UNIVERSITY OF SOUTHAMPTON
FACULTY OF BUSINESS AND LAW
SCHOOL OF MANAGEMENT

WEAK FORM EFFICIENCY AND PRICING DYNAMICS IN A COMPETITIVE GLOBALISED MARKET SETTING

Anastasios Oikonomidis

Thesis for the degree of Doctor of Philosophy (Management)

September 2013

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF BUSINESS AND LAW

Management

Doctor of Philosophy

WEAK FORM EFFICIENCY AND PRICING DYNAMICS IN A COMPETITIVE GLOBALISEDMARKET SETTING

by Anastasios Oikonomidis

Employing data from the football betting market, we explore the impact of institutional structure on price-setting in speculative markets and the extent to which the biases induced by such factors might cause prices to deviate from fundamental values. In Chapter 1, we review the literature on football betting markets with regards to the Efficient Market Hypothesis (EMH) and find that despite sporadic evidence of pricing anomalies, more consistent and persistent evidence of exploitable betting opportunities is required for the EMH to be rejected.

In Chapter 2, we investigate the favourite-longshot bias in the bookmaker market and find that the bias is persistent over a long period of time and related to the parity of the league. A traditional price-setting bookmaker is willing to pay a premium to stimulate overall bettor demand in a competitive market, by setting generous odds on favourites to attract customers. In leagues with more parity among teams, the apparent bias is reduced as the market requires less intervention on the part of the bookmaker.

Heterogeneity in bookmaker operations and price-setting is the focus of Chapter 3. We categorize bookmakers as either position-takers or book-balancers. Position-takers operate a high-margin, low-turnover business model, rarely adjust their odds, and actively eschew informed traders. Book-balancers frequently change their prices, and operate under the alternative, high-turnover, low-margin strategy. Sophisticated traders are not restricted at book-balancing bookmakers, and as such, odds movements at position-takers lag converge in the direction of the odds at book-balancers. We conclude that the sophisticated traders aid the price discovery process at the book-balancing bookmakers, which leads the market to efficiency.

Finally, in Chapter 4 we examine instances of significant dispersion between the two types of bookmakers and show that the generation of positive returns is theoretically possible. However, closer investigation of these finding reveals that market-makers' odds are efficient predictors of event outcomes and therefore, the opportunity to generate profit is provided by biases in position-takers' pricing. Such biases could either be intentional for the purpose of attracting customers or the result of such bookmakers' odds lagging behind in reflecting upcoming information. In all cases, such bookmakers are very likely to pose restrictions successful traders and therefore, the exploitation of the documented anomaly is probably infeasible.

Table of Contents

List of Tables and Figures

Table

Figure

Declaration of Authorship

This PhD thesis is the result of work conducted wholly or mainly by Anastasios Oikonomidis whilst in registered candidature. None of the material presented has been submitted for another degree. Minor aspects of the thesis have been developed from work conducted jointly with other persons who have assisted in supervisory or advisory roles. Details of the contributions made by others are outlined below.

Chapter	***Co-Authors and Advisors***	***Contribution***
One	Professor Johnnie E.V. Johnson (Co-Author) School of Management, University of Southampton, UK.	Provided advice on methodology and feedback on drafts of paper
Two	Professor Johnnie E.V. Johnson (Co-Author) School of Management, University of Southampton, UK.	Provided advice on methodology and feedback on drafts of paper
	Professor Alistair Bruce (Co-Author) Business School, University of Nottingham, UK.	Provided advice on methodology and feedback on drafts of paper
Three	Professor Johnnie E.V. Johnson (Co-Author) School of Management, University of Southampton, UK.	Provided advice on methodology and feedback on drafts of paper
	Dr Andrew Grant (Co-Author) Business Shool, University of Sydney, UK.	Provided advice on the linkages between betting and financial markets
Four	Professor Johnnie E.V. Johnson (Co-Author) School of Management, University of Southampton, UK.	Provided advice on methodology and feedback on drafts of paper
	Dr Andrew Grant (Co-Author) Business Shool, University of Sydney, UK.	Provided advice on the linkages between betting and financial markets

Acknowledgements

First of all, I would like to thank Johnnie for his support and assistance throughout the (long) duration of the PhD period. Secondly, I need to thank Andrew, for going through the papers' results with me, for revealing valuable insights regarding the operations of financial and betting markets, for guiding me through the literature of financial markets and mainly, for suffering long skype calls, in order to provide me with his assistance. Thanks also to Professor Alistair Bruce for his significant contirbution in the second chapter of this thesis and to Dr Mike Smith for suggesting Veitch’s book to me. Moreover, I thank Professor William Benter for offering to me valuable insights concerning the real world of trading and the challenges faced by professional investors. I also thank the staff in the School of Management for their support. I thank *Monalisa* for assisting me in programming issues. Furthermore, I thank Natasha for her consultancy with regards to academic writing, but mainly for being herself. I also thank all of my friends for being around, my mother for her continuous support and my father for his encouragement and insistnce to get this job completed.

Introduction

Scope of the Thesis

This thesis aims to uncover the significance of specific institutional characteristics in the setting of prices in speculative markets and their impact on the degree to which prices reflect the fundamental value of assets. In particular, we investigate the objectives of price-setters and identify major structural differences across market operators. This is in contrast with existing literature, which generally assumes these markets are populated by a homogeneous population of market operators. We explore how these structural differences influence the dynamic flow of information across market prices and demonstrate how they can lead to potentially spurious pricing biases that could be misjudged as trading opportunities. In particular, we show that in markets where the price-setter takes positions against the population of price-layers, the expected value of a trading proposition is likely to be negatively related to its long term exploitability. Such restrictions regarding the exploitation of trading opportunities potentially deter the convergence of asset prices to their true value, effectively generating artificial mispricing. As a result, we suggest that the testing of the Efficient Market Hypothesis (EMH) in such cases is likely to lead to misleading conclusions and potentially risky inferences concerning traders' behaviour.

The Efficient Market Hypothesis

The EMH implies that market prices fully reflect all available information (Fama, 1970). As a result, asset prices should converge to their fundamental values, eliminating the opportunity for systematic, long-term profitable trading. Temporary mispricing is anticipated to be quickly removed by the actions of informed traders, leading prices towards their true underlying level. Hence, in an efficient market, prices constitute an accurate estimate of the assets' fair value, directing capital to be efficiently allocated across diverse investments (Bushman, et al., 2011).

The EMH has been the subject of extensive research in the financial market literature. Several studies have attempted to reject it, providing evidence of apparent anomalies in market prices. However, Fama (1998),reviewing the literature that documents stocks' over- and under-reaction to information, shows that combining the evidence from studies which individually reject the EMH, leads to the conclusion that the EMH still holds overall. In that sense, the evidence that each individual study presents against the EMH seems to be the result of random variation rather than evidence of systematic bias that can lead to the formulation of an alternative robust theory to replace the EMH. Hence, in order for an apparent mispricing to be characterized as market inefficiency, it has to be consistently expressed across different samples, be persistent through time and not be subject to changes in the methodology employed. It is important that these conditions are respected because the implications of rejecting the EMH based on weak evidence are that misleading inferences regarding investors' behaviour may be developed which might, in turn, adversely impact decisions regarding the regulation of markets.

Exploring the EMH in Betting Markets

The investigation of the EMH has been extended to betting markets. Betting markets offer a valuable laboratory for testing hypotheses regarding financial markets, being simpler in form, issues relating to prices can be more clearly viewed in betting markets than in more complex financial markets (Sauer, 1998). As Durham et al. (2005) note, betting markets are similar to financial markets with regards to liquidity and the availability of information, where diverse decentralized market makers offer price quotations on the same event (Marshall, 2009)and the return of each investment depends on an unambiguous result (Law and Peel, 2002). Moreover, in betting, unlike stock-markets, each asset has a well-defined termination point at which its value becomes certain (Thaler and Ziemba, 1988). (For a detailed description of the advantages offered by betting markets in investigating the EMH, see Sung, et al., 2012).

It has to be noted though that the structure and limitations of these markets need to be carefully considered so that inferences regarding investors' behaviour are valid. More specifically, in demand driven markets, such as the parimutuel market and betting exchanges the economic significance of prices is subject to the volume traded in them. Hence, in such markets, perceived mispricing in events that offer low liquidity may simply exist due to the expected profit not being sufficient to attract the interest of informed traders. It could therefore be claimed that such theoretical mispricing does not constitute evidence against the EMH. On the other hand, in supply driven markets, the bookmaker acts as an intermediate operator applying their policy, subject to their own objectives and therefore, the prices they set may not directly reflect demand. As a result, market prices do not necessarily reflect the expectations of the public as it is often assumed by studies that analyse biases in investors' behaviour using bookmaker data.

The betting market literature employs the definition of the EMH from the broader financial market literature. Consequently, in an analogous fashion to the definition of efficient prices in wider financial markets, efficient odds (prices) in betting markets are regarded as those that accurately reflect the corresponding probabilities of the relevant event outcomes[1]. However, many betting market studies analyse data from the bookmaker market where, unlike a demand-driven market (as in most financial markets), the implementation of a trading strategy depends on the trades being consistently accepted by the bookmaker. As a result, it could be argued, that in order for the EMH to be rejected when tested against bookmakers' odds, it has to be shown that positive returns can be generated against such odds. In addition, it must be shown that any apparent mispricing is truly exploitable (i.e. that the unrestricted implementation of a profitable strategy in a substantially liquid market is

[1] Even if the EMH is tested by examining whether profitable betting strategies exist, the determinant of whether such a strategy is profitable is the deviation of odds from the unknown *true* odds (based on the outcomes objective probability).

feasible).Evidence regarding apparent inefficiencies exhibited by bookmakers' odds has been presented in the literature, but, importantly, implementation related issues have been overlooked. However, we argue that there is likely to be a *cause and effect relationship* between the bookmaker's intention to restrict profitable traders and the deviation of its odds-implied estimates from the *true* probabilities. Therefore, we suggest that in such cases, the apparent mispricing should not be considered as an inefficiency, as it is likely that this would not exist without the bookmakers' intention to restrict those who are likely to exploit it. Hence, as in the case of the financial market literature described above, it is important to avoid rejecting the EMH on weak evidence, such as the existence of non-exploitable mispricing. For that purpose, we refer to pricing biases as *theoretical* inefficiencies, which can only be regarded as *true* inefficiencies if they are shown to be exploitable[2].

Overall Contribution

The most popular form of football betting takes place in bookmaker betting markets and these are the most commonly investigated in the football betting literature. A range of studies have examined data related to bookmaker odds and most of these studies make several assumptions regarding the bookmakers' operations. However, none of these studies has scrutinized bookmakers' price-setting mechanisms.

This study analyses unique data sets, where odds offered by diverse bookmakers for football games played in the major European leagues, are collected simultaneously. This collection method ensures that the odds offered coexisted. In addition, the data is collected at several time-points in the market for each game, thus forming a panel data-set of time-varying odds. This data enables the modelling of cross-bookmaker interactions through time for each event. This analysis reveals, for

[2] Of course a strategy can only be proved to be exploitable when it is implemented in the real world. Consequently, it is difficult for an academic study to provide relevant hard evidence. However, the characteristics of the counterparties against which the bets are placed in a theoretically profitable strategy can be enlightening.

the first time, heterogeneity in the bookmaker market with two sets of bookmakers (book-balancers and position-takers) exhibiting structural differences in their operational model. The diverse objectives of the two groups are identified and these are associated with their different pricing policies. In addition, it is shown that the observed mispricing in the market is dependent on the bookmaker's business model, which determines its approach towards informed traders. Consequently, the conclusion to emerge from the analyses conducted in the thesis is that there is insufficient evidence of biases in football betting markets which are likely to be exploitable. As a result, we would argue that viewing the evidence presented in this thesis as a whole, we cannot reject the EMH on the basis of the joint hypothesis of mispricing and exploitability. My overall conclusion, therefore, is that previous studies may have overestimated the significance of pricing biases, as these mainly arise from odds offered by bookmakers which are likely to impose restrictions on bettors who may attempt to exploit them.

Structure of the Thesis

In order to achieve the main aim of this thesis, namely, to identify the key structural characteristics in the operation of diverse bookmakers and examine their impact on price-setting and the efficiency of the market, the thesis is structured as follows: In Chapter 1 the football betting market literature is reviewed, in order to assess the documented evidence regarding price anomalies. Chapter 1 confirms that there is a significant amount of evidence that apparent pricing anomalies exist in football bookmaker markets. Chapter 2 explores the consistency and persistency of a persistent pricing anomaly, namely the favourite-longshot bias, in football fixed-odds betting and investigates its distribution across different European leagues. Subsequently, to examine more thoroughly the source of apparent mispricing in football betting markets, the structural characteristics of diverse bookmakers are explored. In that context, Chapter 3 identifies heterogeneity amongst bookmakers in the football betting market and examines how the different types of bookmakers set their prices and how they interact in a competitive, globalized setting. It is shown that the fundamental differences in the operation of bookmakers often lead to significant price-dispersion in the market. Hence, exploring this price variation is likely to lead to meaningful conclusions regarding the relationship between the bookmakers' structural characteristics and the generation of apparent pricing anomalies. Thus, Chapter 4 presents an analysis of price dispersion across diverse bookmakers and reveals how the observed theoretical inefficiencies in the betting market are mainly associated with the structural characteristics of one type of bookmaker. In the final chapter of this thesis the conclusions drawn from all evidence provided in this study are discussed. This discussion focuses on the sources of the observed instances of mispricing and their association with bookmakers' structural attributes. The inferences of these findings in the context of the EMH are discussed, directions of valuable future research are identified and the potential implications of the findings on broader financial markets are examined.

The results and conclusions from each chapter can be summarized as follows:

Chapter 1

The literature concerning tests of the EMH in football betting markets is reviewed. Following Fama's (1970) definition, information is categorized as weak, semi-strong and strong. It is shown that adopting the traditional concept of market efficiency, several apparent inefficiencies have been documented. However, it is argued that stronger evidence is required regarding the possibility of persistent and consistent long-term returns which are *exploitable* for the evidence to enable the EMH, in the context of football betting markets, to be rejected.

Chapter 2

Using a large dataset of bookmaker[3] odds concerning football games played over a decade across different European leagues, it is shown that consistent deviation of odds from the underlying *true* odds is evident, resulting in significant favourite-longshot bias in bookmakers' odds-implied estimates. This theoretical anomaly is persistent through time, indicating that it is a structural effect of this particular market, rather than the result of erroneous estimation. It is found that the bias is more pronounced in certain leagues and that the level of competition within a league is the main determinant of this heterogeneity. It is argued that this may arise from bookmakers' marketing related objectives, whereby they offer attractive odds for favourites in order to acquire and retain customers, in the face of competition from other bookmakers. It has to be noted that in the first two chapters the word "bookmaker" refers exclusively to traditional bookmakers, classified as position-takers in chapter 3 and that the term "inefficiency" is used conventionally. However, the term "theoretical" or "apparent inefficiency" would be more compatible with the general ideas outlined in this thesis.

Chapter 3

It is discovered that significant heterogeneity exists across different bookmakers who offer odds on football matches. In particular, it is suggested that bookmakers can be classified into two main groups: position-takers and book-balancers. The differences in their characteristics and the objectives of their operations are examined. For the first time in the football betting literature, bookmaker odds are analysed as a time varying panel. To accomplish this, a unique data set of matched

[3]This chapter analyses data from bookmakers, which are classified as position-taking in Chapter 3. The period covered by the sample is crucial for the purpose of this analysis, in order to show that the bias is persistent through time. However, such data could not be matched with data from book-balancing bookmakers, as these did not exist throughout the earlier years of the sample. However, as shown in Chapter 4, in an analysis of a smaller sample of data, the favourite-lonshot bias and apparent inefficient pricing in general, are largely a features of position-taking (cf. book-balancing) bookmakers.

bookmakers’ odds offered at different points in time, for each event is collected. Odds changes for each potential outcome are modelled, employing cross-sectional time-series analysis, in order to explore how the two simultaneously operating types of bookmaker interact. As a result, we demonstrate how structural differences between these different types of bookmaker influence their price-setting. These structural differences are shown to lead to information being transmitted from sophisticated traders directly to book-balancers and indirectly to position-takers, increasing the efficiency of market prices.

Chapter 4

A unique data-set comprising odds concerning outcomes of football events, offered by a range of major bookmakers, is analysed to investigate price variation between different types of bookmaker. The odds available close to the games’ kick-off were collected. This is the period in the market when liquidities maximized, ensuring that any apparent inefficiencies are more likely to be exploitable. Unlike previous studies where the data analysed do not guarantee the simultaneous existence of the various bookmakers’ offers, a multi-threading programming technique is employed, in order to ensure that the various odds coexisted. A multithreading program manages multiple requests simultaneously, without the need of multiple copies of the program running. That way we request the odds data concurrently by different bookmakers. Employing linear optimization, we identify a high frequency of theoretical arbitrage opportunities. However, it is shown that the systematic exploitation of such mispricing instances is unlikely, due to the structural characteristics of position-taking bookmakers (who are shown to be responsible for the generation of the apparent anomalies).

References

Bushman, R.M., Piotroski, J. D., and Smith, A. D. (2011).Capital Allocation and Timely Accounting Recognition of Economic Losses. *Journal of Business, Finance & Accounting*, 38, 1-33.

Durham, G., Hertzel, M., and Martin, J. (2005).The market impact of trends and sequences in performance, new evidence.*Journalof Finance*, 60(5), 2551-2569.

Fama, E. (1970). Efficient Capital Markets: A Review of Theory and Empirical Work. *Journal of Finance*, 25, 383-417.

Fama, E. (1998). Market efficiency, long-term returns, and behavioural finance.*Journal of Financial Economics*, 49, 283-306.

Law, D., and Peel, D. A. (2002).Insider Trading, Herding Behavior and Market Plungers in theBritish Horse-race Betting Market.*Economica*, 69, 327-338.

Marshall, B. R. (2009). How quickly is temporary market inefficiency removed?*The Quarterly Review of Economics and Finance*, 49, 917-930.

Sauer, R. D. (1998). The Economics of Wagering Markets.*Journal of Economic Literature*, 36, 2021-2064.

Sung, M-C, Johnson, J.E.V., and Peirson, J. (2012).Discovering a Profitable Trading Strategyin an Apparently Efficient Market: Exploiting the Actions of Less InformedTraders in Speculative Markets.*Journal of Business Finance & Accounting*, 39, 1131-1159.

Thaler, R., and Ziemba, W. (1988).Pari-mutuel Betting Markets: Racetracks and Lotteries. *Journal of Economic Perspectives*, 2, 161-174.

Chapter 1

Who Can Beat the Odds? The Case of Football Betting Reviewed.

1.1. Historical Development of Football Betting Markets

Sports betting has been an ’integral part of working class structure‘ in the UK since the beginning of the twentieth century (Jones, et al., 2000). At the end of the 1960s, nearly a decade after the formalization of sports betting markets (betting shops were legalized in the UK in the early 1960s), nearly 16,000 betting shops operated in the UK, though concentration resulted in this declining to about 8,800 in 1998 (Jones, et al., 2000). However, betting turnover has been increasing steadily, leading the four major British bookmaking firms to report turnover of £10 billion in aggregate, in 2002 (Levitt, 2004). The Global Betting and Gaming Consultants (2001) indicated that in 1998 about 4 million adults were betting weekly on sports in the UK (Forrest and Simmons, 2003). More specifically, football is, according to the Mintel Leisure Intelligence Report (2001), the fastest-growing form of gambling in the UK and the Gambling Review Report (DCMS, 2001, as cited in Forrest and Simmons, 2003) indicates that most of the sports betting activity concerns football.

Obviously, the popularity of sports betting is not constrained to the UK. Worldwide, betting on sports is extremely popular and football betting has the lion’s share in most countries. The National Gambling Impact Study Commission (1999) estimates that wagering on sporting events in the US approaches $380 billion annually (Levitt, 2004). In 2007/2008, The Jockey Club was the greatest Honk Kong tax payer contributing about US$1.690 billion, approximately 6.5% of all taxes collected by the Hong Kong Government (Wong, et al., 2009). Similar data are reported by So and Kwok (2007) for the season 2005/2006. They also show that football betting turnover in the 2009/10 season was 31.27 billion Hong Kong dollars (about £2.67 billion). Moreover, according to Forrest and Simmons (2003), football industry sources suggest that Far Eastern football betting turnover is about US $1 billion per weekend during the course of the football season, and that about half of this is bet on English Premier League matches.

During the last decade, due to the spread of the internet, internet betting has developed rapidly. According to Jones, Clarke-Hill and Hiller (2000), Sportingbet, the then small, company, established by an independent bookmaker, was the first company to enter the internet sports betting market. Many established companies followed and new ones were formed, creating a very competitive and dynamic market setting. Malaric, Katic and Sabolic (2008) identify 600 different sports betting web sites operating worldwide, representing a $16.6 billion market; which is predicted to rise. The number of internet gambling companies (including casino operations), has been reported to exceed 1,800 (Forrest and Simmons, 2003) and the SportBusiness Group (2001) predict that the e-gaming industry will exceed $100 billion by 2015. China, Hong Kong, Singapore, and Sweden are named as countries which offer high prospects for the growth (Forrest and Simmons, 2003).

