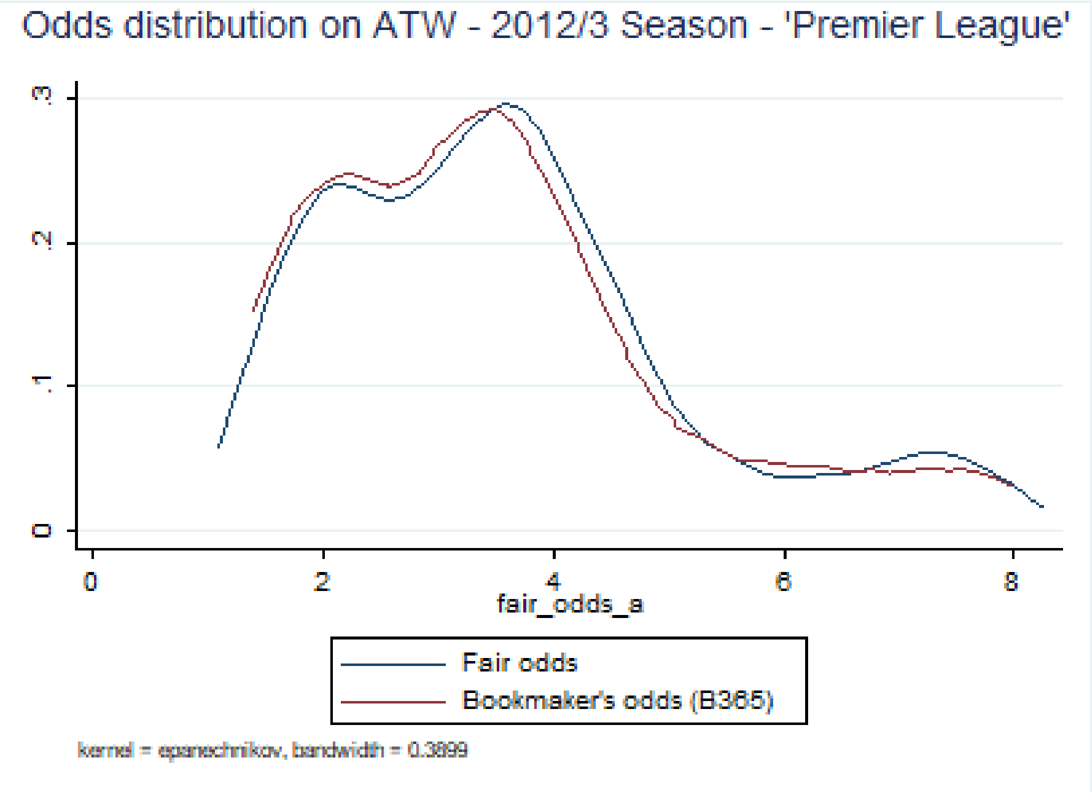
N=357 (for each Fair odds and Bookmakers' odds (B365))

Figure 17: Kernel Density function on ATW - 2007/8 Season - 'Premier League'