Another key feature associated with the evolution of sports betting markets is the establishment of betting exchanges. The function of betting exchanges resembles that of the honest brokers in the 18th century. Hence, unlike the bookmakers, the betting exchange company acts as an intermediary that matches opposing bets between bettors, holds the funds until the outcome is decided and pays the winner, after deducting a small commission. This commission, being risk-free, allows the company to set it at a lower level compared to the bookmakers' usual margin. Sporting Exchange Limited is the major betting exchange company, with the trading name Betfair. It was founded in 1999 and launched its website in June 2000 (Jones, et al., 2004). On a daily basis, Betfair matches about 500,000 bets and had reported turnover exceeding £50 million per week in 2003 (Jones, Hillier, Turner and Comfort, 2004). The company has over two million registered users (Croxson and Reade, 2013). Even though several other betting exchange companies, such as Sporting Options, Betdaq and GGBet have entered the market (Jones, et al., 2004), Betfair accounts for 90% of all exchange-based betting activity worldwide (Croxson and Reade, 2013).

Football betting also takes place in spread-betting, prediction and pari-mutuel markets. The development of betting exchanges and the spread of internet betting has intensified competition, forcing bookmakers to decrease their margins (known as the over-round) significantly on football bets in recent years (Forrest and Simmons, 2003 and Oikonomidis and Johnson, 2008). In addition, recent developments in legislation have been beneficial to bettors (e.g., regarding UK betting tax see Paton, et al., 2002). Thus, overall, modern football betting markets are associated with friendly legislation in many places in the world, high volume, intense competition and very low transaction costs. Consequently, it could be claimed that these markets constitute the ideal setting for the exploitation of profitable opportunities. Testing the efficiency of the football betting market with respect to different sources of information therefore presents an interesting opportunity for researchers. In the following sections, the extent to which this has been achieved will be reviewed. This will be structured around Fama's (1970) categorization of market efficiency; consequently, literature regarding "weak", "semi-strong" and "strong-form" efficiency of football betting markets will be examined in turn.

1.2. Weak Form Efficiency

1.2.1. Introduction

According to Fama (1970), a market is weak form efficient if current prices reflect all information arising from past prices. Consequently, a betting market is weak form efficient if abnormal returns cannot be made using any kind of information related to market odds. A strategy is considered to yield abnormal returns, when it is shown to be consistently profitable over a sufficiently long period of time and over a large number of transactions, sufficient to minimise the probability of randomly achieving similar returns. Consequently, the replication of such a strategy (ceteris paribus) is very likely to lead investors to achieve profits that significantly overcome transaction costs and is subject to minimal risks (as long as a substantial number of trades is employed). A significant number of papers have analysed football betting markets with respect to odds information. Thus researchers have explored the value of odds in predicting football events, the existence of systematic odds-related biases, as well as the degree of variation of odds between different market operators. In the following section, these studies are reviewed.

1.2.2. Odds as Predictors

In a football betting market, odds reflect the estimations of market makers regarding the probability of competing outcomes and some papers have explored the accuracy of such odds. For example, Leitner, Zeileis and Hornik (2008) analyse odds quoted by 45 bookmakers (concerning the European Championship competition 2008) to explore how successfully these predict match outcomes. They employ mixed effects regression (group and bookmaker specific fixed effects and team specific random effects) to model the true odds of each team winning the competition, based on the market odds. They compute pairwise winning probabilities and simulate the tournament, concluding that the estimated odds-based probabilities are highly correlated with the actual outcomes. Leitner, Zeileis and Hornik (2009) apply a similar methodology to obtain winning probabilities for teams in the Champions' league (season 2008/2009), assigning the highest probabilities to Chelsea, Manchester United, Inter-Milan and Barcelona. Three of these teams reached the semi-finals, indicating that predictions arising from the odds based model were good but not perfect. Finally, data from the prediction market (STOCCER championship market) has also been analysed (Luckner and Weinhardt, 2008) and it was found that estimations arising from 'play money' were no less accurate than those arising from betting odds. Similar evidence is provided by Servan-Schreiber et al. (2004) using data from the American National Football League (NFL).

Some researchers have explored whether variation exists in the accuracy of odds-based predictions of football events. For example, Strumbelj and Sikonja (2010) examined odds of ten bookmakers related to 10,699 matches from six major European football leagues. They found that the accuracy of odds in predicting outcomes increased through time but that variations existed in the forecasting ability of different bookmakers and in the cross-league accuracy of the odds in predicting results. However, this later finding could simply be due to cross-league differences in competitiveness. Thus, in more competitive leagues, the outcome of a football event is likely, on average, to be more random (c.f. a less competitive league).

The results of the papers discussed above suggest that football odds exhibit forecasting power; which is not unexpected given the availability of public information regarding the sport and the size of its betting market. However, in order to understand whether odds efficiently incorporate publicly available information and are therefore set at a level to prevent abnormal returns being made, an “accuracy benchmark” is required. This benchmark can be provided by quantitative forecasting models that utilize publicly available, including fundamental information regarding football events. However, such tests move the investigation to a “semi strong” level and, therefore, this subject is reviewed in the corresponding section, later in this chapter.

1.2.3. Odds Biases

Undoubtedly, the most popular object of research is the famous favourite longshot bias (FLB). Initially, the bias was observed in horseracing, but subsequent research has documented its existence in a variety of sports betting markets (for a survey of studies see Sauer (1998) and Vaughan Williams (2005). A FLB exists when the favourites’ winning probability, as implied by the odds, is on average, relatively underestimated compared to their unknown true probability and the probability of longshots is relatively overestimated. However, the reverse phenomenon has also been documented (e.g., Woodland and Woodland, 1994) and is usually referred to as ’reverse’ or ’negative’ FLB. The literature concerning the FLB in the football betting market is now reviewed.

Pope and Peel (1988) investigated the UK market (data derived from the season 1981/1982) and concluded that favourites seem be more profitable, compared to longshots. Cain Law and Peel (2000) analysed a dataset of 2,855 games played in the UK during the 1991/1992 season and found some evidence that market odds underestimate the winning probability of heavy favourites (including the probability of frequent (low) exact scores). Similar evidence was presented by Malaric, Katic and Sabolic (2008), who explored a dataset of 12,218 games played in 10 European leagues in the period 1999-2002. Dechamps (2008) also documented FLB associated with several European leagues in the 2005/2006 season, with more pronounced effects in second division leagues. Finally, Vlastakis, Dotsis and Markellos (2009) observed that the market underestimates merely the winning probabilities of favourites playing away from home.

The ‘draw’ outcome is more frequent in football compared to other sports and therefore, it is interesting to investigate whether its existence influences the FLB in any way. Deschamps and Gergaud (2007) explore a dataset of 8,377 football matches played between seasons 2002/2003 and 2005/2006 in English leagues and observe ‘positive’ FLB, concerning the odds of the home and away teams. However, for the odds on the draw outcome a reverse FLB is identified. Additionally, it is found that the probability for a draw is, in general, underestimated by market odds. Consequently, betting on ‘draw’ yields superior returns than betting on the home or the away team winning. The authors note that the bias is mainly observed in English leagues mainly during the year period 2004-2006, when there was an increase in the frequency of draws that is not accounted for by market prices. However, a wider sample (i.e. including a larger number of leagues over a longer time period) would be desirable in order to conclude that the bias is not the result of random variation in the competitive balance of these leagues from season to season, which can be observed ex-ante, without necessarily being predictable. Moreover, it would be interesting to test the existence of the draw bias on a more recent sample of games, since in recent years the level of transaction costs has decreased and therefore, such a bias could be sufficient to generate betting profit in today’s betting market if it has been sustained.

Variation in the magnitude of FLB has also been associated with the level of transaction costs and with league-specific characteristics, including competitiveness. Paton and Vaughan Williams (1998) found that the fixed-odds football betting market, where transaction costs are higher, exhibits higher FLB compared to the spread betting market (where transaction costs are lower). Oikonomidis and Johnson (2007) suggested that if bookmakers fail to fully account for cross-league differences in competitiveness, heterogeneity in the magnitude of the FLB would be expected. This hypothesis was confirmed by their analysis of a sample of over 56,000 football matches played in 22 European leagues over the last decade; the level of league competitiveness almost completely determined the degree of FLB in each league, with relatively competitive leagues exhibiting significantly higher bias.

Conclusively, it can be stated that significant, 'positive' FLB exists in the football betting market. The fact that it has been documented across different samples and is shown to be persistent across years, points to the fact that it is a structural idiosyncrasy of the market.

1.2.4. Market Variation

As indicated above, the football betting market is currently very large and competitive, with many companies, including bookmakers, betting exchanges and spread betting firms operating with low margins. Hence, it is interesting to explore to what extent odds-based information, arising from these different sources, can be used by bettors to increase their returns. Relevant questions to address are (i) whether there is sufficient variation across market prices to provide bettors zero-risk opportunities to earn profit (i.e. 'arbitrage') and (ii) whether variation of market prices signals the arrival of information concerning the probabilities of particular football events, which can increase the accuracy of bettors' forecasts and as a consequence, enable them to earn abnormal returns.

1.2.5. Arbitrage

Pope and Peel (1988) identified arbitrage opportunities in the football betting market. However, later research (Dixon and Pope, 2004), analysing odds from three different bookmakers, found no opportunities for arbitrage. Dixon and Pope suggest that this may be due to a decrease in the variation of odds between bookmakers, resulting from their prices having become more coordinated under the influence of professional arbitrageurs. Similarly, Vlastakis, Dotsis and Markellos (2008) explore a sample of 12,420 football matches, including odds from five different bookmakers and identify only a small number (63) of arbitrage opportunities. However, for the purpose of such analysis, it would be desirable to analyse odds across a much larger sample of bookmakers, as Oikonomidis and Johnson (2008) estimate that "shopping" for best odds across 45 bookmakers should bring the overall over-round close to 0. In a study of odds quoted by 79 different bookmakers, Deschamps (2008) identified a relatively greater number of arbitrage opportunities (293), across the sample of 6,315 games. Similarly, Deschamps and Gergaud (2007) explored odds from several different bookmakers and found that significant price variation existed; indicating that, "shopping" for best odds can significantly increase the bettors' return.

The studies discussed above used odds from the bookmaker market only. However, Franck, Verbeek and Nüesch (2009) analysed the possibility of arbitrage opportunities arising from simultaneously betting on outcomes of the same event in the bookmaker and the betting exchange markets. They found that the development of betting exchanges has significantly increased the frequency of arbitrage opportunities, since they have increased the variation of prices in the market. In a sample of 5,478 games they found only 10 arbitrage opportunities when considering bookmaker prices alone but 1,450 when the analysis was extended to betting exchanges. The existence of arbitrage opportunities has also been examined in prediction markets. Luckner and Weinhardt (2008) use data from the STOCCER championship market (concerning the FIFA World Cup 2006) and found no significant evidence of arbitrage opportunities.

1.2.6. Signals from Variation of Odds

Some papers have examined to what extent cross market variation of prices is random or whether it signals information regarding the probabilities of events. For example, Dechamps (2008) analysed data from diverse bookmakers and found that outlying odds are informative, even after considering average odds. They provided empirical evidence, which suggests that if a bookmaker is willing to offer very high odds relative to the market, this indicates that market odds are lower than they should be. However, Paton and Vaughan Williams (2005), using data from the spread-betting market found evidence to suggest that bookmakers that offer outlying odds do not possess superior information. More specifically, they found that the average mid-point of the quoted spreads from different bookmakers is a more accurate estimation of the real outcome compared to the outlying spread. This market variation was found to be sufficient to enable profitable trading.

Previous research, discussed above, suggests that the simultaneous operation of several betting companies is likely to provide ‘odds shoppers’ with the opportunity to drastically decrease, or even nullify transaction costs and place nearly fair (or even favourable) bets. However, it should be noted that implementing successful ‘arbitraging’, may involve several difficulties, which are not so obvious when theoretically examining this possibility. For example, bookmakers may change their odds or refuse to accept bets at a high level or liquidity on the desired odds may quickly disappear from the betting exchanges. In all cases, this is a business for the fastest and computationally efficient players (see Marshall, 2009).

1.2.7. Summary

Overall, the literature suggests that several types of weak form inefficiencies exist and it appears possible for the bettors to take advantage of them and to at least decrease their losses. However, in order to assert that these inefficiencies are significant and persistent enough to enable bettors to achieve positive returns, more consistent evidence is required. Additionally, even if the theoretical inefficiencies were shown to exist it remains debatable whether successful exploitation is possible.

1.3. Semi-Strong Form Efficiency

1.3.1. Introduction

A market is semi-strong form efficient if market prices incorporate all relevant, publicly available information (Fama, 1970). Consequently, it should not be possible for bettors to use any kind of publicly available information to estimate football event probabilities more accurately than those derived from odds; forecasting models based on fundamental information should, therefore, not lead to profitable betting strategies. A range of studies have tested the semi-strong form efficiency of football betting markets; the methods employed in these studies and their results are reviewed below.

1. 3.2. Forecasting Methods

Several papers have estimated the winning probabilities of competing outcomes of football games, through modelling (i) the expected goals scored (ii) the goal difference, or (iii) the winning outcome directly. In the first instance, count outcome regression models have been used, such as Poisson or modified Poisson (e.g., Maher, 1982; Dixon and Coles, 1997; Karlis and Ntzoufras, 2003; Dixon and Pope, 2004), negative binomial (Reep, Pollard and Benjamin, 1971; Baxter and Stevenson, 1988) and extreme value distributions (Greenhough, Birch, Chapman and Rowlands, 2002). In order to model the expected goal difference between two opponents, Karlis and Ntzoufras (2009) applied the Skellam's distribution. A number of researchers have employed discrete choice models (mainly ordered probit regression) in order to directly estimate the probability of competing events (Kuypers, 2000; Goddard and Asimakopoulos, 2004). Goddard (2005) performed a statistical comparison of forecasting models and found no significant difference in accuracy between models that forecast goals and these that model result directly. More recently, machine learning techniques have been applied to predict game outcomes (e.g., Vlastakis, et al., 2008; Strumbelj, et al., 2009). Finally, combinations of different types of estimation have also been considered. For example, Vlastakis, et al. (2008) suggested encompassing techniques in order to combine forecasts from Poisson and multinomial regression models, weighted according to the accuracy of predictions.

1.3.3 Home Advantage

Home ground advantage plays a major role in deciding football game outcomes; Home win frequency being about twice that of away wins. Crowd support, stadium familiarity and travelling are factors that have been shown to contribute to the creation of the home advantage (Courneya and Carron, 1992), as has referee bias in favour of the home side (Garicano, et al., 2001). However, the existence of this effect does not appear to bias market odds. Pope and Peel (1988) examined data from the 1981/82 season and found no evidence of inefficiency regarding home advantage. Furthermore, Graham and Stott (2008), using data from the top four English leagues for 2001 to 2006, concluded that the home advantage is relatively constant across teams (in contrast though to an earlier study by Clarke and Norman, 1995) and that, bookmaker prices reflect this lack of between-teams variation in home advantage.

Goddard and Thomas (2006) found that home team advantage was underestimated by market odds in the European Championship 2004. However, the small sample size and the dependency of observations do not enable wider conclusions regarding the bias to be drawn.

1.3.4. Performance Measuring Models

Several models incorporating a wide range of publicly available information have been employed in order to test the semi-strong form efficiency of football betting markets. An overview of these is given below.

A number of indices have been used in order to evaluate the abilities of football teams based on their performance in past games, such as the FIFA ranking[4] and ELO ratings[5]. Such ratings exhibit significant forecasting power (Suzuki and Ohmori, 2008; Lasek et al. 2013) in predicting the outcome of football events. However, there is little evidence that they can outperform predictions derived from the odds in the betting market (Leitner et al., 2010) and therefore, the market seems to be efficient in incorporating such information. Nevertheless, more elaborate measures of performance, such as ESPN's Soccer Power Index[6] and the EA Sports Player Performance Index (McHale et al. 2012) incorporate information concerning the individual players in the teams' line-ups have been developed and would therefore be interesting to test whether such indices are informative enough to improve the accuracy of odds-based estimates.

Dixon and Coles (1997) employed a bivariate Poisson model, whose parameters relate to home advantage and past performance (in terms of goals scored and conceded). They suggested several refinements for low scoring probabilities to fit

[4] See http://www.fifa.com/worldranking/procedureandschedule/menprocedure/index.html

[5] See http://www.eloratings.net/system.html

[6] See http://soccernet.espn.go.com/world-cup/story/_/id/4447078/ce/us/guide-espn-spi-ratings

real data more accurately, and they adjusted the likelihood function to incorporate a proximity parameter, to give more weight to recent observations. The model was fitted using English league and cup data between 1992 and 1995 and was found to yield positive returns in an out-of-sample period (1995/1996 season). Similarly, Dixon and Pope (2004) developed a Poisson model that estimated team-specific parameters, concerning the ability to attack and defend,(based on observed outcomes). They tested the model against bookmaker odds for the correct score and the match outright market and found evidence of market inefficiency.

Employing ordered probit regression and using data derived from matches played in England after 1987, Goddard and Asimakopoulos (2004) built a forecasting model for football results, based on a series of fundamental factors. Recent results, particularly those at home for the home team and away for the away team were identified as key forecasting factors. In addition, Goddard and Asimakopoulos (2004) found that the effect of motivation was significant, while geographical distance of travel for the away team increased home ground advantage, in this case. Elimination from the cup competition appeared to have a negative effect on a team's subsequent league results and teams that attracted higher attendances in their previous games were more likely to be successful in future games (controlling for other performance factors, confirming that this is not an omitted variable bias). A model combining these factors was tested against market odds and found to be profitable for high expected profit bets. Likewise, Kuypers (2000) built a model utilizing similar information, using data from the 1993/94 and 1994/5 seasons from the top four divisions of English football and used this to demonstrate some degree of inefficiency. Forrest, Goddard and Simmons (2005) employed a sample of nearly 10,000 English football games from 1998-2003 to test a similar model; they found that their model only produced superior results to market probabilities in the early years. Thus, they suggest that the football betting market has moved towards efficiency as a result of competition between different bookmaking companies, which has forced them to improve their estimations.

1.3.5. Behavioural Biases

Decisions in markets are made by humans and therefore, it might be expected that biases which characterize human judgment will influence the setting of prices, and may lead to inefficiencies. A review of the studies that examine the efficiency of the football betting market in relation to behavioural biases is now presented.

It is commonly believed that casual bettors behave sentimentally and place bets on the team they support and a number of researchers have examined whether such behaviour biases the odds of popular teams. Forrest and Simmons (2002) found that the winning probabilities of popular teams (i.e. defined as teams that achieve high attendance at their home games) are underestimated by market odds. Similar evidence was provided by Goddard and Asimakopoulos (2004) and Forrest and Simmons (2008), analysing data from the top Spanish (2001-2008) and Scottish league (2001-2005), respectively. They suggested that odds are biased in favour of popular teams, because the bookmakers try to attract sentimental bettors. Franck, et al. (2010) documented the same effect when exploring a sample of 16,000 English football games, between seasons 2000/1 and 2007/8. This effect was not apparent on week-day games. The authors suggested that this result was expected assuming that more casual, ‘sentimental’ gamblers bet at weekends (c.f. weekdays), thus increasing the demand for popular teams. Bookmakers, as a consequence, increase the odds for such teams, in order to sustain the competition in a ‘price sensitive’ market.

In order to investigate whether optimism bias exists in the betting market, Page (2009) set the opposite hypothesis, compared to the studies reviewed in the previous paragraph. The author suggested that the existence of an optimism bias among bettors would lead UK betting companies (which are more likely to have a majority of British bettors) to lower the odds for UK teams in international matches (due to the likely high demand). However, it should be noted that this would result from the optimism bias if bookmakers balance their books, but not if they try to attract sentimental bettors, as suggested by the studies reviewed above. Page (2009) analysed odds derived from 161 different betting companies for 3,585 international football matches and 5,301 European cup matches between 1998 and 2007 and found no evidence of optimism bias. On the other hand, Bernile and Lyandres (2011) investigated returns of European football clubs traded in the stock market and found that investors overestimated their teams' expected performance; leading to abnormally negative returns. Consequently, as bookmakers' odds reflect to some extent their desire to attract bets on the popular teams and prediction markets prices are completely demand led, this points to structural pricing-related differences between prediction and bookmaker markets.

In the total goals market, a utility bias has been observed. For example, Rodney and Weinbach (2009) analysed over 15,000 football games played in 22 European leagues. They examined the most common form of betting, i.e. to bet on whether more ("over") or fewer ("under") than 2.5 goals will be scored in the game. They found odds in this market to be significantly biased, as the expected loss for a random bet on 'over' was more than twice the size of the expected loss of a bet on 'under'. They suggested that bettors exhibited a behavioural bias, as they appeared to show a preference for betting on 'over'. Oikonomidis and Johnson (2008) analysed a similar dataset and identified a similar over/under bias. However, this bias was shown to decrease through time. This was shown not to arise from changes in goal scoring frequency but from bookmakers offering significantly lower odds on 'under' (and higher on 'over'), in later years.

1.3.6. Subjective Estimations

Various public media, including newspapers, radio and television programmes and web sources provide bettors with predictions regarding football events. Even though the mechanisms by which such advice is transmitted may be different, they all reflect subjective estimations of a person or a group of persons, involved with the world of football. A number of studies have challenged the value of these subjective predictions. Their findings are presented below.

Using a sample of 1,694 English football games, Forrest and Simmons (2000) tested the value of newspaper tipsters' services. Even though some forecasting ability was observed (tipsters' predictions were significantly better than random), Forrest and Simmons concluded that tipsters fail to adequately account for publicly available information concerning teams' strengths. Moreover, they tested whether tipsters' predictions were informative after performance-measures of team strength were considered and found this to be the case for only one of the tipsters. Thus, these studies offer no convincing evidence that the guidance offered by tipsters is of great value, unless there is no other information available. Andersson, Edman and Ekman (2005) organized a survey concerning football predictions for the 2002 World Cup. 251 participants took part, varying from football fans, journalists and coaches to non-experts. Both experts and non-experts were found to predict better than random, but there was no evidence that experts predicted more accurately, than non-experts. Surprisingly, a simple prediction rule, based on world rankings, achieved superior predictions than most of the participants.

Spann and Skiera (2009) explored data that included stock prices on a prediction market, sports journal tipster predictions and bookmaker odds regarding football matches played in the German Bundesleague in the period 1999 to 2002. They found that predictions based on betting odds and the prediction market achieve approximately the same level of accuracy (which is significantly more accurate than tipsters' forecasts). However, some suggested rules for combining these three prediction sources led to improved forecasts which could be profitable in a 'friendlier' jurisdiction, where transaction costs are lower than the 25% faced in the German market. However, more research is required to confirm this result, as the sample of bets employed in the study was limited.

1.3.7. Betting In-Running

Betting during the course of a football game has become extremely popular and Hill (2009), estimated that half of betting activity takes place in-running. As a result a number of researchers have examined whether the market is able to efficiently incorporate the continuous, dynamic flow of information arising from live football action. Using a dataset of 4,000 English football matches, Dixon and Robinson (1998) developed a birth process model to predict football outcomes during the course of a football game, based on the home advantage, the attacking and the defensive abilities of the teams, the current score and the time remaining until the end of the game. They found that the scoring rate increased through the game and therefore, that a non-homogeneous process is appropriate to model the expected result. They also found that the scoring rate depended on the current score, and, in particular, the scoring rates of both teams decreased significantly when the home side held a narrow lead. Dixon and Robinson (1998) tested the model against spread betting odds and found some evidence of inefficient pricing.

In a more recent study, Croxson and Reade (2013) used in-running data (concerning 1,206 football matches played in various competitions) from the betting exchange market. The response of the market to significant updates of information (i.e. goals scored) was compared to updated theoretical odds based on a Poisson model fitted to historical data. No evidence of inefficiency in price setting was found, while no relationship between liquidity and inefficient pricing was identified.

1.3.8. Summary

As indicated above, odds have been shown to be successful predictors of football outcomes. In this section, it is asserted that their forecasting power is comparable to that of sophisticated, fundamental models that utilize a range of publicly available information. However, some researchers found evidence concerning the existence of semi-strong inefficiency and others do not, both sets employing similar information, across different samples. In parallel, odds have been shown to be more efficient in responding to several types of information compared to others It seems to be clear though that in the recent years odds-setting has improved, posing a more difficult challenge to those intent to making profit from betting on football.

1.4. Strong Form Efficiency

A market exhibits strong form information efficiency if prices fully reflect all publicly and privately held information (Fama, 1970). Consequently, the football betting market is strong form inefficient if some market operators possess ‘superior’ information regarding the ‘true’ odds of football events. Thus, in football, where allegations of match fixing have surfaced in recent years, it may be the case that individuals involved in match fixing are exclusively aware of the fact that some outcome is very likely to occur; they may, therefore, use this information to make profit from betting- leading to a strong form inefficient market In the following section we examine cases of match fixing in football and their association with the betting market.

Numerous betting-related football scandals have been revealed through the years. In early 1960s, a group of players, organized by Jimmy Gauld, fixed the outcomes of several football games, in order to profit from betting against bookmakers (Preston and Szymanski, 2003). More recently, the goalkeepers of Liverpool and Wimbledon were accused of accepting bribes to fix games in between 1993 and 1994 and this was linked with Asian betting syndicates (Preston and Szymanski, 2003). In Italy, individuals have confessed to attempting to fix games, with the purpose of profiting from betting, during 1979/80 period, and in Germany, the referee Robert Hoyzer was convicted of match-rigging in 2005. Betting-related match fixing cases have also been identified in Malaysia (Hill, 2009).

The consequences of match-fixing are detrimental to all participants in the betting market, excluding the match-fixers themselves. From the bookmakers’ point of view, match fixing may lead to significant betting losses (to the match fixers) and the demand for betting from honest bettors may decrease if such events trigger doubts concerning the fairness of the game (Hosmer-Henner, 2010). From the bettors’ perspective, they may lose money directly to the match fixers in the betting exchange market. Thus, football fans and authorities, bettors and the betting industry all have an interest in keeping the game ‘clean’. Consequently, we discuss below how match fixing can still take place.

1.4.1. How Are Matches Fixed?

Hill (2009) conducted over 220 interviews with match fixers, players, referees, sports and law enforcement officials and agents in the gambling industry. He created the ‘Fixed Match Database’, which includes matches presumed to have been played honestly together with 130 legally certified fixed football matches. The database also includes a sample of 117 players who were approached to fix matches (of whom, 24 refused). Hill (2009) shows that for a match fixer to be successful five stages of corruption have to be completed successfully: ‘access’, ‘set up’, ‘calling the fix’, ‘performance’ and ‘payment’. Initially the match fixer needs to gain access (directly or through an agent) to at least one influential player, who will then organise a network within the team to undertake the match fixing operation. Then, the most suitable way to set up the arrangement has to be identified (who to approach and how). Hill (2009) shows that, depending on the type of the game, the match fixers’ approach may be more or less personal. Hill’s data suggests that corrupted players tend to underperform to achieve the desired outcome and that match fixers pay some money to the players in advance, but the main payment is made after the desired result is achieved(usually in cash and not through any sophisticated network).

Match-fixing by gambling syndicates has been documented, even without the involvement of the main participants in the game (i.e. players, coaches, managers and referees). For example, in 1999 an Asian gambling syndicate sabotaged the lighting systems of English football stadiums while the score of the game was favourable to them, resulting in high gambling profits (Hosmer-Henner, 2010).

To profit from betting on a fixed game, the match fixers also need to operate successfully in the betting market. They need to explore the type of bet that will maximize their profit, identify ways to place high stakes (usually disguising their identity) and to ensure that the players will bring about exactly the desired outcome. According to Hill (2009), in order to remain unnoticed, the match fixers choose games in which the betting market liquidity is high; so that their actions will not cause significant moves in the odds. Alternatively, they may bet on favourites, and profit by ensuring the realization of an expected result, which will naturally cause little suspicion. Spreading rumours concerning fixing related to the team opposing the one actually approached has also been documented; the aim being to stimulate bets on the opposing team and increasing the odds on the team they intend to bet. Finally, the results of Hill's (2009) research suggests that the corruptors are also more likely to enter the market late, so as not to signal information that may decrease the odds of the team they intend to back.

1.4.2. Identifying Potentially Fixed Games

The Union of European Football Associations (UEFA) has set up betting fraud detection systems across Europe in order to investigate 27,000 matches played across all the associations (see http://news.bbc.co.uk/sport2/hi/football/europe/7964790.stm). Betting companies and have also established ''early warning systems' aimed at the identification of fixed matches (Hill, 2009). It might be thought that the analysis of fundamental statistics regarding football events may also assist in the discovery of fixed games. However, Hill (2010) compared 137 fixed matches to 120 matches that were (or at least assumed to have been) played honestly and uncovered little statistical evidence against dishonest players. The problem is that players intending to fix a result appear to prefer to under-perform rather than to conduct serious, notable errors, such as own goals or conceding penalties (which are too readily identified). Nevertheless, it was found that the goal-scoring rate in fixed games was higher at the beginning of the game and decreased near its end; the opposite trend to that observed in non-fixed games (Dixon and Robinson, 1998).

1.4.3. Summary

There is evidence that fixing matches of football matches is possible, and that it has taken place in different types of competition in many countries. The large size of the gambling market induces match fixers to attempt to profit from fixing football outcomes. In more recent years, football and betting authorities have established intelligence systems to identify suspicious games and therefore, the match fixers' task has possibly become more difficult. However, it is very difficult for authorities to prevent it completely, as long as the betting market is characterized by high liquidity. Thus, it seems likely that strong form inefficiency in the football betting market will continue to exist.

1.5. Conclusions

The study of the literature confirms that the football betting market, like most other markets exhibits several types of information inefficiency. Thus, information concerning the odds, fundamental data associated with teams' performances, psychological biases and even, inside information may be utilized by gamblers to achieve positive returns. However, it is clear that the football betting market is dynamic, and the observed inefficiencies are not necessarily persistent through time. Moreover, it is likely that any strategy that aims to exploit market inefficiencies will be subject to difficulties associated with implementation. Thus, even though opportunities for profit theoretically exist, only the fastest, most efficient and highly determined players are likely to convert theory to practice and benefit from inefficient pricing in the football betting market.

References

Andersson, P., Edman, J., and Ekman, M. (2005). Predicting the World Cup 2002 in soccer: performance and confidence of experts and non-experts. *International Journal of Forecasting*, 21, 565-576.

Baxter, M., and Stevenson, R. (1988). Discriminating Between the Poisson and the Negative Binomial Distributions: An Application to Goal Scoring in Association Football. *Journal of Applied Statistics*, 15, 347-348.

Bernile, G., and Lyandres, E. (2011). Understanding Investor Sentiment: The Case of Soccer. *Financial Management*, 40, 357-380.

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

Cain, M., Law, D., and Peel, D. (2001). The incidence of insider trading in betting markets and the Gabriel and Marsden anomaly. *Manchester School*, 69, 197-207.

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

Clarke, S. R., and Norman, J. M. (1995). Home ground advantage of individual clubs in English soccer. *Journal of the Royal Statistical Society, Series D*, 44(4), 509-521.

Courneya, K. S., and Carron, A. V. (1992). The home advantage in sport competitions: A literature review. *Journal of Sport and Exercise Psychology*, 14 , 13-27.

Croxson, K., and Reade, J. J. (2013). Information and Efficiency: Goal Arrival in Soccer Betting. *The Economic Journal, 124(575), 62-91.*

Deschamps, B. (2008). Betting Markets Efficiency: Evidence from European Football. *The Journal of Gambling Business and Economics,* 2(1), 66-76.

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

Dixon, M. J., and Coles, S. G. (1997). Modelling Association Football scores and inefficiencies in the football betting market. *Applied Statistics*, 46, 265-280.

Dixon, M., and Pope, P. (2004). The value of statistical forecasts in the UK association football betting market. *International Journal of Forecasting*, 20, 697-711.

Dixon, M. J., and Robinson, M. E. (1998). A birth process model for association football matches. *The Statistician*, 47, 523-538.

Fama, E. F. (1970). Efficient Capital Markets: A Review of Theory and Empirical Work. *The Journal of Finance,* 25(2), 383-417.

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

Forrest, D., and Simmons, R. (2000). Forecasting sport: The behaviour and performance of football tipsters. *International Journal of Forecasting*, 16, 317-331.

Forrest, D., and Simmons, R. (2002). Outcome Uncertainty and Attendance Demand in Sport: The Case of English Soccer. *The Statistician,* 51, 229-241.

Forrest, D., and Simmons, R. (2003). Sport and gambling. *Oxford Review of Economic Policy*, 19(4), 598-611.

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

Franck, E., Verbeek, E., and Nüesch, S. (2009*).* Inter-Market Arbitrage in Sports Betting. *NCER Working Paper No 48, Series from National Centre for Econometric Research.*

Franck, E., Verbeek, E., and Nüesch, S. (2010). Prediction accuracy of different market structures - bookmakers versus a betting exchange. *Sports Forecasting*, 26(3), 448-459.

Garicano, L., Palacios, I., and Prendergast, C. (2005). Favoritism Under Social Pressure. *Review of Economics and Statistics*, 87, 208-216.

Global Betting and Gaming Consultants (2001). 1st Annual Review of the Global Betting and Gaming Market. West Bromwich: Global Betting and Gaming Consultants.

Goddard, J. (2005). Regression models for forecasting goals and match results in association football. *International Journal of Forecasting*, 21, 331-340.

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

Goddard, J., and Thomas, S. (2006). The Efficiency of the UK Fixed-Odds Betting Market for Euro 2004. *International Journal of Sport Finance*, 1, 21-32.

Graham, I., and Stott, H. (2008). Predicting bookmaker odds and efficiency for UK football. *Applied Economics*, 40, 99-109.

Greenhough, J., Birch, P. C., Chapman, S. C., and Rowlands, G. (2002). Football goal distributions and extremal statistics. *Physica A*, 316, 615-624.

Hill, D. (2009). How Gambling Corruptors Fix Football Matches. *European Sport Management Quarterly*, 9(4), 411-32.

Hill, D. (2010). Strategies of Corruption: How Corrupt Football Players and Referees Perform Fixed Matches. In M. Kreutner (Ed.), *Practice Meets Science: Contemporary Anti-Corruption Dialogue IACSS 2009* (pp. 35-50). Wien 2010: Manzsche Verlags- und Universitatsbuchhandlung.

Hosmer-Henner, A. (2010). Preventing Game Fixing: Sports Books as Information Markets. *Gaming Law Review and Economics*, 14(1), 31-38.

Jones, P., Clarke-Hill, C. M. and Hillier, D. (2000). Viewpoint: back street to side street to high street to e-street: sporting betting on the Internet. *International Journal of Retail & Distribution Management*, 28(6), 222-227.

Jones, P., Hillier, D., Turner, D., and Comfort, D. (2004). Betting on the exchanges: changing customer relationships in the sports betting market in the UK. *Management Research,* 27, 95-103.

Karlis, D., and Ntzoufras, I. (2003). Analysis of Sports Data by Using Bivariate Poisson Models. *The Statistician,* 52(3), 381-393.

Karlis, D., and Ntzoufras, I. (2009). Bayesian modelling of football outcomes: using the Skellam's distribution for the goal difference. *IMA Journal of Management Mathematics,* 20, 133-145.

Kuypers, T. (2000). Information and efficiency, an empirical study of a fixed odds betting market. *Applied Economics,* 32, 1353-1363.

National Gambling Impact Study Commission (1999). Final Report, Washington, D.C.

Lasek, J., Szlávik, Z., and Bhulai, S. (2013). The predictive power of ranking systems in association football. International Journal of Applied Pattern Recognition, 1(1), 27-46.

Luckner, S., and Weinhardt, C. (2008). Arbitrage Opportunities and Market-Making Traders in Prediction Markets. *IEEE Joint Conference on E-Commerce Technology (CEC'08) and Enterprise Computing, E-Commerce and Eservices (EEE '08)*. USA: Washington, D.C.

Leitner, C., Zeileis, A., and Hornik, K. (2008). Who is going to win the EURO 2008? A statistical investigation of bookmakers odds. *Report 76, Department of Statistics and Mathematics, Wirtschaftsuniversität Wien, Research Report Series.*

Leitner, C., Zeileis, A., and Hornik, K. (2009). Forecasting the winner of the UEFA Champions League 2008/09. In R. Koning, and P. Scarf (Eds.), *Proceedings* *of the 2nd International Conference on Mathematics in Sport* (pp. 94-99). The Institute of Mathematics and its Applications.

Leitner, C., Zeileis, A., and Hornik, K. (2010). Forecasting sports tournaments by ratings of probabilities: A comparison for the EURO 2008. International Journal of Forecasting, 26(3), 471-481.

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

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

Malaric, R., Katic, T., and Sabolic, D. (2008). The market efficiency of the soccer fixed odds internet betting market. *Applied Economics Letters,* 15(3), 171-174.

Marshall, B. R. (2009). How quickly is temporary market inefficiency removed? *The Quarterly Review of Economics and Finance,* 49(3), 917-930.

McHale, I. G., Scarf, P.A., and Folker, D. A. (2012). On the Development of a Soccer Player Performance Rating System for the English Premier League. Interfaces, 42(4), 339-351.

Mintel Leisure Intelligence Report (2001). Online Betting. London: Mintel International Group Limited.

Oikonomidis, A., and Johnson, J. E. V. (2008). Exploring the efficient market hypothesis through time; Evidence from the total goals football betting market. *Symposium of the Southern Economic Journal: Gambling, Prediction Markets and Policy Implications.* UK: Nottingham, September 2008.

Oikonomidis, A., and Johnson, J. E. V. (2009). An exploration of influences on transaction costs in betting markets. *14th International Conference on Gambling and Risk Taking.* USA: Reno , 25 - 29 March 2009.

Oikonomidis, A., and Johnson, J. E. V. (2010). Subjective Judgment vs Market Idiosyncrasies; Exploring Influences on the Favourite Longshot Bias in

European Betting Markets. *24th European Conference on Operational Research*. Portugal: Lisbon, 11-14 July 2010.

Page, L. (2009). Is there an optimistic bias on betting markets? *Economics Letters,* 102(2), 70-72.

Paton, D., and Vaughan Williams, L. (1998). Do Betting Costs Explain Betting Biases? *Applied Economics Letters,* 5, 333-335.

Paton, D., Siegel, D. S., and Vaughan Williams, L. (2002). A Policy Response to the E-commerce Revolution: The Case of Betting Taxation in the UK. *Economic Journal,* 112(480), F296-F314.

Paton, D., and Vaughan Williams, L. (2005). Forecasting outcomes in spread betting markets, can bettors use quarbs to beat the book? *Journal of Forecasting,* 24, 139-154.

Pope, P. F., and Peel, D. A. (1988). Information, prices and efficiency in a fixed odds betting market. *Economica*, 56, 323-341.

Pope, P., and Peel, D. (1990). Information, prices and efficiency in a fixed-odds betting market, *Economica*, 56, 323-341.

Preston, I., and Szymanski, S. (2003). Cheating in Contests. *Oxford Review of Economic Policy,* 19, 612-624.

Reep, C., Pollard, R., and Benjamin, B. (1971). Skill and chance in ball games. *Journal of the Royal Statistical Society, Ser. A,* 131, 581-585.

Rodney, P. J., and Weinbach, A. P. (2009). Are Behavioral Biases Consistent Across the Atlantic? The Over/Under Market for European Soccer. *The Journal of Gambling Business and Economics,* 3(2), 89-101.

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

Servan-Schreiber, E., Wolfers, J., Pennock, D. M., and Galebach, B. (2004). Prediction Markets: Does Money Matter? Electronic Markets, 14(3), 243-251.

So, H. T., and Kwok, N. K. (2007). No longer an entertainment: Social costs of problem gambling in Hong Kong. *Journal of Macau Studies*, 41, 94-104.

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

SportBusiness Group (2001). Online Gambling: The Regulation of Global Sports Betting: Executive Summary, available at www.sportbusiness.com

Strumbelj, E., and Sikonja, M. R. (2010). Online bookmakers' odds as forecasts: The case of European soccer leagues. *Sports Forecasting*, 26(3), 482-488.

Strumbelj, E., Sikonja, M. R., and Kononenko, I. (2009). Learning Betting Tips from Users' Bet Selections. In P. Perner (Ed.), *Machine Learning and Data Mining in Pattern Recognition* (pp. 678–688). Berlin: Springer-Verlag.

Suzuki, K. and Ohmori, K. (2008). Effectiveness of FIFA/Coca-Cola World Ranking in predicting the results of FIFA World Cup Finals. Football Science, 5, 18-25.

Vaughan Williams, L. (Ed.) (2005). *Information Efficiency in Financial and Betting Markets*. New York: Cambridge University Press.

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

Vaughan Williams, L., and Paton, D. (1998). Why are some favourite-longshot biases positive and others negative? *Applied Economics,* 30, 1505-1510.

Vlastakis, N., Dotsis, G., and Markellos, R. N. (2008). Nonlinear modelling of European football scores using support vector machines. *Applied Economics,* 40(1), 111-118.

Vlastakis, N., Dotsis, G., and Markellos, R. N. (2009). How efficient is the European football betting market? Evidence from arbitrage and trading strategies. *Journal of Forecasting,* 28, 426-444.

Wong, Y-L, R., Leung, Y. K. T., and Lau, C. W. D. (2009). Behind the allure of gambling: A qualitative exploration of the existential yearnings of Chinese men with problem gambling in Hong Kong. *International Gambling Studies,* 9(3), 189-205.

Woodland, L. M., and Woodland, B. M. (1994). Market efficiency and the favorite-long shot bias: the baseball betting market. *Journal of Finance,* 49(1), 269-280.

Chapter 2

Bias and Efficiency in European Markets for State Contingent Claims

Abstract

This paper uses the lens of homogeneity to examine the apparent diversity in efficiency outcomes which characterizes the financial market literature. To achieve this we employ a large sample of similar markets for state contingent claims, geographically and temporally dispersed, to examine the extent to which they differ in terms of their weak form efficiency/inefficiency. Specifically, we compare the degree to which past prices are effectively discounted in current prices in betting markets associated with the outcome of 52,865 soccer matches played between 1999 and 2008 in twenty two leagues across eleven European countries. We observe a significant degree of bias in prices across the European market as a whole but conclude that important differences exist in the degree of efficiency in the separate markets; in particular, we identify significant differences in the size of the favourite longshot bias between leagues. We test and reject the hypothesis that the observed variations in efficiency are caused by different levels of transaction costs in these markets. However, we find that the efficiency differences are affected by outcome predictability.

Keywords

Market efficiency, transactions costs, cross-market differences, favourite longshot bias

2.1. Introduction

Financial markets are regarded as efficient only when the prices of assets reflect all market-relevant information. As such, a condition of efficiency is that expected returns to all assets are identical. According to Fama's (1970) categorization, weak form efficiency implies that prices fully reflect any information embodied in past prices and price paths. Semi-strong form efficiency requires that prices reflect all publicly available information, while the strong form of market efficiency additionally requires that prices reflect information which is restricted to a certain group of individuals

Empirical studies offer a diversity of results relating to financial market efficiency and a range of explanations for apparent deviations from efficiency. Fama (1998) notes that the empirical literature is characterized by both market over- and under-reaction to information. However, he is skeptical as to whether, overall, the efficient market hypothesis can be rejected. Fama (1998) argues that three key conditions need to be present in order to reject the efficiency hypothesis: (i) *Endurance*: the inefficiency should remain significant across a sample covering several years, (ii) *Homogeneity*: the inefficiency should be apparent, and similarly manifested, in different markets, and (iii) *Robustness*: the methodology should be robust enough to validate the existence of the inefficiency with a significant level of confidence. In other words, if an anomaly is long lasting, homogeneous across different markets and validated by robust tests, its existence can be taken to imply the presence of inefficiency.