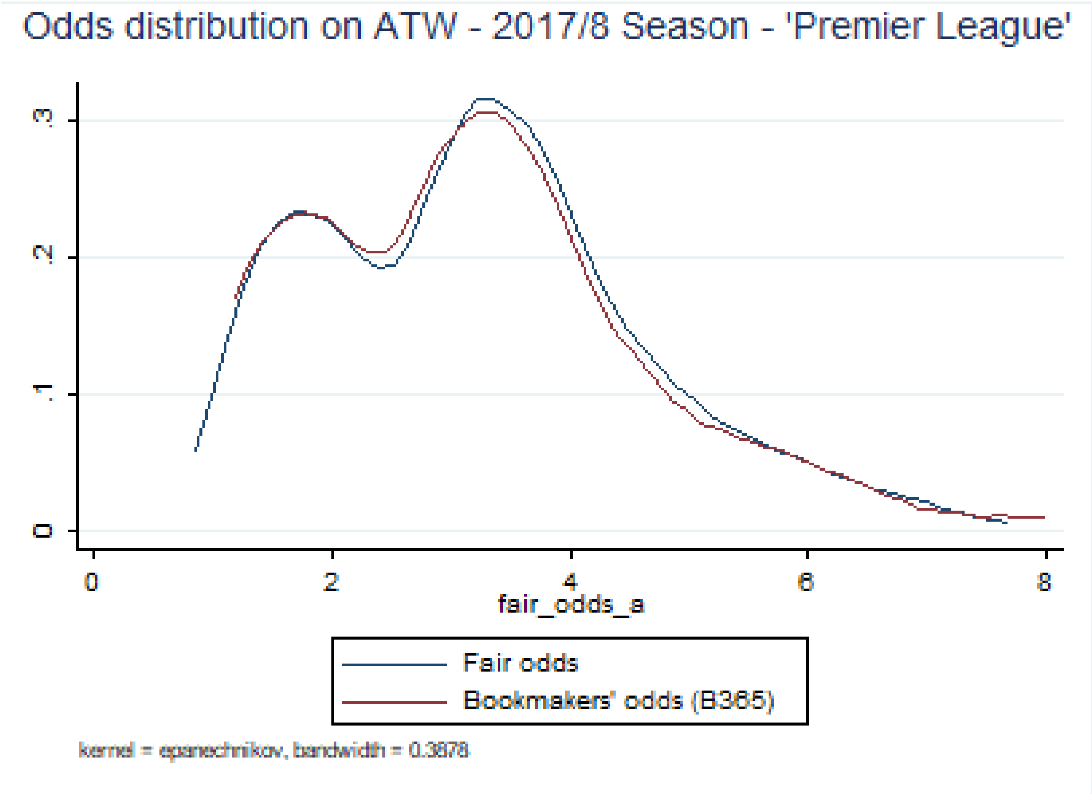
N=323 (Fair odds) and 326 (B365)

Figure 18: Kernel Density function on ATW - 2012/3 Season - 'Premier League'



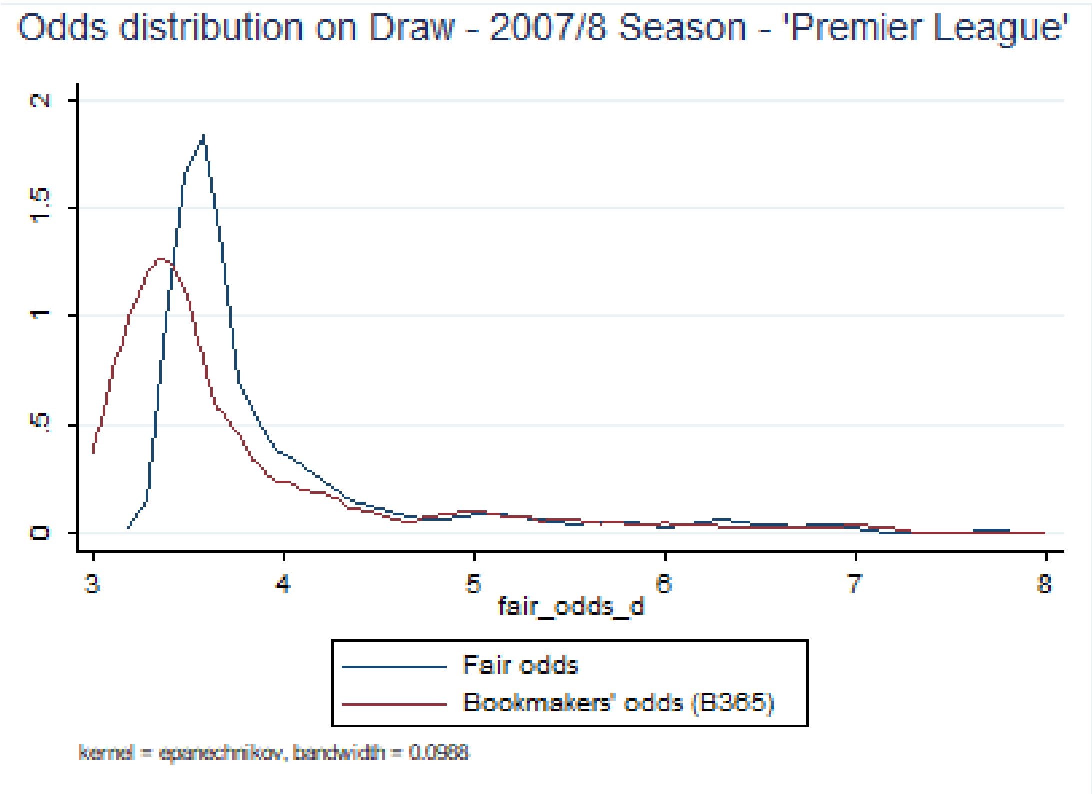
N=321 (Fair odds) and 323 (B365)