This paper focuses on the homogeneity condition in exploring weak form efficiency in one form of financial market across eleven European countries. This approach starts from the premise that adopting a tight focus on a particular market context maximises the prospect that the homogeneity condition will hold across a range of national locations. As such, it offers a marked contrast with the wider body of empirical literature on financial market efficiency, which is characterised by a high level of diversity in terms of the forms of market under scrutiny, with a corresponding diversity in the degree and nature of efficiency identified. Within this context, we make a number of important contributions in this paper: First, we observe a significant degree of bias in prices across the European market as a whole. Second, we identify a lack of homogeneity in market efficiency across directly comparable markets. Third, we are able to reject a previously cited cause of heterogeneity in market efficiency (variations in transaction costs) and, finally, identify ‘outcome predictability’ as a key source of variation in market efficiency.

The context which forms the focus for the analysis is a market for state contingent claims which has received a significant level of empirical attention - the betting, or wagering, market. An immediate advantage of this type of market in testing for homogeneity is that, as Thaler and Ziemba (1988) observe, it is better suited to analysing market efficiency and rational expectations than stock or other asset markets. This arises because in wagering markets (unlike stock markets) each asset or bet has a well-defined termination point at which all uncertainty is resolved and its value becomes certain. In particular, an unequivocal outcome is generated within a finite time frame and this provides an objective benchmark against which to measure the quality of the decision to purchase (place) a particular claim (bet) (Law and Peel, 2002). Consequently, wagering markets “can provide a clear view of pricing issues which are more complicated elsewhere” (Sauer, 1998, p.2021), and as the number of events on which bets can be placed is large, there is a large pool of similar markets available for analysis. As a result, betting markets have been employed by numerous researchers to shed light on investors’ behaviour in wider financial markets (e.g., Dowie, 1976; Schnytzer and Shilony, 1995; Law and Peel, 2002; Levitt, 2004). This is facilitated by the characteristics shared by betting markets and wider financial markets, including ease of entry, extensive market knowledge and large numbers of participants; market makers, noise traders and informed traders (Snyder, 1978). In addition, the factors which influence a bet’s prospects (or an asset’s value) are interdependent and complex.

The richness of betting markets as a medium for the investigation of financial market efficiency is augmented by their growing contemporary significance. The last decade has witnessed a huge growth in operating revenue, tax revenue and levels of participation across all jurisdictions where betting is legal.

The paper proceeds as follows. A review of the literature relating to weak form efficiency in betting markets is provided in section1. Section 2 explores the nature of fixed odds betting markets, which form the focus of our enquiry Section 3 describes the data and methods employed in the investigation of cross-market efficiency differentials and explains the basis for examining the influence of transactions costs and event competitiveness on the observed differences. Results of the analyses are presented and discussed in section 4. Some concluding remarks follow in section 5.

2.2. Weak form efficiency in betting markets: The literature

2.2.1. Horserace betting markets

In betting markets, weak form inefficiencies occur when the expected returns to assets (bets) with particular prices ('odds') are significantly more favourable than others. The most documented form of this phenomenon is the traditional or 'positive' favourite-longshot bias (FLB), first identified in horserace betting markets, whereby favourites/longshots (horses with the shortest/largest odds in their market)have been shown to win more/less often than the subjective probabilities implied by their odds suggest. This phenomenon has attracted significant interest over the last forty years, with its presence in a variety of horserace betting contexts being empirically validated (e.g., *USA*: Ali, 1977; Asch et al., 1982; Snowberg and Wolfers, 2005; Snyder, 1978; Thaler and Ziemba, 1988; *Australia*: Bird and McRae, 1994; Tuckwell, 1983; *New Zealand*: Gander et al., 2001; *UK*: Bruce and Johnson, 2000; Crafts, 1985; Dowie, 1976; VaughanWilliams and Paton, 1997). Exceptions have been identified in the Hong Kong and Japanese horserace betting markets (e.g. Busche and Hall, 1988; Busche, 1994).

2.2.2. American football betting markets

Investigation of the FLB has been extended to wider sports-betting markets, with American football attracting most attention. Results vary depending on the sample investigated. Where inefficiency has been identified, it is generally in the form of a negative FLB. For example, Vergin and Scriabin (1978) and Tryfos et al. (1984), concluded that betting on big underdogs was profitable. Similarly, Golec and Tamarkin (1991), found that odds consistently under-estimated the chances of underdogs in National Football League (NFL) and collegiate American football games over a fifteen year period(1973-1987).On the other hand, many papers have suggested that the American football market is efficient. For example, Zuber et al. (1985) found that point-spreads were unbiased predictors of NFL game outcomes (albeit, using data from one season only).Similarly, Dare and McDonald (1996) found that the NFL and college betting markets were efficient (with the exception of Superbowls). Finally, Paul et al. (2003) found that the collegiate football betting market was efficient in general, even though the winning probability of certain types of longshot tended to be underestimated.

2.2.3. Other sports betting markets

Colquitt et al., (2004) analysed nearly 16,000 bets in the American collegiate basketball (NCAA) point-spread market and found that the longshot failed to cover the point spread more than 50% of the time, providing evidence of positive FLB. Similarly, a positive FLB has been observed by Forrest and McHale (2007) in a sample of 8,500 men’s singles tennis matches played between 2002 and 2005 and by Cain et al. (2003) in boxing, snooker and cricket betting markets. By contrast, Metrick (1996) found that the heaviest favourites in NCAA pools were over-bet, although the size of this bias fell slightly in larger pools. Similarly, negative FLB has been found in betting markets associated with major (US) league baseball (Woodland and Woodland, 1994 and 2003) and with the National (US) Hockey League (Woodland and Woodland, 2001).

2.2.4 Soccer betting markets

Most studies of soccer betting markets have reported positive FLB, though some exceptions have been identified. For example, positive FLB was identified in soccer betting markets by Kuypers (2000) (four English leagues, 1993-1995), by Malaric et al., (2008) (12,128 soccer games, 10 European leagues (German, English, Danish, Spanish, Italian, French, Scottish, Austrian, Belgian and Dutch,1999-2002), and by Graham and Scott (2008) (11,000 matches, four main English leagues, 2001-2006). By contrast, Cain, et al. (2003) found the football betting market to be generally efficient.

It is clear from this brief review of the literature that betting markets mirror wider financial markets in yielding a diversity of results in relation to the degree of weak form efficiency/inefficiency. In a sense this is unsurprising given the diversity of institutional detail across individual countries' betting markets.

A distinctive feature of this paper is its potential for addressing directly the degree to which similar markets are homogenous in terms of market efficiency and for exploring the causes of any observed heterogeneity. This is achieved via a simultaneous analysis of international fixed odds betting markets relating to the outcomes of soccer fixtures. Specifically, we analyse activity in markets relating to games played in 22 leagues across 11 European countries throughout a 10 year period, with the focus being on the existence, nature and consistency across locations of weak form inefficiencies. Further analysis then addresses the factors influencing cross-location differences.

2.3. The fixed-odds soccer betting market

The win-lose-draw fixed odds soccer betting market corresponds to a market for contingent claims with three states which correspond to the outcomes of the game (home win, away win, draw). In state contingent claims terms, the purchase price of a claim on outcome i in game j (q_{ij}) which pays £1 if outcome i occurs and nothing if it does not occur, is given by $1/(1+O_{ij})$, where O_{ij} represents outcome *i*'s quoted market odds (Shin, 1993). Bookmakers typically publish the odds relating to each possible outcome of a game some days before the game takes place. There are cases when these odds change prior to the start of a game, but this is not generally the case. Placing a bet at these odds is only possible prior to the game and the odds at which a bet is settled will be those which prevailed at the time the bet was made. Odds can be expressed in both decimal and fractional form. Apart from the UK, where the fractional presentation of the odds is still dominant, the decimal form of the odds is generally employed. Fractional odds for outcome *i* in game *j*, O_{ij} express the net profit to a successful bet per unit of stake; the equivalent decimal form, $X_{ij} = O_{ij} + 1$, gives the gross profit (i.e. including returned stake) to a successful unit stake bet. In the case of an unsuccessful bet, the loss is equal to the stake.

From the bettor's perspective, there are only two possible outcomes in fixed odds betting, win or lose, and the bettor's loss is limited by the stake. In other words the net expected profit of a unit stake bet on outcome *i* in game *j* with odds X_{ij} is:

$$E(X_{ij}) = (X_{ij} - 1)p(W_{ij}) - (1 - p(W_{ij})) \quad (1)$$

where $p(W_{ij})$ = probability of outcome *i* occurring.

For a rational, risk neutral bettor, (1) indicates that if $E(X_{ij}) = 0$, the bettor is indifferent to whether he/she bets on event *i*. However, a bet should be placed when $E(X_{ij}) > 0$ and avoided when $E(X_{ij}) < 0$. The sign of the expected profit depends on the probability of event *i*. In a weak form efficient market the expected profitability of each bet is independent of the bet's odds and in the absence of trading costs the probability of each outcome would be equal to the inverse of the odds: $p(W_{ij}) = 1/X_{ij}$. Substituting this expression in (1) gives $E(X_{ij}) = 0$. In other words if trading costs are zero and the only available information is the price (odds) of the bet, the expected profitability of each bet is 0. However, the existence of positive transaction costs modifies the situation. In the context of fixed odds markets, transaction costs are expressed as the amount by which the sum of the probabilities implied by the odds relating to the alternative outcomes exceeds unity (the 'over-round'). This, in effect, increases the market maker's (bookmaker's) prospects of generating a profit. Positive transaction costs effectively decrease all odds by the value of the over-round. Thus, where c is the market's over-round, $p(W_{ij}) = 1/X_{ij}(1+c)$. Substituting this expression in (1) indicates that, for a unit stake bet, $E(X_{ij}) = -c/(1+c)$. The quantity $c/(1+c)$ is the discount rate for each bet associated with an event in which transaction costs are equal to *c*. Consequently, in a weak form efficient market, each bet has negative expected profitability (which depends on the level of trading costs) and this remains constant across all odds levels. This implies that in the long term (after a statistically significant number of bets) bettors who tend to bet on favourites will suffer the same losses as those who tend to bet on longshots. Therefore, if analysis reveals sustained differential returns to bets in different odds categories, this constitutes evidence for the existence of weak form inefficiency.

2.4. Data and methodology

2.4.1. Data

Odds and match outcome data are drawn from twenty two leagues across eleven European countries. The data are sourced from Gamebookers[7], one of the most popular international internet bookmakers with a worldwide customer base of approximately 145,000.The nature of the customer base is important to emphasise. In particular, whilst the focus for analysis is betting activity across twenty two leagues and eleven countries, betting in relation to each league represents the decisions of an international clientele, rather than simply customers located in the relevant country. The use of a single bookmaker's odds, particularly a bookmaker of the size and

[7] We chose this bookmaker (https://sports.gamebookers.com) because the data was slightly more consistent compared to others, due to this company's early focus on the internet market. The empirical part of this paper was replicated employing data from William-Hill and Ladbrokes and this produced similar results.

influence of Gamebookers, is unlikely to introduce bias in exploring market efficiency, since Dixon and Pope (2004) and Forrest et al. (2005) found only minor differences between odds offered by diverse bookmakers[8].This is not surprising because if one bookmaker's odds deviate from the rest of the market it is likely that arbitrageurs will force them back to the mean. In fact Marshall (2009) found that the median duration of arbitrage opportunities in sports betting markets was only 15.4 minutes.

The dataset is one of the largest ever employed in a sports betting market study and as such it overcomes the criticisms of inadequate data levelled at previous enquiries. The data cover 52,865 individual games/markets between 1999 and 2008. Data for each game include the odds for each of the three possible outcomes, together with the outcome of the game. The simultaneous analysis of activity across such a diversity of betting markets is a distinctive feature of this study.

2.4.2. Methods

This section details two distinct stages of analysis. The first, comprising of two strands of enquiry, is designed to investigate evidence for cross-market efficiency differentials. The second stage allows an investigation of the influence of (i) transactions costs and (b) league competitiveness in explaining observed efficiency differentials.

2.4.2.1. Exploring Homogeneity: Efficiency in European fixed odds soccer betting markets

In order to test the following hypothesis, *the fixed odds soccer betting market across Europe is weak form efficient and homogeneous*, we conduct two strands of analysis: First, we explore whether the average level of profit (loss) for a unit stake bet

[8] All of these studies, as well as the current one, employ data from bookmakers, which in Chapter 3 we classify as position-takers. As it will be shown in Chapters 3 and 4, inferences concerning the operations of position-takers do not necessarily apply to book-balancers.

remains constant across all odds levels both for the European market as a whole and in separate markets associated with each league. Second, we conduct conditional logit (CL) regression analyses to compare the degree of FLB in the betting markets associated with the various leagues. Each of these approaches is detailed below.

Variations in returns across odds categories: In each element of the analysis the units of observation are the odds of match outcomes. We first group odds (irrespective of the match outcome to which they relate) into twelve categories, with the boundaries between categories designed to ensure (for the purposes of statistical robustness) sufficient observations per category. Average profit for odds category *h* is calculated by assuming that a unit stake is placed on each betting opportunity with odds X_{ij} (for outcome *i* in game *j*) in category *h*. If a bet is successful the profit is equal to X_{ij}-1 and if the bet proves unsuccessful the profit (loss) is equal to -1. This procedure is undertaken for both the aggregate sample and for each league separately. In a perfectly efficient market, without trading costs, the expected profit in all odds categories should be zero. In practice, with transactions costs of *c*, expected profits are calculated by subtracting the discount factor discussed above ($c/(1+c)$) from zero, giving a constant expected profit less than zero across all odds categories. If abnormal profits are observed in any odds category this will suggest evidence of weak form inefficiency. We also examine the variation in the actual/expected return ratio in each odds category. When the rate exceeds (is lower than) 1, this is an indication of the market's underestimation (overestimation) of betting opportunities in that category.

Modelling winning probabilities to assess FLB: The second strand of analysis involves the modelling of winning probabilities in order to measure the degree of FLB evident in the odds available for matches across all the 22 leagues and for matches within each league.

To achieve these objectives a conditional logistic regression is employed, where the outcome of each game is the dependent variable, which takes value 1 for the event that occurred (e.g. draw) and 0 for those that did not occur (e.g. home/away win). Since each game has three possible outcomes, the probability of outcome *i*in game *j* occurring, is then given by:

$$p(Y_{ij}=1)=e^{Z_{ij}}\Big/\sum_{i=1}^{3}e^{Z_{ij}} \qquad (2)$$

where Z_{ij} is a function of variables which could influence the probability of different outcomes occurring. We define p_{ij}^{s} as the probability of outcome *i* occurring for game *j*, as implied by the bookmaker's odds. As demonstrated above, in a betting market for game *j* with transaction costs equal to c_j, $p_{ij}^{s}=1/X_{ij}(1+c_j)$, where X_{ij} is the odds of outcome *i* in game *j* expressed in decimal form. Consequently, if we set $Z_{ij}=b*Ln(p_{ij}^{s})$ in (2) we obtain:

$$p(Y_{ij}=1)=e^{b*Ln(p_{ij}^{s})}\Big/\sum_{i=1}^{3}e^{b*Ln(p_{ij}^{s}}=(p_{ij}^{s})^{b}\Big/\sum_{i=1}^{3}(p_{ij}^{s})^{b} \qquad (3)$$

A positive (negative) FLB means that as the subjective (implied by odds) probability (p_{ij}^{s}) of an event increases the objective (actual) probability of the same event (p_{ij}^{o}) increases at a higher (lower) rate. In other words, the FLB is positive only if:

$$p_{if}^{s}>p_{il}^{s}\Rightarrow p_{if}^{o}\big/p_{if}^{s}>p_{il}^{o}\big/p_{il}^{s} \qquad (4)$$

Substituting the expression for p (Y_{ij}=1) derived from (3) for p_{ij}^{o} in (4) suggests that for a positive FLB:

$$p_{if}^{o}\big/p_{if}^{s}>p_{il}^{o}\big/p_{il}^{s}\Rightarrow(p_{if}^{s})^{b}\Big/p_{if}^{s}\sum_{i=1}^{3}(p_{if}^{s})^{b}>(p_{il}^{s})^{b}\Big/p_{il}^{s}\sum_{i=1}^{3}(p_{il}^{s})\Rightarrow(p_{if}^{s})^{b-1}>(p_{il}^{s})^{b-1} \quad (5)$$

For $p_{if}^{s}>p_{il}^{s}$, expression (5) is only valid where *b*>1. Consequently, the FLB is positive (favourites are underestimated) only if *b* in (3) is significantly greater than 1. Similarly it can be shown that when*b*<1 the FLB is negative (favourites are overestimated) and when*b*=1 the market is weak form efficient. We employ a maximum likelihood procedure to estimate *b*in (3) in order to assess the degree of FLB in the aggregate sample.

In order to test the homogeneity of the market across the 22 leagues, we estimate a CL model defined by (2), where $Z_{ij} = \sum_{t=1}^{N} b_t q_t Ln(p_{ij}^s)$. In the expression for Z_{ij}, t represents the league (t=1, 2….22) and q_t is a binary variable which takes the value 1 for the league under scrutiny and 0 for all the other leagues. For instance if game j refers to a game in the Italian league 1, $q_t = 1$ for t=Italian league 1 and $q_t = 0$ for the rest (N-1) of the leagues. In this way, each league derives a unique coefficient for $Ln(p_{ij}^s)$ and this enables us to compare the degree of FLB in the betting markets associated with each of these leagues.

2.4.2.2. Exploring the causes of efficiency differentials

Assessing the role of transaction costs: Various explanations have been presented in the literature for the under-estimation of the chances of favourites in betting markets (for a comprehensive review, see Vaughan Williams, 2005). However, few of these explanations are able to simultaneously explain the absence of FLB in some markets, which is characteristic of the results presented below. A recent theme in the literature which has the potential to address this phenomenon relates to the role of transactions costs, their influence on the degree of informed trader activity and the importance of informed trader activity as a cause of positive FLB (see, for example, Hurley and Mcdonough,1995; Terrell and Farmer, 1996; Bruce et al., 2008). In a similar way, Vaughan Williams and Paton (1998) demonstrated that in the absence of trading costs, longshots are underestimated but as trading costs increase, the bias turns from negative to positive beyond a critical point. Empirical evidence in Vaughan Williams and Paton (1998) and Smith et al. (2006) also supported the view that the level of transaction costs determines the direction and the volume of the bias.

Transaction-costs have, therefore, been shown to explain the inter-market variability of FLB and it could also be argued that they might equally be informative regarding intra-market differences in the degree of FLB. In order to test the hypothesis that *any observed FLB in betting markets is associated with differences in the level of* *transaction costs,* the most straightforward approach would be to set $Z_{ij} = b_1 * Ln(p_{ij}^s) + b_2 c_j$ in (2), (where c_j is the level of transactions costs in game*j*), and to observe the significance and sign of b_2. However, the terms$b_2 c_j$ in the numerator and denominator of equation (2) would cancel each other out because these do not vary across the possible outcomes of a single game. To overcome this problem, we set:

$$Z_{ij} = b_1 * Ln(p_{ij}^s) + b_2 c_j \alpha_{ij} + b_3 c_j \beta_{ij} \qquad (6)$$

where α_{ij} is a binary variable, which takes the value 1 if the *i*th outcome is the most likely outcome of game *j*, 0 otherwise and β_{ij} is a binary variable which takes the value 1 if the *i*th outcome is the least likely outcome of game *j*, 0 otherwise. If the market is efficient then no coefficient in (6) other than b_1 should be significant. However if the hypothesis that positive FLB is associated with higher transaction costs is correct then b_2 should be positive and b_3 negative, so that increases in transaction costs are associated with increases in the probability that favourites win and decreases in the probability that longshots win.

In order to confirm any association between transactions costs and the degree of FLB we also estimate a CL function which incorporates an interaction term between the odds implied probability and the level of trading costs. Consequently, in (2) we set:

$$Z_{ij} = b_3 * Ln(p_{ij}^s) + b_4 c_j Ln(p_{ij}^s) \qquad (7)$$

If the level of transaction costs are not associated with the degree of FLB then b_4 should be insignificant. The advantage of this formulation over that indicated above for assessing the association between the level of transaction costs and the FLBis that the odds implied probabilities of winning of the favourites are accounted for directly. In particular, in the earlier approach, two outcomes which are both favourites (but with different odds implied probabilities, p and αp), will be treated identically in the regression. However, in (7), favouritism is measured on a continuous scale, so that the model probability of outcome i in game j is given by $\exp(\beta \ln p_{ij}) \big/ \sum_{i=1}^{3} \exp(\beta \ln p_{ij}) = p_{ij}^{\beta} \big/ \sum_{i=1}^{3} p_{ij}^{\beta}$, where $\beta = b_3 + b_4 c_{ij}$. Consequently, as β increases, the FLB increases (becomes more positive) and therefore the higher the level of transaction costs, the higher the bias. Consequently, if the bias is caused by an increase in transaction costs, b_4 should be significantly greater than 0.

Assessing the role of league competitiveness in explaining differences in market efficiency.

Section 5 reports the results of the analysis described in the previous section. These results demonstrate that variation in transactions costs does not account for the within-bookmaker variation in FLB observed between the various European leagues. This illustrates the need for an alternative perspective on variation in FLB. Levitt (2004) argued that because odds on soccer matches do not vary significantly following their publication, this may indicate that prices are not adjusted to fully reflect relative demand. In addition, he found evidence that favourites (cf. longshots) are more popular bets. Moreover, Forrest and Simmons (2008), Marshall (2009) and Franck, et al. (2013) argue that competition among bookmakers is likely to lead to more favourable odds for popular bets. As a result, it could be argued that bookmakers offer 'generous' odds for favourites because they wish to stimulate demand and turnover. Consequently, they may be prepared to compromise their margin per game in the expectation that the increased demand may increase the absolute size of their profits. This can explain the fact that the majority of the soccer betting markets exhibit positive FLB. Whilst each of these arguments may contribute to an understanding of the existence of aggregate inefficiency, neither satisfactorily explains cross-market efficiency differences. Therefore, in this section, we focus on differences in endogenous characteristics of the European soccer leagues which have the potential to explain the differential incidence of FLB.

The results presented below demonstrate that, with the exception of Spain, those markets identified as efficient are all associated with games played in lower divisions of leagues. Anecdotal evidence suggests that in these lower divisions there is not such a large difference between the talents of teams which compete (perhaps due to greater similarity between their economic resources than for teams which compete in higher leagues).To the extent that this is true, it is possible that odds setters' may undervalue the difference in competitiveness between different leagues, so that odds are set in a similar way when a 'relatively strong' team faces a 'relatively weak' team irrespective of the league in which the game takes place. Clearly, if there is a significant difference in the competitiveness of games played in certain leagues, the average probability of the relatively weaker team winning will be higher in the more competitive leagues. Since the demand for betting on favourites is higher (Levitt, 2004), leagues that are dominated by a group of strong teams (who are likely to be strong favourites in games they play against weaker opposition) are expected to face more asymmetric demand (cf. that experienced in more competitive leagues). This could potentially lead to the FLB being more pronounced when within-league competition is lower, as bookmakers compete to satisfy the bettors' preferences. Consequently, the following hypothesis is suggested: *There will be a tendency towards more pronounced positive FLB in the betting markets associated with the less competitive leagues.*

In order to test this hypothesis we develop a proxy for a league's competitiveness, based on the average of the absolute value of the goal differences of the teams in this league (for a discussion of league competitiveness see Koning and Markidakis, 1999). More specifically, for each team the mean absolute goal difference per game is determined for a given season. The mean of this value for all teams in a given league in a given season is then calculated. This is used as a proxy for the degree of competition for a given league in a given season; it is assumed that leagues with lower means are more competitive, other things being equal. Averages by league through all seasons then provides a proxy for the overall level of competition that each league exhibited in the sample period; we refer to this measure as the league's overall mean absolute goal difference per game (mean AGD). We examine the correlation between the mean AGD for a given league and a measure of that league's degree of FLB. This latter measure is given by the coefficient of the index function, Z_{ij} in a CL model with the form of (2), where $Z_{ij} = \sum_{t=1}^{N} b_t q_t Ln(p_{ij}^s)$. In the expression for Z_{ij}, t represents the league (t=1, 2….22) and q_t is a binary variable which takes the value 1 for the league under scrutiny and 0 for all the other leagues. The coefficients of Z_{ij} for each league, resulting from estimating the CL model over all 52, 865 games across the 22 leagues from 1999 to 2008, are given in Table 2.3.

In order to ensure that the results relating competitiveness of a league to the efficiency of its associated betting market are robust, we also employ an alternative measure of competitiveness. This measure assumes that if a league is highly competitive, results are likely to be less predictable and, therefore, bookmakers' odds are less likely to fully account for the true winning probabilities of each team. Consequently, we use the proportion of variability in the actual outcome that can be explained by the probability indicated by the odds as an alternative proxy for the league's level of competitiveness. This in turn is determined, for each league, by the McFadden R^2 statistic associated with the estimated CL function, with natural log of the odds implied probability as the sole predictor. The McFadden R^2 statistic compares the log-likelihood of a model with no predictors (M_{no}) and the model containing the predictors (M_{pred}): $R^2_{McF} = 1 - (LL(M_{pred}) / LL(M_{no}))$ .The McFadden R^2 statistic is preferred as a measure of competitiveness to the log-likelihood arising from the estimation of the relevant CL model, because it is not dependent on the number of observations.

2.5. Results

2.5.1. Variations in returns across odds categories

Table 2.1 details the mean profit for unit stake bets across all odds categories for betting markets associated with the 52,865 soccer games in the aggregate sample. The expected profit generally decreases as the odds increase indicating the existence of positive FLB. The over-round is on average about 0.11, and the results displayed in table 2.1 show that betting at odds lower than 2.4 generally provide higher than expected returns, whereas betting at high odds (especially those greater than 3.5) results in returns much lower than the expected level.

The aggregated results, however, mask substantial heterogeneity in the magnitude and direction of the FLB between betting markets associated with different leagues. This is illustrated, for example, by considering the mean profit and the ratio of actual to expected returns for just two leagues, the Spanish second division and the Italian first division (see Table 2.2).

Table 2.1: A comparison of mean profit and actual/expected returns across odds categories for betting markets associated with 52,865 soccer games across 22 European soccer leagues between 1999 and 2008.

Decimal odds category	Mean profit per bet	Actual/expected return
≤ 1.5	-0.034	1.085
>1.5, ≤ 1.8	-0.072	1.043
>1.8, ≤ 2.0	-0.069	1.046
>2.0, ≤ 2.4	-0.081	1.033
>2.4, ≤ 2.8	-0.111	0.999
>2.8, ≤ 3.1	-0.118	0.991
>3.1, ≤ 3.2	-0.134	0.973
>3.2, ≤ 3.3	-0.108	1.002
>3.3, ≤ 3.5	-0.112	0.998
>3.5, ≤ 4.0	-0.161	0.943
>4.0, ≤ 5.0	-0.157	0.947
>5.0	-0.28	0.809

Table 2.2: A comparison of mean profit and actual/expected returns across odds categories for Spanish league II and Italian League I betting markets between 1999 and 2008.

	Spanish League II N=3619		Italian League I N=2714	
Decimal odds category	Mean profit per bet	Actual/expected return	Mean profit per bet	Actual/expected return
$\leq$1.5	-0.132	0.975	0.02	1.146
>1.5, $\leq$1.8	-0.134	0.973	-0.065	1.051
>1.8, $\leq$2.0	-0.121	0.988	-0.082	1.031
>2.0, $\leq$2.4	-0.109	1.001	-0.06	1.056
>2.4, $\leq$2.8	-0.11	1.000	-0.084	1.029
>2.8, $\leq$3.1	-0.067	1.048	-0.075	1.039
>3.1, $\leq$3.2	-0.173	0.929	-0.175	0.927
>3.2, $\leq$3.3	-0.04	1.079	-0.32	0.764
>3.3, $\leq$3.5	0.003	1.127	-0.386	0.690
>3.5, $\leq$4.0	-0.132	0.975	0.02	1.146
>4.0, $\leq$5.0	-0.134	0.973	-0.065	1.051
>5.0	-0.121	0.988	-0.082	1.031

It appears from the results displayed in table 2.2 that profits and the ratio of actual to expected returns steadily diminish as odds increase in betting markets associated with the Italian first division, whilst profits and the ratio of actual to expected returns generally increase as odds increase for the Spanish second division. This preliminary analysis suggests that distinguishing between different divisions/leagues may provide useful evidence concerning the heterogeneity of FLB in the European soccer betting market.

2.5.2. Modelling winning probabilities to assess FLB

Whilst the results presented in tables 2.1 and 2.2 are suggestive of FLB in the aggregate European football betting market, a more formal test is needed to assess the significance of the evidence. Consequently, we estimate the CL function given by equation (3) using data from betting markets associated with all the 52,865 games played between 1999 and 2008 in the 22 leagues across 10 European countries. The coefficient (*b*) of the odds implied probabilities is highly significant in this model (*b*=1.141, Std. Error = 0.013, Z=92.82) suggesting that the odds incorporate a significant amount of information concerning the probability of each outcome of a game. In addition, the 95% confidence interval for *b* (1.116-1.165) indicates that this is significantly greater than 1. This supports the conclusion that if the European soccer betting market is considered as a single unit it exhibits positive FLB. As a consequence, betting on favourites across European leagues will achieve higher returns compared to betting longshots.

In order to explore whether the degree and nature of the FLB is consistent across the 22 leagues, we estimate, as indicated above, a CL model with the form of (2), but where $Z_{ij} = \sum_{t=1}^{N} b_t q_t Ln(p_{ij}^s)$. In this way, each league derives a unique coefficient for $Ln(p_{ij}^s)$. The results of maximum likelihood estimation of this model are presented in table 3. The 95% confidence intervals for the betting markets associated with most of the leagues indicate that they display a positive FLB. However, for9 of the 22 leagues the results suggest that the efficient market hypothesis cannot be rejected, as the corresponding coefficients of the interaction term between the natural logarithm of the odds implied probability and a binary, league specific variable, are not significantly different from 1. In three of these leagues the interaction term coefficient is less than 1 (Spanish league II, English league III and Scottish league I), but these coefficients are not sufficiently less than 1 to conclude that a negative FLB exists in the associated betting markets. Overall, these results indicate that there are significant differences in the nature and degree of the FLB in betting markets associated with different European soccer leagues as indicated by the confidence intervals of the coefficients.

Table 2.3: Results of estimating a CL model with an index function made up of interaction terms between natural log of the odds implied probabilities of match outcomes with binary variables which indicate the league in which the particular game is played, employing the 52,865 soccer games across 22 European soccer leagues played between 1999 and 2008. (* *Coefficient greater than 1 with 95% confidence*)

Interaction terms[1]: binary variable identifying league x $Ln(p_{ij}^{s})$	Coefficient	Std. Error	z-value	P>z	95% Conf. Interval
Spain2	0.936	0.064	14.70	0.000	0.811-1.061
England3	0.989	0.053	18.71	0.000	0.885-1.093
Scotland1	0.999	0.082	12.12	0.000	0.837-1.161
England1	1.019	0.049	20.61	0.000	0.922-1.116
England2	1.029	0.050	20.51	0.000	0.931-1.127
Spain1	1.044	0.050	20.95	0.000	0.946-1.141
Scotland2	1.044	0.085	12.20	0.000	0.877-1.211
Germany2	1.103	0.067	16.45	0.000	0.971-1.234
Germany1	1.117*	0.055	20.13	0.000	1.008-1.225
England Conf.	1.125	0.083	13.52	0.000	0.962-1.288
France1	1.128*	0.059	19.10	0.000	1.012-1.244
Portugal1	1.136*	0.058	19.45	0.000	1.021-1.250
Turkey1	1.170*	0.057	20.66	0.000	1.059-1.281
England Prem.	1.191*	0.048	24.78	0.000	1.097-1.286
Scotland Prem.	1.194*	0.057	21.10	0.000	1.083-1.305
France2	1.204*	0.066	18.11	0.000	1.074-1.334
Belgium1	1.208*	0.054	22.58	0.000	1.104-1.313
Netherlands1	1.247*	0.051	24.66	0.000	1.148-1.346
Italy1	1.253*	0.052	24.27	0.000	1.152-1.354
Italy2	1.270*	0.060	21.06	0.000	1.152-1.389
Greece1	1.300*	0.057	22.30	0.000	1.167-1.392
Scotland3	1.330*	0.075	17.78	0.000	1.181-1.473

[1.] *Country N: Soccer league N in that country x* $Ln(p_{ij}^{s})$

It is clear from table 2.3 that most of the European leagues investigated exhibit significant degrees of positive FLB and it could be that this is the prevailing norm, those which display no FLB simply arising by chance. This view is tested using a sign test. The test statistic, which corresponds to a value of the standard normal distribution (N(0,1)), is calculated as follows: Z* = $\frac{k-0.5-0.5*n}{0.5*\sqrt{n}}$, where, k is the number of leagues where the FLB is significantly positive and n is the number of leagues under investigation (22). We find Z*=0.64, implying that the hypothesis that a betting market associated with a randomly selected European soccer league is efficient cannot be rejected ($p < 0.1$). Consequently, we conclude that even though most markets exhibit significant positive FLB, market efficiency should not be considered an uncommon situation when investigating one particular league (or a small number of leagues), rather than a cross-league sample.

We explore whether there is a clear distinction in terms of efficiency between those betting markets associated with leagues where the coefficient of the interaction term is significantly greater than 1 and those where it is not. This is achieved by estimating a CL model with the form of (2), with $Z_{ij} = \sum_{r=1}^{2} b_r h_r Ln(p_{ij}^{s})$, where h_r is a dummy variable, such that h_1 takes the value 1 if the betting market associated with a given league displays a significant positive FLB (from the results displayed in table 3), and 0 otherwise; h_2 takes the value 1 if the betting market associated with a given league displays no significant FLB (from the results displayed in table 3), and 0 otherwise. The results of estimating these CL models indicate that there is a clear distinction between the betting markets associated with leagues which display a positive FLB in table 2.3 and those that do not. In particular, the coefficient b_1 is estimated to be 1.2020 (Std. error= 0.0158, Z= 76.16, 95% confidence interval: 1.1710-1.2323), indicating a significant positive FLB, whereas the coefficient b_2 is estimated to be 1.0233 (Std. error= 0.02083, Z= 49.09, 95% confidence interval: 0.9815-1.0631), suggesting that these markets are weak form efficient. These results are confirmed by a likelihood ratio test which compares the amount of information concerning winning probabilities contained in (a) probabilities derived from the CL function given by equation (3) (which is estimated using data from betting markets associated with all the 52,865 games played between 1999 and 2008, combined; no distinction made between the leagues in which the games are played), from (b) probabilities derived from the CL model outlined above, with the form of (2), with $Z_{ij} = \sum_{r=1}^{2} b_r h_r Ln(p_{ij}^s)$; which is again estimated using data from betting markets associated with all the 52,865 games played between 1999 and 2008, but where a distinction is made between games played in leagues whose associated betting markets display a significant positive FLB from those that appear weak form efficient. The resulting likelihood ratio is 47.18, which is significant at the 1% level ($\chi_1^2(.01) = 6.64$). This result confirms that there is a clear distinction between betting markets associated with leagues which display a positive FLB and those which do not.

In summary, two conclusions emerge from the analysis presented above. First, the traditional (positive) form of FLB is dominant in the European soccer betting market if it is considered as a single entity. The results offer sufficient evidence to reject the efficient market hypothesis and conclude that overall, the winning probabilities of favourites are underestimated in the European soccer betting market. However, when this market is sub-divided into those betting markets associated with each of the constituent 22leagues, we conclude that odds in betting markets associated with nine divisions *are* efficient predictors of match results. The odds in these betting markets appear to be significantly less biased than those for the 13 betting markets associated with the remaining leagues. Taken together, these results lead us to reject the hypothesis that the fixed odds soccer betting market across Europe is weak form efficient and homogeneous. Consequently, whilst the European soccer betting market as a whole may be inefficient, this would not necessarily be discerned by examining one or even a limited sample of betting markets associated with particular European soccer leagues. This is an important conclusion given the nature of the limited samples employed in some previous studies.

2.5.3. Investigating the roots of cross-market differences in efficiency

Previous evidence concerning the role played by transactions costs in influencing the FLB leads us to explore the extent to which the differences we have observed between weak form efficient betting markets and those which display marked positive FLB can be explained by differences in transaction costs. This is achieved by estimating a CL model with Z defined by equation (6), using data from all the 52,865 soccer games in the aggregate sample. The results, displayed in table 2.4, demonstrate that the signs of the interaction terms between over-round for the betting market associated with a particular game (c_j) and the binary variables which capture whether the outcome is the most or least likely outcome (α_{ij} and β_{ij} respectively) correspond with the view that transactions costs influence the FLB; that is the sign of these coefficients suggest that the probability of the most (least) likely outcome occurring increases (decreases) as transactions costs increase. However, neither of these coefficients is significant at the 5% level, casting doubt on the significance of transactions costs as an influential factor in the determination of the direction and the strength of the FLB.

Table 2.4: Results of estimating a CL model with an index function made up of the following terms: the natural log of the odds implied probabilities of match outcomes and two interaction terms between the over-round (*c*) and binary variables which, take the value 1 if the *i*th outcome is the most likely (α_{ij}) or least likely (β_{ij}) outcome of game *j*, 0 otherwise (see (6)).

Predictor	Coefficient	Std. Error	z-value	*P*>z	95% Conf. Interval
Lnp	1.1584	0.0190	61.07	0	1.1212-1.1956
α_{ij} *c	0.0051	0.0157	0.32	0.75	-0.0257-0.0358
β_{ij} *c	-0.0147	0.0095	-1.56	0.12	-0.0333-0.0038

In order to explore further the association between transactions costs and the degree of FLB we estimate a CL function which incorporates an interaction term between the natural logarithm of odds implied probability and the level of transactions costs; that is where the index function is given by (7).

The results of estimating this CL function indicate that neither the coefficient for the natural logarithm of odds-implied probability nor the coefficient of the interaction term between transactions costs (c_j) and the natural log of odds implied probabilities are significantly different from zero at the 5% level (Coef.of $Ln(p_{ij}^s)$ = 0.0116, Std. error = 0.8130, z-value=0.01; Coef. of $c_j Ln(p_{ij}^s)$ =1.0125, Std. error = 0.7309, z-value= 1.39). A comparison of these results with those obtained from estimating a CL model with $Ln(p_{ij}^s)$ as the sole predictor variable indicates that the significance of the natural logarithm of odds-implied probability declines markedly when the interaction term is included as an additional predictor variable. This suggests that, due to colinearity, the standard errors of the two predictors increase significantly and the coefficients of the two variables are not robustly estimated. Therefore, interpreting the results based on the coefficients of the variables and the corresponding significance statistics will lead to misleading conclusions.

In order to overcome this difficulty, we explore whether a model which *does not* distinguish between weak form efficient and inefficient markets, but accounts for transactions costs, incorporates as much information concerning winning probabilities as one which *does* distinguish between efficient and inefficient markets and, in addition, accounts for transactions costs. Clearly, if differences in transaction costs can explain differences between betting markets which do and do not display marked positive FLB then there should be no difference in the information content of these two models. Consequently, we compare the log likelihood of a CL model with an index function incorporating $Ln(p^s_{ij})$ and $c_j Ln(p^s_{ij})$ (estimation results given above; log likelihood = -53392.6) with the log likelihood of a CL model of form (2), with $Z_{ij} = c_j Ln(p^s_{ij}) + \sum_{r=1}^{2} b_r e_r Ln(p^s_{ij})$, where e_r is a dummy variable:e_1 takes the value 1 if the betting market associated with a given league displays no significant FLB (from the results displayed in table 2.3), and 0 otherwise;e_2 takes the value 1 if the betting market associated with a given league displays a significant positive FLB (from the results displayed in table 2.3), and 0 otherwise. The results of estimating this latter model are given in table 2.5. Once again, the individual coefficients and significance statistics reported in table 5 are misleading because the standard error of the coefficients increases due to colinearity. However, this does not influence the overall model log-likelihood.

Table 2.5: Results of estimating a CL model with an index function made up of the following terms: an interaction term between the transaction costs in a given game and the natural log of the odds implied probabilities of match outcomes and two interaction terms between the natural log of odds implied probabilities and binary variables which account for whether a particular market is weak form efficient (e_1) or displays positive FLB(e_2).

Predictor	Coefficient	Std. Error	z-value	*P*>z	95% Conf. Interval
$c_j Ln(p_{ij}^s)$	0.7978	0.7357	1.08	0.28	-0.6442-2.2398
$e_1 Ln(p_{ij}^s)$	0.1358	0.8178	0.17	0.87	-1.4670-1.7386
$e_2 Ln(p_{ij}^s)$	0.3142	0.8188	0.38	0.70	-1.2907-1.9190

Model log-likelihood= -53369.4

A likelihood ratio test confirms that the log-likelihood of the CL model which accounts for the efficiency of the betting market associated with the league in which the game is played is significantly larger than the log-likelihood of the CL model which does not account for this factor (LL= -53379.4 and -53392.6, respectively; Likelihood ratio= 46.43, $\chi_2^2(.01) = 9.22$).This result confirms that even when transactions costs are fully discounted in the CL model (as they are in both these models) there is still a significant difference between those markets where positive FLB is detected and those which exhibit no FLB. If transactions costs account for differences in the degree of FLB then there should not be a significant distinction between efficient and inefficient markets where transactions costs are considered. Consequently, we reject the hypothesis that transactions costs account for the differences in FLB observed in the betting markets examined here.

We now turn to examining another potential explanation for the variation in bias across different markets, a league's competitiveness. As indicated in the Methods section above, we use a league's overall mean absolute goal difference per game (mean AGD) as a proxy for its degree of competitiveness. The calculated values for mean AGD for each of the 22 European leagues are shown in table 2.6. We explore the degree of correlation between a league's competitiveness and the degree of FLB in this league (measured by the relevant coefficient relating to that league shown in table 2.3). The resultant correlation coefficient (r) is 0.627, which is significant at the 1% level (t = 3.6, where $t = r\sqrt{[(n-2) / (1 - r^2)}$, n =22).