Figure 19: Kernel Density function on ATW - 2017/8 Season - 'Premier League'



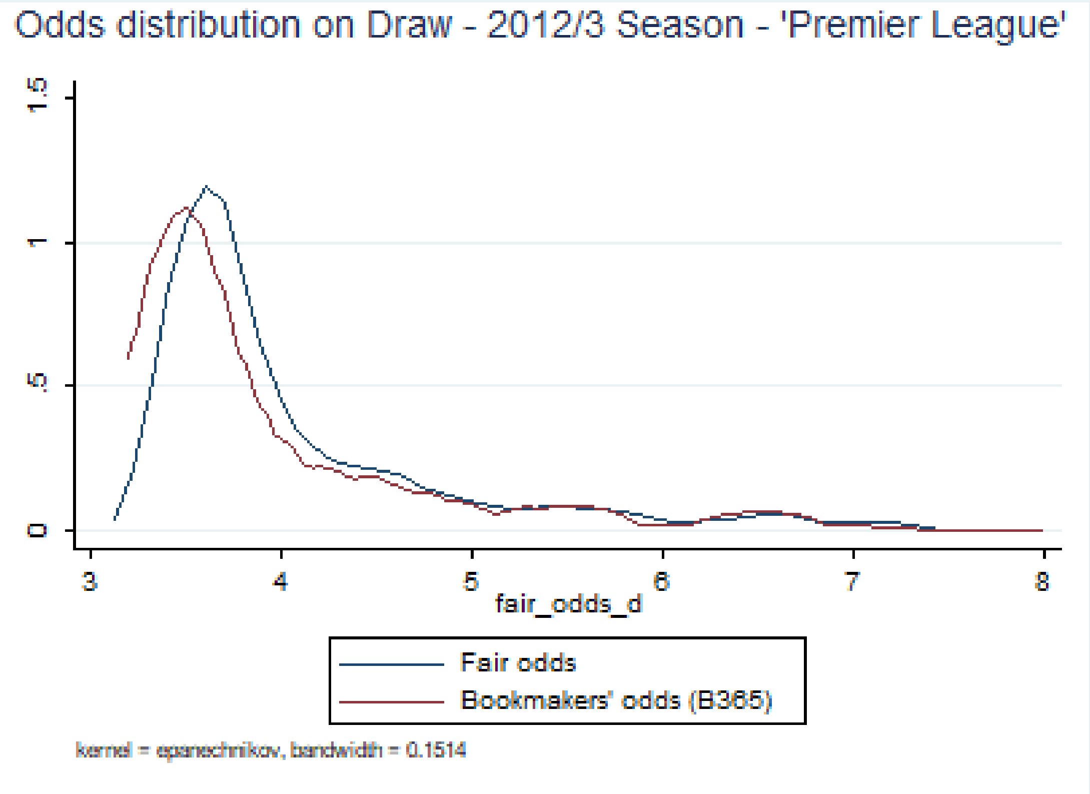
N=301 (Fair odds) and 303 (B365)

Figure 20: Kernel Density function on Draw - 2007/8 Season - 'Premier League'