Table 2.6: Coefficients of interaction terms between natural log of the odds implied probabilities of match outcomes with binary variables which indicate the league in which the particular game is played, in CL functions (estimated employing the 52,865 soccer games across 22 European soccer leagues played between 1999 and 2008), together with associated McFadden R^2 values and the mean absolute goal difference for the respective leagues

League	Coefficient identifying degree of FLB	McFadden R^2	Mean absolute goal difference
Spain2	0.936	0.0305	0.2034
England3	0.989	0.0417	0.2740
Scotland1	0.999	0.0522	0.3540
England1	1.019	0.0475	0.2722
England2	1.029	0.0474	0.2816
Spain1	1.044	0.0767	0.3074
Scotland2	1.090	0.0571	0.3357
Germany2	1.103	0.065	0.3204
Germany1	1.117	0.0886	0.3457
England Conf.	1.125	0.0557	0.2325
France1	1.128	0.0653	0.2642
Portugal1	1.136	0.1040	0.3564
Turkey1	1.170	0.1198	0.3935
England Prem.	1.191	0.1136	0.3704
Scotland Prem.	1.194	0.1500	0.4722
France2	1.204	0.0621	0.2387
Belgium1	1.208	0.1193	0.4887
Netherlands1	1.247	0.1430	0.4837
Italy1	1.253	0.1276	0.3732
Italy2	1.27	0.0927	0.2612
Scotland3	1.314	0.1226	0.5770
Greece1	1.344	0.1873	0.4313

We explain above that we also employ, as a proxy for league competitiveness, the McFadden R^2 statistic associated with the league's estimated CL function, with natural logarithm of the odds-implied probability as the sole predictor. The results of these CL estimations are given in table 3 and the resulting coefficients for the natural log of odds implied probabilities are related to the associated McFadden R^2 statistic in table 2.6. From table 2.6 it appears that a strong association exists between the McFadden R^2 statistics and the magnitude of the coefficient of the natural log of odds implied probabilities in the CL models developed for betting markets for corresponding leagues. For example, efficient markets like the Spanish, Scottish and English second divisions and the English third division are associated with low R^2 values, whereas leagues exhibiting significant FLB, such as Greece's first division, are associated with high R^2 values. The correlation coefficient between the McFadden R^2 statistic and the coefficient of the natural logarithm of odds-implied probabilities in the CL model for the corresponding league is 0.842, which is significant at the 1% level (t = 6.98).

In summary, FLB is negatively correlated with both the measures of league competitiveness employed here; greater positive FLB occurs in the least competitive leagues, whereas there is no FLB in the most competitive leagues. These results support our hypothesis that there is a tendency towards more pronounced positive FLB in the betting markets associated with less competitive leagues. This suggests that bookmakers' prices do not account sufficiently for the competitive differences between leagues and, as a result, the chances of weaker teams in less competitive leagues are over-estimated.

2.6. Discussion

Analysing data from twenty-two divisions across eleven European national soccer leagues, this study concludes that weak form inefficiencies, as evidenced by the FLB, constitute a significant feature of soccer betting markets in Europe. However, a considerable degree of heterogeneity in the degree of efficiency between betting markets associated with different leagues, is observed. The positive form of the FLB dominates, in line with the findings of the majority of earlier efficiency studies across a range of betting markets. However, the degree of inefficiency is variable across markets associated with different leagues and absent in several. Consequently, in terms of Fama's perspective on market inefficiency, the homogeneity condition is clearly not satisfied.

In probing the causes of the differential incidence of bias, the role of transactions costs, a significant factor in earlier accounts of inter-market FLB variance, is investigated. However, the results suggest that this is not an explanatory variable of intra-market FLB variance. An alternative source of potential influence, the variable degrees of competitiveness or predictability across leagues, is considered. This analysis reveals a significant and positive correlation between competitiveness and the degree of FLB, as it seems that bookmakers have to pay a higher price in order to stimulate demand in less competitive leagues.

The main contributions of this study, therefore, lie in its identification of heterogeneity across European soccer betting markets in terms of degrees of weak form efficiency, its rejection of variations in transactions costs as a possible explanation for intra-market, as opposed to inter-market variation, and the identification of competitiveness or predictability as a more promising explanation for these differences in market efficiency. Clearly, further evidence is required to investigate whether the degree of competitiveness or 'predictability' in betting events influences the degree of FLB in contexts other than those examined here. However, this study represents a useful first step in this direction, as well as offering a contemporary perspective on the efficiency characteristics of soccer betting markets across Europe.

Endnote

[1] Results for betting markets associated with other leagues available on request.

References

Ali, M. M. (1977). Probability and utility estimates for racetrack bettors. *The Journal of Political Economy* 85, 803-815.

Asch, P., Malkiel, B.G., and Quandt, R. E. (1982). Racetrack betting and informed behaviour. *Journal of Financial Economics*, 10, 187-194.

Bird, R., and McCrae, M. (1994). The efficiency of the racetrack betting market: Australian evidence. In D.B. Hausch, V.S.Y. Lo, and W.T. Ziemba (Eds.), *Efficiency of Racetrack Betting Markets* (pp. 575-581). London: Academic Press.

Bruce, A. C., and Johnson, J. E. V. (2000). Investigating the roots of the favourite-longshot bias: An analysis of decision making by supply- and demand-side agents in parallel betting markets. *Journal of Behavioral Decision Making,* 13, 413-430.

Bruce, A. C., Johnson, J.E. V., Peirson, J., and Yu, J. (2008). An examination of the determinants of biased behaviour in a market for state contingent claims. *Economica*, 76(302), 282-303.

Busche, K., and Hall, C. D. (1988). An exception to the risk preference anomaly. *Journal of Business*, 61, 337-46.

Busche, K. (1994). Efficient market results in an Asian setting. In D.B. Hausch, V.S.Y. Lo, and W.T. Ziemba (Eds.), *Efficiency of Racetrack Betting Markets* (pp. 615-16). London: Academic Press.

Cain M., Law, D., and Peel, D. (2003). The favourite-longshot bias, bookmaker margins and insider trading in a variety of betting markets. *Bulletin of Economic Research,* 55, 263-73.

Colquitt, L. L., Godwin, N. H., and Swidle, S. (2004). Betting on long shots in NCAA basketball games and implications for skew loving behaviour. *Finance Research Letters,* 1, 119-126.

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

Dare, W. H., and McDonald, S. S. (1996). A generalized model for testing the home and favorite team advantage in point spread markets. *Journal of Financial Economics*, 40, 295-318.

Dixon, M. J., and Pope, P. F. (2004). The value of statistical forecasts in the UK association football betting market. *International Journal of Forecasting,* 20, 697-711.

Dowie, J. (1976). On the efficiency and equity of betting markets. *Economica*, 43, 139-150.

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

Fama, E. F. (1998). Market efficiency, long-term returns, and behavioral finance. *Journal of Financial Economics*, 49*,* 283-306.

Forrest, D., Goddard, J., and Simmons, R. (2005). Odds-setters as forecasters: The case of English football. *International Journal of Forecasting*, 21, 551-64.

Forrest, D., and McHale, I. (2007). Anyone for tennis (betting)? *The European Journal of Finance*, 13, 751-768.

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

Franck, E., Verbeek, E., and S. Nüesch, E. (2013). Inter-Market Arbitrage In Betting. *Economica*, 80, 300-325.

Gander, J. M., Zuber R. A., and Johnson, R. S. (2001). Searching for the favourite-longshot bias down under: An examination of the New Zealand pari-mutuel betting market. *Applied Economics*, 33, 1621-29.

Golec, J., and Tamarkin, M. (1991). The degree of inefficiency in the football betting market. *Journal of Financial Economics*, 30, 311-23.

Graham, I., and Stott, H. (2008). Predicting bookmaker odds and efficiency for UK football. *Applied Economics*, 40, 99-109.

Hurley, W., and McDonough, L. (1995). A Note on the Hayek Hypothesis and the Favorite Long Shot Bias in Parimutuel Betting. American Economic Review, 85(4), 949-955.

Kuypers, T. (2000). Information and efficiency: an empirical study of a fixed odds betting market. *Applied Economics,* 32, 1353-1363.

Koning, R. H., and Markidakis, S. (1999). Balance in competition in Dutch Soccer. *The Statistician, Journal of the Royal Statistical Society: Series D*, 49(3), 419-43.

Law, D., and Peel, D. A. (2002). Insider trading, herding behaviour and market plungers in the British horse-race betting market. *Economica*, 69, 327-38.

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

Malaric, R., Katic, T., and Sabolic, D. (2008). The market efficiency of the soccer fixed odds internet betting market. *Applied Economics Letters*, 15, 171-174.

Marshall, B. R. (2009). How quickly is temporary market inefficiency removed? *The Quarterly Review of Economics and Finance*, 49(3), 917-930.

Metrick, A. (1996). March madness? Strategic behaviour in NCAA basketball tournament betting pools. *Journal of Economic Behavior and Organization*, 30, 159-172.

Paul, R. J., Weinbach, A. P., and Weinbach J. (2003). Fair bets and profitability in college football gambling. *Journal of Economics and Finance*, 27, 236-242.

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

Schnytzer, A., and Shilony, Y. (1995). Inside information in a betting market. *The Economic Journal*, 105, 963-71.

Shin, H. S. (1993). Measuring the incidence of insider trading in a market for state-contingent claims. *The Economic Journal*, 103, 1141-53.

Smith, M. A., Paton, D., and Vaughan Williams, L. (2006). Market efficiency in person-to-person betting. *Economica*, 73, 673-689.

Snowberg, E., and Wolfers, J. (2010). Explaining the favourite-longshot bias: Is it risk-love, or misperceptions? *Journal of Political Economy*, 118(4), 723-746.

Snyder, W. W. (1978). Horse racing: Testing the efficient markets model. *The Journal of Finance*, XXXIII, 1109-18.

Terrell, D., and Farmer, A. (1996). Optimal betting and efficiency in parimutuel betting markets with information costs. *The Economic Journal*, 106, 846-868.

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

Tryfos, P., Casey, S., Cook, S., Leger, G., and Pylypiak, B. (1984). The profitability of wagering on NFL games. *Management Science*, 30, 123-132.

Tuckwell, R. H. (1983). The thoroughbred gambling market: Efficiency, equity and related issues. *Australian Economic Papers*, 22(40), 106-118.

Urrutia, J. L. (1995). Tests of random walk and market efficiency for Latin American equity markets. *Journal of Financial* Research, 18, 299-309.

Vaughan Williams, L. (Ed.) (2005). *Information Efficiency in Financial and Betting Markets.* Cambridge: Cambridge University Press.

Vaughan Williams, L., and Paton, D. (1997). Why is there a favourite-longshot bias in British racetrack betting markets? *The Economic Journal*, 107, 150-158.

Vaughan Williams, L., and Paton, D. (1998). Why are some favourite-longshot biases positive and others negative? *Applied Economics*, 30, 1505-1510.

Vergin, R. C., and Scriabin, M. (1978). Winning strategies for wagering on National Football League games. *Management Science*, 24, 809-818.

Woodland, L. M., and Woodland B. M. (1994). Market efficiency and the favourite long shot bias: The baseball betting market. *Journal of Finance*, 49, 269-80.

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

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

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

Chapter 3

Bettors *vs*. Bookmakers: 1-0! Examining the Origins of Information in Football Betting Markets

Abstract

This paper examines the sources of information in football betting markets. We identify a clear distinction between two types of bookmakers in these markets, namely book-balancing and position-taking bookmakers and argue that these differences hold the key to the important sources of information within this market. We argue that book-balancing bookmakers act in the same manner to market makers in financial markets, effectively matching buyers with sellers by adjusting their odds according to the amounts traded on different game outcomes. As a result, their profit is a function of the generated turnover and their prices reflect a volume-weighted average of the public's opinion, potentially dominated by 'smart-money'. Position-taking bookmakers on the other hand, attempt to maximize their profit margin rather than minimize exposure against a large customer base, deliberately filtered to avoid 'skilled bettors'. We analyse a unique longitudinal dataset relating to the odds posted by a leading book-balancing and a leading position-taking bookmaker for 2,132 games. For each game we collected the odds offered on the potential outcomes at nine separate points. The data allows us to examine the movements through time of the odds posted by the two different types of bookmakers and we are able to show that book-balancers move their odds relatively often. The results indicate that book-balancers are relatively proactive in their price changes and the position-takers are reactive. This finding contradicts the generally espoused proposition that bookmakers are superior over bettors in predicting event outcomes, since the position-takers are the ones who follow trends in the book-balancers' (demand driven) prices, rather than vice-versa. Finally, we show that this dynamic transmission of information from bettors to bookmakers improves the forecasting ability of market odds. This finding is consistent with the efficient market hypothesis.

3.1. Introduction

The deductive approach implies that "the researcher on the basis of what is known about in a particular domain and of theoretical considerations in relation to that domain, deduces a hypothesis (or hypotheses) that must then be subjected to empirical scrutiny" (Bryman and Bell, 2003, pp. 9-10). From an inductive stance, "theory is the outcome of research. In other words, the process of induction involves drawing generalisable inferences out of observations" (Bryman and Bell, 2003, p. 12). In the case of betting markets, inductive reasoning can be applied when attempting to explain patterns observed in market prices. However, in exploring the behaviour of bookmakers, we employ the deductive approach. In particular, we draw from the literature and identify two contrasting models of bookmaking that have been suggested in the previous studies. In the real world, where a large number of bookmakers operate, it is very likely that the two business models will coexist. Hence, we provide a theoretical framework regarding the simultaneous operation of the two types of businesses and the interactions between them. In addition, the theoretical framework accounts for interactions between the bookmakers and the betting public, informed and uninformed. We then analyse the observed patterns in market prices in order to test the proposed theory. Hence, we employ deductive reasoning in using market data as empirical evidence to validate the theoretical prototype concerning the flow of money in the market.

Previous papers examining betting behaviour tend to simplify the structure of betting markets, assuming they are places where two homogeneous entities interact: (i) bookmakers, who are regarded as motivated to maximize profit or to earn a risk-free return by trading and (ii) bettors, who are generally regarded as less-informed than bookmakers, seeking to gamble for pleasure or to maximize their well-defined utility functions. This over-simplification of the structure of betting markets may have led to some misunderstanding of the nature of prices and price movements in these markets. For example, in a recent influential publication (Levitt, 2004) it was suggested that bookmakers are better at predicting game outcomes than the typical bettor. It is the aim of this paper to examine how a more accurate portrayal of the nature of the modern bookmaker market can lead us to a better understanding of the nature of prices and price movements in these markets and as a result to lead to very different conclusion from Levitt (2004).

The rapid expansion of online betting has led to a fundamental shift in the entire gambling industry. Modern betting takes place in a globalized setting, where billions of dollars are traded on a weekly basis (Forrest, 2006, 2012) between a heterogeneous population of bettors and bookmakers. This rapid expansion has coincided with a situation in which prices offered by bookmakers are increasingly competitive,[9] which has in turn attracted sophisticated bettors seeking to take advantage of new 'investment' opportunities. Due to the ease, and relatively low-cost of shifting their capital between bookmakers in the online setting, high-stakes bettors, who are likely to be price sensitive, have the opportunity to seek out bookmakers who will provide sufficient liquidity to accommodate their large trades.

It is the contention of this paper that the bookmakers who are willing to accept large stakes (who are almost exclusively online-based) will typically employ a relatively low-overhead, low-margin, high-turnover strategy. In order to do so, it is argued that they actively manage their book to ensure that liabilities across outcomes for a particular event are relatively equalized. This will ensure that their profits on a given event will roughly equal the transaction costs which they incorporate into their odds, so that profits are assured regardless of the outcome of the event. We refer to this strategy as 'book-balancing' throughout this paper.

It is argued here that there also exist in the market a different category of bookmakers, which are termed 'position-takers' here. These generally have longstanding reputations and they largely cater for recreational (less informed) bettors.

[9] Forrest (2012) notes a decrease in the over-round from 11.13% offered by the major U.K. bookmaker Ladbrokes on English Premier League matches in 2000-01 to 6.1% in 2010-11, which could have been decreased even further by the bettor 'shopping-around' for the best odds.

These bookmakers include those that provide physical-world betting services (i.e., betting shops). They actively encourage new accounts with small account opening bonuses and advertise low-probability (and low-liquidity) bets, such as accumulators and other exotics prominently. The greater physical-world presence (and overheads associated with operations such as betting shops) means that these bookmakers do not have the luxury of operating with the low transaction costs (over-rounds) of their exclusively online competitors, and thus tend to operate relatively low-turnover, high-margin strategies. With a target market of unsophisticated clientele, and relatively high margins, these 'established bookmakers' will have less incentive to change their odds frequently than the book-balancing bookmakers and they have less incentive to maintain a balanced book. This arises because they believe that they generally have superior information to their customers, which enables them to set odds in such a way as to enable them to maximize profits over time (and these odds are generally associated with high margins. It is argued here that these established bookmakers will change prices in response to public information (when they believe this represents genuine information which they have not incorporated into their odds), and to avoid excessive imbalance in their book (perhaps to avoid being on one side of a synthetic Dutch book). However, due to the larger 'cushioning' effect of their high over-rounds it is expected that their odds changes will be less frequent than those of the book-balancing bookmakers. To achieve our aims we compare the evolution and efficiency of prices in European football betting markets of a major Asian bookmaker that we identify as a 'book-balancer' with those of a major U.K.-based bookmaker that we identify as a 'position-taker'. A unique data set is employed to examine differences in the nature of the prices and the evolution of prices of these two bookmakers. This incorporates odds collected at nine points in time in the 24 hours leading to the kickoff of 2,132 matches in the 2012-13 seasons of six major European leagues: English Premier League, Spanish La Liga, Italian Serie A, German Bundesliga, French Ligue 1, and Dutch Eredivisie. We find that the odds of the book-balancing bookmaker are different at point *t* compared to its odds at point *t-1* on 77% of occasions; however, the odds of position-takers only differ from the preceding time period on 12.4% of occasions. This result suggests a clear difference in their approach to odds setting.

Using a random-effects model, we show that lagged odds changes at the book-balancer are significant predictors of the odds changes at the position-taker bookmaker, but not vice-versa. A reduced-form model shows that the changes in odds at the position-taker can be attributed, in large part, to the lagged differences in odds between the two types of bookmaker. As a result, the closing odds of the position-taker are shown to be a function of the day-ahead odds of the book-balancer, but the day-ahead odds of the position-taker do not significantly affect the book-balancer's closing odds. This is consistent with the trades of sophisticated bettors moving prices at the book-balancer and this information diffusing to the position-taker. In addition, using a conditional logit model, the odds of the book-balancer (cf. the position-taker) are shown to better forecast match outcome compared to those of the position-taker and the closing prices (observed 1 second before kickoff) are more efficient predictors of actual match outcomes than these observed 24 hours prior to kickoff, for both types of bookmaker. This is consistent with the betting behaviour of bettors revealing new information about game outcomes over time. These results lead us to conclude that the weight of money from sophisticated bettors in the low-margin, high-turnover markets informs prices in the high-margin, low-turnover position-taking market, which is mainly populated by uninformed traders.

The remainder of this paper is structured as follows: Section 2 includes a review of the theory and empirical evidence regarding the behaviour of bookmakers and bettors. This review is then employed to develop the hypotheses in Section 3. The details of the unique data set employed, a detailed discussion of nature of the specific bookmakers whose odds are examined in this study, and methodology are presented in Section 4. Results are presented in Section 5. A discussion of the implications of the results and suggestions for future research are provided in Section 6.

3.2. Literature Review

3.2.1 Bookmakers – Theory and Evidence

Position-Taking Bookmakers

Several papers suggest that prices in betting markets are set by bookmakers, who take positions in order to maximize their expected profit. Levitt (2004), for example, argues that bookmakers are better at predicting game outcomes than the typical bettor. As a consequence, bookmakers are able to set prices in order to exploit their superiority over the bettors in forecasting event outcomes. This can yield greater profit than could be obtained if the bookmakers acted like traditional market makers and attempted to set prices to balance supply and demand. Levitt provides empirical evidence of a U.S. bookmaker holding an unbalanced book (taking a position) against a large pool of NFL bettors. Further supporting evidence for U.S. sports (all of which employ point spreads and are therefore typically close to even-odds) has been presented by Paul and Weinbach (2007) and Humphreys (2010) for the NFL betting market, and Paul and Weinbach (2008) for the NBA betting market.

Kuypers (2000) explains that bookmakers seek to maximize profits and can even set odds that deviate from those indicated by unbiased probability estimates as a result of coming to different conclusions regarding how bettors will place their bets. Similarly, Marshall (2009) investigating instances of cross bookmaker price dispersion, states that bookmakers could remove odds discrepancies themselves after viewing the odds of their competitors but there is little reason to do this if they believe their odds better reflect the outcome probabilities than do those of their competitors; suggesting in other words that the objective of bookmakers' price-setting is to maximize profit rather than remove risk. Finally, in the same context, Franck, Verbeek, and Nüesch (2013) suggest that bookmakers purposely quote some odds at a level above those of their competitors as a marketing ploy to attract customers to their website. These actions, as pointed out by Marshall (2009),[10] would be consistent with the Salop and Stiglitz (1977) and Varian (1980) theory of spatial price dispersion.

In addition, Franck, Verbeek, and Nüesch (2013) suggest that such bookmakers do not necessarily maximize their profit per game. Rather, they aim to maximize their profits across their whole customer base, intending to earn a greater

[10] Interestingly, price dispersion remains persistent in the internet age (Baye et al, 2004) and Baylis and Perloff (2002) find that some online sellers persistently offer both high prices and poor services.

long term profit, even if several bets are accepted which are expected to result in a loss. This may include loss-leading marketing strategies. Since the bookmakers reserve the right to close the accounts or refuse the bets of those bettors who only bet against them at the bets where the bookmaker's odds are particularly good (higher than the market average or that perceived by the bookmaker to be fair odds), the policy of maximizing the customer base, is likely to maximize long-term profits for the bookmaker

Book-balancing Bookmakers

A more conventional view of bookmakers than that explored above is that they set their prices to eliminate risk by balancing the potential liabilities on all possible outcomes. As a result, they guarantee their payoff is close to their over-round irrespective of outcome of the event. This is an approach advocated by Sidney (2003), in a text designed to educate bookmakers. Magee (1990) also states that bookmakers adjust their odds regularly, in order to achieve a book which is as balanced as possible. Consequently, as Woodland and Woodland (1991) explain the market odds in this case are a reflection of the money staked on each possible outcome (or at least the bookmaker's early prediction of liabilities for either side[11]). It is evident that according to this business model, the bookmakers effectively act as market makers whose profits are only a function of the volume traded in their books.

Theoretical models of bookmaker behaviour also often take this perspective. For example, Fingleton and Waldron (1999) model the bookmaker as an infinitely risk-averse market maker, who seeks to avoid holding a liability on any outcome.

[11] Such bookmakers are likely to use information signaled by the stakes of bettors profiled as successful in order to estimate the volume anticipated to be traded on each potential game outcome. They are likely to use this information to set their early odds accordingly, even if these do not reflect the balance of the stakes up to that point. For example, assume that up to point *t*, $100,000 is staked on outcome A and $10,000 on its complementary, 'not A' (1 - A). This should push the book-balancer to increase the odds of A and decrease those of 1 - A. However, the bookmaker might do the opposite if the $10,000 is staked from bettors profiled as skilled and the $100,000 is staked by the general public. This may occur because they anticipate that betting volumes on 1 - A in the forthcoming period are likely to be significantly larger (based on the information derived from the current bets of informed traders) and is expected to wipe out the bookmaker's exposure, as this stands at point *t*.

They provide empirical evidence, based on 1,696 horse races in Ireland, that bookmakers are more likely to be risk-averse book-balancers than the risk-neutral profit-maximisers of Shin (1993) (who simply set prices to avoid losing to insider traders).

Cain, Law, and Lindley (2000) simultaneously model a bookmaker and bettor's aiming to engage in optimal betting. They find optimal price setting mechanisms for a bookmaker seeking to maximize his expected gains (a risk-neutral bookmaker), or seeking to ensure a sure gain (an infinitely risk-averse bookmaker). A third possibility of the bookmaker who seeks to maximize the minimum of gains on any outcomes is also considered (a maximin strategy). Their general finding is that bettors need to hold diverse opinions for the bookmaker to ensure a profit under any of these strategies, and that bookmakers may increase their profits by varying their overround. However, Cain et al (2000) do not consider the evolution of bookmakers' odds over time, nor do they consider how bookmakers may optimally set prices across different games to maximise their returns.

In a more recent paper, Hodges and Lin (2009) model the problem of a book-balancing bookmaker in a multi-period setting. The bookmaker sets odds in a manner to avoid excessive liabilities on any particular outcome, in a similar fashion to the market maker's bid-ask spread in the Stoll (1978) and Ho and Stoll (1981, 1983) inventory models of market microstructure. The model of Hodges and Lin (2009) predicts that the book-balancing problem becomes easier for the bookmaker as uncertainty relating to outcome liability is resolved near the start of a match. An interesting implication of this finding would be that, in a competitive marketplace, bookmakers seeking to balance their book should reduce their over-round.

From the bettor's point of view, a book-balancing bookmaker constitutes a similar form of market to a betting exchange. In a betting exchange, trades are conducted directly between the different bettors and the betting exchange (which facilitates the transaction) obtains a commission from the winner. Clearly, in this environment, bettors have the ability, via the facility of the betting exchange, to make their own markets. A similar mechanism to that seen in the betting exchange operates within a book-balancing bookmaker. In particular, they effectively simply act as an intermediary, receiving a small share of turnover in exchange for providing liquidity to match competing orders from different bettors. One could consider this role analogously with that of a market-maker on a stock exchange.

In general, however, the bookmaker-driven football betting market is far more liquid than the comparable betting exchange market. Duffie (2012) explores some reasons as to why over-the-counter (OTC) markets may prevail over exchange-based mechanisms. An obvious reason he points out is that betting public may prefer the bookmaker mechanism because it allows for flexibility in their offerings; an individual trader, for example, may not want to face the counterparty risk from offering low-probability, high dollar payouts (e.g., from exotic bets,) and may, therefore, avoid accepting such bets on an exchange. Franck et al. (2013) show, even for single-game bets on in the highly popular football betting market, betting exchange liquidity remains significantly lower than that of the bookmaker market. For example, Franck et al. (2013) report, for the 2010 data employed in their study, an average exchange-traded volume of only £64,907 per game in the top five leagues in Europe of. Forrest (2012) on the other hand quotes the liquidity available through a syndicate of Asian bookmakers of up to €300,000 on a single second-division Belgian game, and up to €1 billion on the 2011 Champions League final. Drawing on an analogy with the OTC bond trading market (Duffie, 2012), we note that the economic significance of offers in the bookmaker market are more consistent than those from betting exchanges. This is the case because bookmakers, in acting as market makers, guarantee liquidity. Consequently, unlike in betting exchanges, the execution of a trade does not depend on the reverse order having been placed earlier.

3.2.2. The Population of Bettors

The betting public is comprised of a heterogeneous population, having diverse backgrounds and exhibiting different behavioural characteristics. Gainsbury, Sadeque, Mizerski and Blaszczynski (2012), for example, analyse data covering 11,394 customers of a large Australian bookmaker and find differences in the frequency of betting, the size of the stakes and the success level (in the sense that several groups tend to lose more than others) across the betting population. Whilst there may be many sub-populations of bettors, for the purpose of our study, it is important to distinguish between two types of bettors; namely, the casual betting public whose bets exhibit negative expected returns on average and minority much smaller group of sophisticated bettors whose expected returns may be positive. The latter group may include bettors who (i) possess inside information, as suggested by Shin (1991, 1992, 1993); the existence of such a group in terms of football games has been observed by Forrest (2012); (ii) arbitrageurs; i.e. those who attempt to benefit from substantial pricing differences across different wagering operators (Hausch and Ziemba, 1990; Edelman and O'Brian, 2004; Marshall, 2009; Franck, et al., 2013), and (iii) bettors capable of successfully applying mathematical models to profit from betting; it is well documented that individuals of this sort operate in betting markets (e.g., Benter, 1994; Thorp, 2000).

Franck et al. (2013) suggest that some bookmakers are likely to restrict trade with those bettors whom they believe hold superior information. Bookmakers may either implement restrictions on the size of stake they are willing to accept, the type of bet they are willing to accept (such as restrictions on arbitrage betting), or simply cancel traders' accounts for 'commercial reasons'. Veitch (2009, pp. 231-233) presents his own experience regarding the lengths to which bookmakers can go in order to restrict the bets of successful bettors[12]. Franck et al. (2013) investigate the degree to which arbitrage opportunities arise between betting exchanges and bookmakers' odds and find that these occur in 19.2% of all matches analysed. They argue that this results as a consequence of the bookmakers intentionally mispricing events in order to attract customers. However, the observed arbitrage cases are effectively non-exploitable in the long run because bookmakers are likely to restrict or eliminate the activity of those bettors profiled as skilled in spotting these

[12] Veitch argues "people are shocked to hear that under current legislation bookmakers can advertise prices and boast freely about their willingness to lay large bets, only to refuse point blank to take a single penny if they are too wary of the person placing the bet" (p. 231).

opportunities. Consequently, the intentional mispricing could simply be regarded as part of the acquisition cost of new clients.

Marshall (2009) measured the median duration of arbitrage opportunities in football betting markets during the period 2003 – 2005 and found this to be 15.4 minutes. It seems fair to assume that nearly a decade later, with the popularity of internet-betting at its peak and the availability of software programs that claim to be able to explore arbitrage opportunities flooding the market, that if such instances of price dispersion were not intentional, they would have quickly been removed. Thus Franck et al.'s (2013) theory that the generation of such arbitrage opportunities are intentionally created by the bookmaker and are non-exploitable seems sensible. Hence, interestingly, instead of indicating market inefficiency, this finding suggests that several prices in the bookmaker markets may be there for some bettors but not for others. This implies that bookmakers may apply discriminatory policies against skilled bettors, such as those described by Veitch (2009).

Levitt (2004), analysing bettor-specific data supplied by a bookmaker, suggests that there is little evidence of individual bettors who are able to systematically beat the bookmaker. This result, Levitt claims, is consistent with his hypothesis that bookmakers take positions, as it would not make sense for them to do so if they did not exhibit superior forecasting ability compared to their clients. However, considering the option available to bookmakers to discriminate against potentially successful clients (Franck et al., 2013), the fact that Levitt finds no winning bettors in the data supplied by the bookmaker could be the result of the bookmaker's self-fulfilling prophecy rather than the lack of bettors capable of beating the market. In other words, it is possible that the latter group exists, but is just not welcome by a position-taking bookmaker such as the one whose data Levitt examined.

Conversely, for a book-balancing bookmaker, such as that described in Woodland and Woodland (1991), the objective is to maximize trading volume, as this is the only determinant of such an operator's profitability. Hence, unlike the position-taking bookmaker, a book-balancing operator has no incentive to eliminate potentially successful bettors as such an operator takes no position against them[13]. According to Forrest (2012), it might not even be possible for a book-balancing bookmaker to know if a large bet is arising from a professional gambler, from a regional Asian bookkeeper offsetting a single bet against the online bookmaker arising from aggregated bets collected on the streets, or even from an insider trader. It could even be argued, that since successful bettors are likely to apply reinvestment strategies, such as the Kelly Criterion (an investment strategy which optimizes long run wealth and involves increasing stakes as the investor's bankroll grows), their stakes on individual games are expected grow over time, until they are bound by the market's staking limits. Consequently, such clients are likely to be massive suppliers of liquidity (and therefore, profitability) for a business model in which trading volume is the decision variable.

The application of the Kelly Criterion results in the exponential growth of successful bettors' capital. This, together with the book-balancing bookmaker's turnover-maximization strategy, leads to the book-balancer uses staking limits to manage the ratio of volume fed by professional bettors against stakes from the unsophisticated betting public. It also results in them using regular movements of odds as a tool to minimize potential exposure. In other words, this type of bookmaker sets a limit on the size of the stake that it is willing to accept at a given level of odds. In addition, they reduce the odds on whichever outcome receives a sizeable bet and increase the odds correspondingly on other outcomes (as a result the probability of receiving sizeable bets on those outcomes increases). The result of this process may be that the bookmaker is guaranteed a profit somewhat lower than the over-round, since the stakes on all outcomes might be higher when the corresponding odds were above the average level for a given offer during its life cycle[14]. It is also likely, that the level

[13] Pinnaclesports a major bookmaker following this business model states on its website "our success derives from the economy of scale that a high volume of bets generates – think Walmart or Tesco. This approach means that we welcome all bets, so unlike most online bookmakers, winners are welcome" (http://www.pinnaclesports.com/betting-promotions/winners-welcome).

[14] Imagine the bookmaker sets odds of (1.96, 1.96) for (A, 1-A) and receives a bet of $10,000 on A. It may respond by changing its odds to (1.94, 1.98) and it may then receive a $10,000 bet on 1 – A. The result of the two bets is a fully hedged position for the bookmaker. However, the outcome is that the bookmaker's expected profit is lower than that implied by the over-round (as the trades were ($10,000, $10,000) at (1.96, 1.98), which corresponds to a 1.5%, rather than 2% over-round.

of the odds' adjustment after a high stake may vary depending on the profile of the bettor who placed it, provided the individual can be identified; i.e. a greater decrease in the odds should be expected after a bettor identified as skilled (rather one profiled as average) places a large bet (Forrest, 2012). Levitt (2004) argues that it should be expected that "the most talented individuals would be employed as the odds makers" (p. 245) and as a result bookmakers will always be able to forecast event outcomes more accurately compared to bettors overall. We suggest that for Levitt's claim to be valid, assuming rational expectations, it has to be shown that it is not possible for talented bettors to beat the market (e.g. that winning bettors are not welcome in any form of market) or that bookmakers can remunerate odds-setters to a greater extent than could be achieved via a successful betting system. The second of these conditions is unlikely to be true given the magnitude of the financial success enjoyed by professional gamblers such as Benter and Thorp (e.g., Benter, 1994; Thorp, 2000). Consequently, in order to test Levitt's proposition, we investigate the evolution of prices in a market setting where both book-balancing bookmakers (whose odds are driven by the flow of money and *smart money* in particular) and position-taking bookmakers (whose business model matches that described by Levitt (2004)) operate in order to explore whether the evidence suggests that it is not possible for talented bettors to beat the market.

To achieve this objective, we examine price movements in the football betting market. Specifically we compare price movements of (i) a major Asian bookmaker, SBOBet, which we use as an example of a 'book-balancing' bookmaker who sets low over-rounds and who shifts their odds almost continuously in response to stakes placed by bettors and (ii) Ladbrokes, as an example of a major European, position-taking bookmaker, who charge a relatively high over-round to bettors and change their prices with far lower frequency.

3.3. Hypotheses Development

Levitt (2004) suggests that bookmakers are superior forecasters of sport event outcomes compared to bettors and they set odds that efficiently reflect the corresponding probabilities of the alternative results. Consequently, he argues, they do not have to move these odds often in order to balance their books, as taking positions will lead them to higher profits. According to this theory, a bookmaker should rarely move their odds and if they do so, this will happen in order for them to better reflect the outcomes' probabilities rather than to minimize their exposure. As a result, a bookmaker that very frequently adjusts its odds, reflecting the volume of trading on different outcomes, does not fit with Levitt's model of a bookmaker. Rather they might better be described as a book-balancer, the sort described by Woodland and Woodland (1991). However, according to Levitt, bookmakers are superior forecasters compared to the population of bettors as a whole and consequently, a bookmaker that adjusts prices responding to the stakes placed is unlikely to predict outcomes with higher accuracy than the 'expert' bookmaker described by Levitt (2004). As explained by Marshall (2009), even in the extreme that the resulting price differences are significant enough to generate arbitrage opportunities, still the expert-bookmaker has no reason to move its odds, since this will result in distancing them from the efficient line, decreasing its profit in the long run. However, Smith, Paton and Vaughan Williams (2006, 2009) and Franck et al. (2010, 2013) provide evidence that odds derived from betting-exchange markets constitute superior forecasts of match outcomes compared to bookmaker markets. This finding is at odds with Levitt's claim, as it suggests that a demand-driven market evaluates the outcomes of sport events better than the expert bookmakers. According to Kuypers (2000), Forrest and Simmons (2008) and Franck, et al. (2013) this difference in efficiency between betting exchanges and bookmaker markets could be attributed to inefficient price-setting. This in turn, they argue, might be structural in bookmaker markets, due to bookmakers trying to take advantage of bettors' sentimental betting or accepting bets at a disadvantage in order to attract customers. According to this theory, these bookmakers are still operating as position-takers that create such inefficiencies intentionally in order to achieve their long term objectives. As a result, they are not expected to move their odds responding to bettors' demand, even if this results in the generation of arbitrage opportunities, since, as pointed out by Franck et al. (2013) they can always refuse bets from arbitrageurs or other skilled bettors.

Book-balancing bookmakers, whose profits are maximised by maximising turnover, have little incentive to eliminate or restrict sophisticated or potentially successful customers such as insider traders, efficient model-based traders and arbitragers. On the other hand, as discussed above, position-taking bookmakers are likely to restrict the activities of such bettors (Franck et al., 2013). Consequently, the population of successful bettors are only likely to have access to book-balancing bookmakers in order to place large stakes or, but to a much lesser degree, to betting exchanges. Their access to betting exchanges is restricted to those events where there is sufficient volume on the outcomes against which they wish to bet and where there are no restrictions regarding the distribution of profit [15]. The implication of this is that information from skilled bettors is very likely to be passed onto book-balancing bookmakers. The theories advocated by Levitt (2004), Kuypers (2000), Forrest and Simmons (2008) and Franck, Verbeek, and Nüesch (2010) would suggest that this should have no impact on the prices of position-taking bookmakers. In particular, this arises because, according to Levitt (2004) they are superior forecasters to bettors (and, as a consequence to book-balancing bookmakers who simply rely on developing odds based on the weight of money on the various outcomes) and according to Kuypers (2000), Forrest and Simmons (2008) and Franck et al. (2010) because their price inefficiencies are intentional. However, we argue that the client-discrimination model suggested by Franck, et al. (2013) implies a belief by the bookmakers that their odds can be beaten. Moreover, by eliminating skilled bettors, they lose direct access to the only market players likely to yield superior forecasts than theirs, something that would help these bookmakers move their odds closer to the objective probabilities of the outcomes and consequently, to gain a higher margin from their clients (i.e. casual bettors). Nevertheless, if the position-taking bookmakers believe that their odds can be improved, based on information arising from the informed betting public, they are

[15] According to its regulations, Betfair (the most popular betting-exchange) can withhold up to 60% of a winning player's profit (http://www.betfair.com/www/GBR/en/aboutUs/Betfair.Charges/). This is an obvious deterrent for skilled bettors, who are likely to be alienated by such a policy. For related criticism, follow the links below:
http://www.independent.co.uk/sport/racing/betfair-faces-criticism-for-massive-rise-in-charges-924359.html
http://www.theguardian.com/sport/2008/sep/09/horseracing1
http://www.theguardian.com/sport/2011/jun/29/betfair-premium-charge-increase

expected to adjust their odds according to changes in the book-balancing market, which is expected to welcome skilled bettors. Forrest (2012), based on anecdotal information, suggests that traditional position-taking bookmakers' odds do indeed follow trends in Asian bookmakers' odds who act as market makers.

Of course, we do not suggest that book-balancers never take positions (i.e. by purposely not adjusting prices in order to balance their books in cases where they have a strong opinion on the outcome of the event) and that position-takers never move prices for hedging purposes. Hence, there is a spectrum between book-balancing and position-taking among which different companies position themselves. It is possible that even the same bookmaker might move merely towards book-balancing or position-taking from time to time, based on the dynamics of their customer base and market trends. However, we characterize bookmakers as book-balancers or position-takers according to the predominant policy, which is indicated by the objectives of each operator; i.e. maximise volume or maximise the profit margin, as implied by their pricing policy (for further details see below).

To examine the arguments explored in the preceding discussion we develop three testable hypotheses: First, we examine to what extent bookmakers are heterogeneous in their operations. As indicated above, we believe that 'book-balancing' market makers will aim to realise a risk-free profit by earning profits based on the over-round, requiring a high turnover. On the other hand, position-taking bookmakers (e.g. the only bookmakers to be referred to by Levitt, 2004) will set a price and aim to realise profits based on their superior forecasting ability. Therefore, we expect that book-balancers will accommodate larger stakes from sophisticated bettors, and move odds regularly to attract volume to attract bettors to the underweighted side of their book. On the other hand, position-taking bookmakers will operate with a higher over-round to compensate for adverse selection, avoiding taking bets from sophisticated bettors, and rebalance odds less frequently in response to inventory mismatches. The position-taking bookmaker's lower turnover vis-à-vis the book-balancing bookmaker is likely to offset by a higher profit margin.

To examine the veracity of these views we test the first hypothesis, namely: The book-balancing bookmaker (SBOBet) changes odds more frequently and charges a lower transaction cost (over-round) per dollar bet than the position-taking bookmaker (Ladbrokes). Sbobet and Ladbrokes are leading bookmakers, having managed to establish reputable brand names in Asia and Europe respectively, attracting a significant customer base and trading volumes as a consequence (see pp. 106 – 109 of this thesis). Hence, these two bookmakers are characteristic representations of the two types of bookmaking. Moreover, the correlation in the prices between the bookmakers in the book-balancing and position-taking groups are such, that the choice of the bookmaker is unlikely to influence the results of the study (see Table 4.4, p. 174 of this thesis).

As discussed above, the sophisticated bettors are only likely to trade with the book-balancing bookmakers due to restrictions placed upon them by the position-taking bookmakers. Consequently, information is expected to flow from the prices exhibited in the book-balancers' odds (which are adjusted to accommodate the information of the informed bettors), to those of the position-takers. The prices of the position-taking bookmakers are therefore expected to lag those of the book-balancing bookmaker. Informed bettors will therefore be responsible for price movements in both types of bookmaker market. This contrasts with Levitt's (2004) conjecture that bettors are mainly noise traders with an inability to influence market prices. To test this view we test hypothesis 2a, namely that: Price changes at the position-taking bookmaker (Ladbrokes) converge to *lagged* price changes at the book-balancing bookmaker (SBOBet) and hypothesis 2b, that price changes at the book-balancing bookmaker are not influenced by *lagged* prices of position-takers. Hence, due to the lead-lag relationship between the book-balancing bookmaker and the position-taking bookmaker, the closing prices of the position-taker will be significantly related to early prices set by the book-balancer and their own early prices, but closing prices of the book-balancer will not be significantly related to those of the position-taking bookmaker, after controlling for their own early prices. Thus, we explore this by testing hypothesis 2c, that the closing prices of the position-taking bookmaker (Ladbrokes) will be significantly related to both their own early prices *and* the early prices of the book-balancer (SBOBet), whereas the closing prices of the book-balancing bookmaker will be related to their own early prices, but not the early prices of the position-taking bookmaker

The collective information of all bettors will influence prices over time. Information that is probably related to the outcome of the event is revealed closer to kick-off (i.e. market close). For example, the teams' final line-ups, weather and pitch conditions, the attendance figures etc may only become apparent in the last few minutes before the kick-off.. Moreover, the maximum stakes accepted by book-balancing bookmakers increase significantly closer to kick-off [16]. Consequently, bettors who wish to avail themselves of the maximum information and volume (and this is likely to apply to the skilled bettors who aim to maximize their return from betting) are likely to bet closer to the market close. As a result, due to the progressive increase in the trading volume, as well as the availability of information as time approaches the game's kick-off, we expect later odds to be more informative of true match outcomes than earlier odds for both book-balancing (driven by the flow of *smart money*) and position-taking bookmakers (following significant odds moves of book-balancers). However, because the sophisticated bettors trade mainly with the book-balancing bookmaker, closing prices of the book-balancing bookmakers are likely to be more efficient predictors of match outcomes. This leads to two related hypotheses, namely (i) hypothesis 3a: Closing odds (those collected just prior to kickoff) will be more efficient predictors of match outcomes than early odds (those collected one day before kickoff), for both the book-balancing bookmaker (SBOBet) and the position-taking bookmaker (Ladbrokes), and (ii)hypothesis 3b: Closing odds from the book-balancing bookmaker (SBOBet) will be more efficient predictors of match outcomes than the closing odds of the position-taking bookmaker (Ladbrokes).