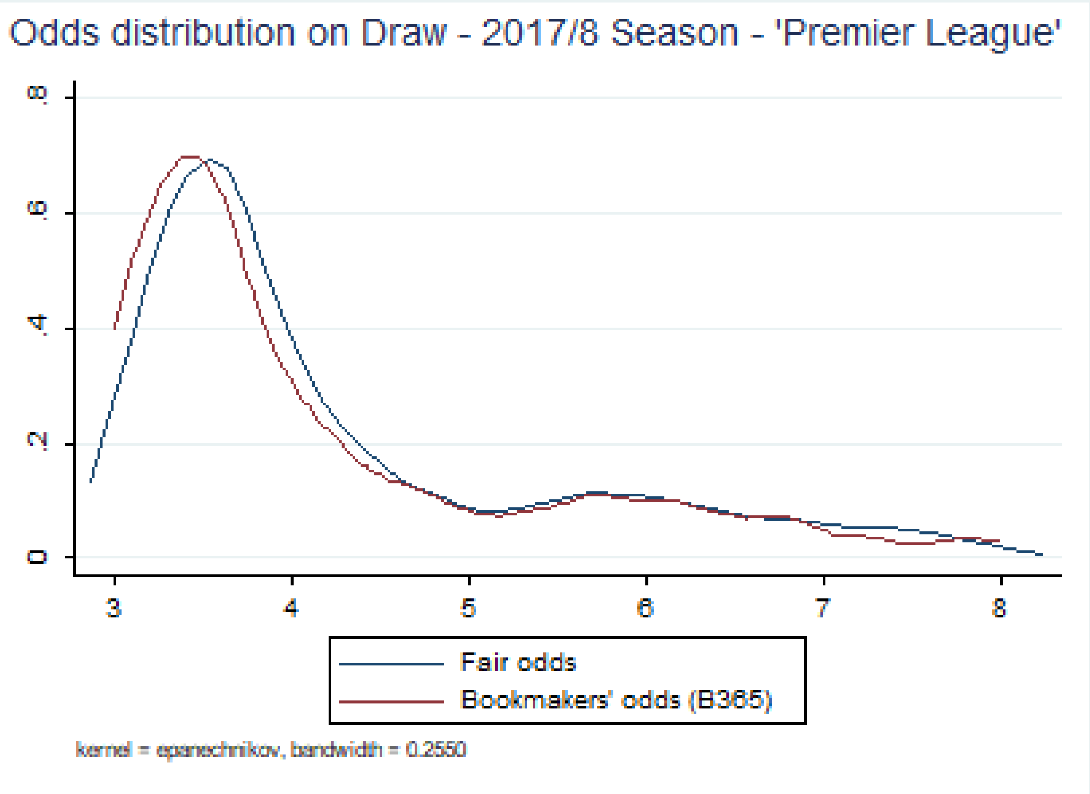
N=379 (Fair odds) and 378 (B365)

Figure 21: Kernel Density function on Draw - 2012/3 Season - 'Premier League'



N=377 (Fair odds) and 379 (B365)

Figure 22: Kernel Density function on Draw - 2017/8 Season - 'Premier League'



N=364 (Fair odds) 363 (B365)

Table A.1: Comparison of results on Betting Strategies

Betting strategies	Conditions	Number of bets		Return flow
Betting strategy 1	**1. Raw bookmaker outcome probability - Estimated real outcome probability < 0**	**120**	**Return based on average odds**	**Return based on maximal odds**
	2. Outcome probability of event xi >= 55%			
German League		**53**		
English League		**36**		
Spanish League		**31**		
Home Team-win		**74**	**12.57%**	**19.60%**
Home Team-win		29	31.94%	40.06%
Home Team-win		22	-14.27%	-9.23%
Home Team-win		23	13.84%	21.39%
Away Team-win		**46**	**4.74%**	**11.15%**
Away Team-win		24	20.90%	28.88%
Away Team-win		14	-26.71%	-22.85%
Away Team-win		8	11.24%	17.50%
Betting strategy 2	**1. Overround <=0.01%**	**780**		
	2. Outcome probability of event xi >= 65%			
German League		**182**		
English League		**313**		
Spanish League		**278**		
Home Team-win		**583**	**0.02%**	**2.67%**
Home Team-win		118	-5.05%	-2.37%
Home Team-win		259	0.85%	2.37%
Home Team-win		206	1.86%	5.95%
Away Team-win		**190**	**-5.75%**	**-0.52%**
Away Team-win		64	3.28%	3.79%
Away Team-win		54	-5.49%	14.24%
Away Team-win		72	-14.06%	-15.41%
Draw		***7***		*-44.29%*

N=893