[16] See chapter IV of this thesis for a comparison between the staking limits offered a day prior to kick-off and those offered within hours of the kick-off.

3.4. Data and Methodology

3.4.1. Data

We analyse time-stamped home team, draw, and away team (1x2) prices collected consistently from two bookmakers (Ladbrokes and SBO) at points 1 day, 16 hours, 8 hours, 4 hours, 2 hours, 1 hour, 30 minutes and 1 second prior to kick-off for all games in our data set. Our database of matches consists of every game played in season 2012/13 in the 6 most prominent European football leagues: The English Premier League, the Spanish La Liga, the Italian Serie A, the German Bundesliga, the French Ligue 1 and the Dutch Eredivisie. This constitutes a sample 2,132 games, for each we collected odds offered at the above stated time points, for a total of 2,132×8 = 17,056 1x2 group of odds offers, leading to a total of 17,056×3 = 51,168 individual odds offers per bookmaker on potential game outcomes.[17]

We had to post a request to the bookmakers' servers to obtain the data. Consequently, the choice of eight odds points was made as a result of a trade-off between data availability and reliability. An increased frequency in the collection of data would have meant many more requests to servers, particularly when matches may kick off contemporaneously. Increasing the number of lags per game would have increased the number of requests we made to the bookmakers' servers. This posed the risk that the bookmakers might restrict our access to their websites, by blocking our Internet Protocol address. Hence, we decided not to exceed the eight time-points per game. Moreover, we did not want to overly strain the servers at the two bookmakers by collecting more odds than we believed necessary to complete this study. With this proviso we ensured that we collected data at points with increasing frequency as match

[17] Although the traditional match outcome (Home team win, draw, away team win) betting is most popular in Europe, SBOBet specialises in Asian Handicaps. We chose to use match outcome data as these were most likely to be liquid at the position-taker, and hence most likely to lead to reliable odds movements. Moreover, it should be noted that there are missing offers in the data (about 10% of the sample), as our data collection programs have occasionally failed, due to various spontaneous problems such as changes or overload in the bookmakers' websites, temporary loss of the internet connection and power-cuts. However, the missing data do not influence the conclusions drawn in this paper as the confidence level provided by the statistical significance of all results is very high.

kick-off approached, as we aimed to get an even distribution of moves across the time periods. As kick-off nears, the activity of bettors tends to increase and so the exponential frequency of odds collection aims to capture a similar amount of betting activity within each period.

3.4.2. Identifying characteristics to distinguish the type of bookmaker

The time-stamped price betting data was collected from two different bookmakers, Ladbrokes (LAD) and SBOBet (SBO). These two bookmakers were intentionally chosen to provide good examples of a position-taking bookmaker (Ladbrokes) and a book-balancing bookmaker (SBOBet). We provide a brief overview of their operations here in order to distinguish the key features of their operations that allowed us to identify the type of bookmaker they represent:

Ladbrokes

Ladbrokes is a traditional UK bookmaker (incorporated in Gibraltar), dating back to 1886, with over 16,000 employees, and they claim to be the most recognised betting brand in the United Kingdom.[18] They operate more than 2,800 retail betting shops in the UK, Ireland, Belgium, and Spain. Ladbrokes' website claims to have attracted over 1 million active clients.[19] According to their financial statements, over £17 billion was staked by customers over the entirety of Ladbrokes' operations in the 2012 financial year; the company achieved net revenue in excess of £1 billion that year. Over 80% of their revenue is sourced from betting shops and in 2012 the company reported that 17% of their revenue was generated online. Ladbrokes offers a diversified range of gambling services, including racing and sports betting, but also online casino games, poker, bingo, and in-store machines. Sports-betting is likely to be used as a marketing tool by Ladbrokes to attract customers to these other (less-risky) operations. From their 2012 annual report, customer acquisition costs were placed at

[18] Ladbrokes 2012 Annual Report, p.15
[19] http://www.ladbrokesplc.com/about-ladbrokes.aspx

£107 per customer in 2012 by Ladbrokes, inclusive of promotions and bonuses netted from online revenue. Ladbrokes expects to recoup these costs over the long run as casual betting customers tend to remain with a single bookmaker, even if more favourable terms may be available elsewhere.

As Franck et al. (2013, p. 311) point out, Ladbrokes actively discourages sophisticated clientele, and ‘reserve the right to refuse part or all of a bet.’ Using customer information[20] based on historical trades, Ladbrokes is able to create customer profiles in an effort to identify and restrict the activities of potential arbitrageurs or sophisticated traders who only (or mostly) bet when prices are overly favourable. Veitch (2009) names Ladbrokes (and William Hill) as a bookmaker that eliminated his direct access to its books and applied policies to restrict even his potential indirect access (p. 232). Consequently, such bookmakers are able to mitigate potentially large losses due to adverse selection through trade discrimination. This allows for Ladbrokes to operate under the high-margin, low-turnover model of the traditional position-taking bookmaker.

Ladbrokes notes under its key risks in its 2012 annual report (p. 23): “the online gambling market is characterized by intense and substantial competition and by relatively low barriers to entry for new participants. In addition, Ladbrokes faces competition from market participants who benefit from greater liquidity as a result of accepting bets from jurisdictions in which Ladbrokes chooses not to operate.” These restricted territories include the United States, Greece, Italy, and China. As Forrest (2012) notes, much of the betting on football now comes from South-East Asia (directly or indirectly), and Ladbrokes appears have made the conscious decision to not aim to compete with the Asian bookmaking market.[21] Ladbrokes also note (2012 Annual Report, p. 24) that they face “a relatively high fixed-cost base as a proportion of its total costs, consisting primarily of employee, rental and content costs associated

[20] Franck, Verbeek, and Nüesch (2013) note that identification practices may include cookies, log files, clear gifs, and the engagement of third parties.

[21] The website analytics service, Alexa, reported on Sept. 11, 2013, that 56% of the traffic to Ladbrokes website came from the United Kingdom, with 5% from the United States, followed by minor percentages of traffic from Germany, Japan, and Sweden. They also note a high representation of visitors to the site ‘who did not go to college.’

with its betting shop estate.” Although the high fixed-component of Ladbrokes cost is not necessarily a driver of its odds-setting process, it does illustrate that they are unlikely to be able to operate at the ultra-efficient levels of an online-only sports betting agency. Other bookmakers (mainly European), including those without physical betting shops may also operate as position-taking bookmakers, competing in the market for relatively unsophisticated traders and engaging in similar practices to Ladbrokes to ensure they can retain a high-margin, low-turnover business. This includes the sample of bookmakers considered in the study of Franck et al. (2013), who generally offer ‘signing-up bonuses’ and other side products, and all restrict activities to seemingly-sophisticated players.

SBOBet

SBOBet (the first three letters standing for “Sports Bookie Online”) is a major online bookmaker licensed in the Phillipines (Cagayan Special Economic Zone) and Isle of Man. It is a subsidiary company of Celton Manx, Ltd, a private company, founded in 2008. As such, there is less publicly-available information relating to the history and profitability of the bookmaker or its specific operations. SBOBet was awarded the prize for Asian Operator of the year in 2009 and 2010.

The Institute of International and Strategic Relations (IRIS) Report (2012) explores corruption in sports and presents a detailed analysis of betting markets in Asia (see also Forrest, 2012). They identify one of the four major Asian players as SBOBet.com, and note (p. 44) that it ‘represents heritage of an activity begun in 1994 in Singapore that spread to Malaysia, Indonesia and then the Philippines, where the sports betting business acquired an online betting license in the economic area of Cagyan, a very lax jurisdiction.’ At present, the web analytics service, Alexa, reveals that the majority of visitors to SBOBet.com hail from South-East Asia; the top five nationalities of visitors, which make up over 68% of traffic originate from Indonesia, Thailand, Japan, The Philippines, and Malaysia. However, as the IRIS report notes, the few Asian sites[22] represent the overground section of a vast pyramid scheme. SBOBet represents part of the highest level of a large pool of regional bookmakers, who collect bets from the wider population through a set of localized bookmakers. These regional bookmakers then hedge their own risks online. Hence, SBOBet (among others) are accustomed to accepting very large bets, as an amalgamation of a portfolio of small bets. The IRIS report (2012) in particular notes, that ‘these sites offer a particularly high rate of return to the bettor (around 97%), the low margin being offset by the very high volume of bets.’ The figures quoted suggest that a single client can place a stake at one of the large Asian bookmakers of around 20 times the amount that a European bookmaker would accept for a major European championship.

The executive director of SBOBet.com, Bill Mummery, in an article through EGR magazine, also explains some key differences between the Asian and European markets. He states, “The gaming culture in Europe is one of ‘I want to place a very small bet for a life-changing experience’ (whereas in Asia bettors are able to) make a value judgment very quickly and reckon to make 2-3% return on their bet.” Football betting is particularly popular in both markets due to the high quality of the matches, and easy accessibility of the televised product.

SBOBet, as with the other Asian bookmakers, positions itself in the globalised football betting market as being particularly focused on Asian Handicaps – a product that reduces the usual three-outcome (1x2) betting on football by creating a ‘handicap’ of fractional goals. For example, a favourite might be given a handicap of a half-goal, and thus needs to win by a clear goal for a bet on the favourite to pay out (a drawn match would be a win for the longshot). This reduces the complexity of the book-balancing problem for the bookmaker, as there are only two lines to move, which helps drive the high-volume, low margin model.

As Figure 3.1 shows (below, bottom left corner), SBOBet differs from most European bookmakers in that they publish the maximum amount they would be willing to accept on particular outcome. The quoted figure from SBOBet here is ‘Brest, -0.5, @2.09, MaxBet GBP2,223.’ A bettor can place a stake of up to £2,233 on the

[22] The others identified in the IRIS report are 188bet.com, ibcbet.com, and 12bet.com

Asian Handicap for the French League Cup match Brest vs. Auxerre, giving Auxerre a half-goal start, at gross odds of 2.09 per unit stake. If the bettor were to seek a larger stake, it would create an imbalance in SBOBet's liabilities – price concessions to the bookmaker would be required beyond this point.

In reporting market depth, SBOBet shows that it is relatively indifferent to the identity of the counterparty (this price and volume is available for any bettor on the site). Bettors are able to observe both prices and depth movements by refreshing the site. Although trade-by-trade data is not publicly available, the lack of discrimination in counterparties and more transparent structure helps us identify SBOBet as a book-balancer.

Figure 3.1: Screenshot of SBOBet's trading window.

Key distinguishing features of Position-takers and Book-balancers

Football betting is very popular in both markets, but the two bookmakers operate in very different fashions. The key characteristics that help to classify a bookmaker as either a position-taker or book-balancer are summarised in Table 3.1.

Table 3.1: Key distinguishing features of Position-takers and Book-balancers

Characteristic	Position-Taker (e.g. Ladbrokes)	Book-Balancer (e.g. SBOBet)
Transaction Cost	High Margin	Low Margin
Turnover	Low Volume	High Volume
Clientele	Restrict Service to Sophisticated Traders and Arbitrageurs	Do Not Discriminate Against Sophisticated Traders
Clientele	European	Asian
Speciality Product	1x2 (Home, Draw, Away) Betting	Asian Handicap
Maximum Stake	Low Maximum Stake, Non-Transparent Pre-Trade	High Maximum Stake, Transparent Pre-Trade
Customer Acquisition	Loss-Leading Promotional Odds	Do Not Offer Promotional Odds
Cost Base	Relatively High Fixed Costs (incl. Physical Locations)	Relatively Low Fixed Costs (Strictly Online)

3.4.3. Methodology

3.4.3.1. Differences between bookmakers

Calculation of transaction costs for each bookmaker

The level of transaction costs for each bookmaker is generally measured by the extent to which the sum of their odds-implied probabilities across all match outcomes exceeds unity. This is known in betting parlance as the over-round. The level of the over-round is related to the expected loss of the average, uninformed bettor against the bookmaker. Hence the higher the over-round the higher the bettor's expected loss. We measure the over-round, $\rho_{i,k,t}$, at each time point, t in the lead-up to game i, for each bookmaker $k \in \{SBO, LAD\}$, by adding the inverse of the gross payoffs per dollar bet $X_{i,k,t,o}$ for each of the three outcomes $o \in E = \{H, D, A\}$ in the football match (H denoting a home win, D denoting a draw, and A denoting an away win) [23].

$$\rho_{i,k,t} = \frac{1}{X_{i,k,t,H}} + \frac{1}{X_{i,k,t,D}} + \frac{1}{X_{i,k,t,A}} - 1 \qquad (1)$$

Thus we have a measure of the transaction cost at nine points in the lead-up to the kick-off to game i, for both bookmakers. The average overround, $\bar{\rho}_{k,t}$ for each bookmaker, across all matches, for all points in time is computed and the bookmakers' average trading costs are compared using the pooled t-test, with test statistic T:

$$T_{k,t} = \frac{\bar{\rho}_{SBO,t} - \bar{\rho}_{LAD,t}}{S_{SBO,LAD} \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \qquad (2)$$

where

$$S_{SBO,LAD} = \sqrt{\frac{(n_1 - 1)S_{SBO}^2 + (n_2 - 1)S_{LAD}^2}{n_1 + n_2 - 2}}$$

is the estimate of the common standard deviation of the two samples, and n_1 and n_2 are the number of observed matches for SBO and LAD, respectively. Our first hypothesis predicts that the transaction costs for the book-balancing firm, SBO, should be significantly lower than their position-taking counterpart, LAD, and therefore the test statistic in (2) should be significantly negative nearer to match kickoff.

[23] The average uninformed bettor who bets across the odds range against the bookmaker is expected to lose an amount equal to $\frac{\rho_{i,k,t}}{\rho_{i,k,t}+1}$ on a unit stake.

Estimation of Frequency of Odds

The average frequency of odds changes across the sample of games for each game is calculated and the hypothesis that the book-balancing bookmaker moves its odds significantly more often compared to the position-taking bookmaker is tested. First, we calculate the difference in prices for each bookmaker, for each match-outcome-time triple:

$$\Delta X_{i,k,t,o} = X_{i,k,t,o} - X_{i,k,t-1,o}$$

As we are only interested at this point in whether the odds move or not, rather than the actual size of the movement, we tally each match-outcome-time triple based on a categorical score, $M_{i,kt,o}$:

$$\begin{aligned} M_{i,k,t,o} &= 1 \quad if \quad \Delta X_{i,k,t,o} \neq 0 \\ M_{i,k,t,o} &= 0 \quad if \quad \Delta X_{i,k,t,o} = 0 \end{aligned} \tag{3}$$

A simple binomial test is carried out to determine if the frequency of odds movements at SBO is higher than the frequency of odds movements at LAD, or $\frac{M_{i,SBO,t,o}}{n_1} > \frac{M_{i,LAD,t,o}}{n_2}$.

3.4.3.2. Transmission of Information within the Market

Serial Correlation in Odds Movement

We conduct unit-root tests to determine whether the first-ordered differences in the bookmakers' odds are stationary and utilize the Bayesian Information Criterion, in order to conclude how many lags are appropriate for further analysis of changes in odds. It could be argued, for example, that a greater frequency of odds movements does not imply anything beyond noise around the bookmakers' true probabilities, similar to bid-ask bounce observed in financial market microstructure. Odds movements would therefore be related to small changes in bookmaker liability, rather than information-based price moves, and the time series of odds movements would be unpredictable based on information arising from odds at previous time-points. Alternatively, if information is driving price movements, we would expect lagged movements to be important in predicting odds changes. Moreover, the time series of price movements may exhibit higher degrees of serial correlation if bookmakers' odds either underreact or overreact to information flowing from bettors.

Odds movements across outcomes within games are complementary, inasmuch that an increase in odds on the favourite winning will generally coincide with a decrease in odds on the longshot winning. Draw outcomes are notoriously difficult to predict for both experts and models (e.g. Pope and Peel, 1989; Goddard, 2005). Moreover, movements in draw odds are similarly likely to be driven by bets on one of the other outcomes, rather than based on specific information concerning the likelihood of this outcome. Thus we restrict our analysis to only the favoured team (defined as the team with the higher gross payoff-reciprocal, or odds-implied probability at SBO, one-day before kickoff in each game[24]).

We conduct a Fisher-type unit root test on panel data (Choi, 2001) for each bookmaker separately to test odds movements for stationarity. Each panel consists of match-time point observations of bookmaker prices. The Bayesian Information Criterion (BIC) is used to find the optimal number of lags to include in our regression model to test hypothesis 2; this allows us to find a trade-off between model fit and parsimony (see the next subsection for further discussion of the BIC).

Modelling the Changes in Odds Movement at the Two Bookmakers

In order to test hypotheses 2a, that position-takers' odds are expected to converge to those of book-balancers and 2b, that book-balancers' odds are not influenced by position-takers' odds, we employ a random-effects model to look at the

[24] In our dataset, there are no cases in which the draw is the favoured outcome. We take the odds one day before kickoff to determine the favourite in case the favoured team changes in the lead-up to kickoff.

determinants of odds changes at the two bookmakers. A random effects model is chosen to account for any unobserved variation in odds due to match-specific factors. As with the examination of lag length, the panel in the random-effects model consists of match-time observation points for each bookmaker. The general form of the random-effects models is as shown in Equations (4a) and (4b):

$$\begin{aligned}\Delta X_{i,t,LAD} = c &+ b_{11}X_{i,t-1,LAD} + b_{12}X_{i,t-1,SBO} \\ &+ b_{21}X_{i,t-2,LAD} + b_{22}X_{i,t-2,SBO} + \cdots \\ &+ b_{n2}X_{i,t-n,LAD} + b_{n2}X_{i,t-n,SBO} + U_i \\ &+ e_{i,t}\end{aligned} \tag{4a}$$

$$\begin{aligned}\Delta X_{i,t,SBO} = c &+ b_{11}X_{i,t-1,SBO} + b_{12}X_{i,t-1,LAD} \\ &+ b_{21}X_{i,t-2,LAD} + b_{22}X_{i,t-2,SBO} + \cdots \\ &+ b_{n2}X_{i,t-n,LAD} + b_{n2}X_{i,t-n,SBO} + U_i \\ &+ e_{i,t}\end{aligned} \tag{4b}$$

where $\Delta X_{i,t,LAD}$ and $\Delta X_{i,t,SBO}$ are the first differences in odds $(X_{i,t,LAD} - X_{i,t-1,LAD})$ and $(X_{i,t,SBO} - X_{i,t-1,SBO})$ for the favoured team at Ladbrokes and SBOBet, respectively, U_i is the game-specific error term, accounting for unobserved random variation across games, and $e_{i,t}$ is the i.i.d. error term from the regression. The generalised form of the model will be reduced to a more parsimonious form, with the appropriate number of lagged terms of bookmakers' odds on the right-hand-side of (4a) and (4b) selected using the Bayesian Infomration Criterion:

$$\mathrm{BIC} = \mathrm{n}\ln\left(\frac{1}{n}\sum_{i=1}^{n}\left(\Delta X_{i,t,k} - \Delta\hat{X}_{i,t,k}\right)^2\right) + g\ln n \tag{5}$$

where n is the number of observations[25], $\Delta X_{i,t,k}$ is the realised first difference in odds between time $t-1$ and time t at bookmaker k, and g is the degrees of freedom (the

[25] It is debatable whether this should be the number of observations or the number of groups in a panel data set. In this case, due to the low correlation of the 'within-panel' odds-differences, we use the number of observations.

number of parameters in the model minus one.) The first term in the BIC penalises poor model fit, while the second term penalises the number of parameters required to achieve the model fit. Thus a smaller value of the BIC means the model is preferred.

We expect that the models described by Equations (4a) and (4b) can be simplified. In particular, since hypothesis 2 is concerned with whether a particular bookmaker's prices will converge to those of the other bookmaker, the difference in the two bookmakers' prices in the previous lag is the main determinant of the odds. Therefore, equation (3) is expected to reduce to

$$\Delta X_{i,t,LAD} = c + B\left(X_{i,t-1,SBO} - X_{i,t-1,LAD}\right) + U_i + e_{i,t} \qquad \text{(6a)}$$

$$\Delta X_{i,t,SBO} = c + B\left(X_{i,t-1,LAD} - X_{i,t-1,SBO}\right) + U_i + e_{i,t} \qquad \text{(6b)}$$

where *B* is a positive coefficient indicating the degree of convergence of the bookmaker whose odds-changes are being forecast to those of the other bookmaker. For example, in Equation (6a) the coefficient *B* indicates the degree of convergence of Ladbrokes' odds to those of *SBObet*.

According to the arguments which lead to hypothesis 2a, if sophisticated gamblers can only access book-balancing bookmakers, it is expected that prices in that market will reflect quality information, which is unavailable to position-taking bookmakers. Therefore, the latter are expected to react to odds changes in the book-balancers' market and as a consequence adjust their odds when these deviate from those offered by book-balancers. Therefore, a significantly positive value for *B* is expected to be estimated for Equation (6a). Along these lines, Equation (6b) tests the influence of the position-taking bookmakers' odds (i.e. Ladbrokes) on the odds changes of the book-balancer (SBOBet). If the sophisticated traders are driving the market, hypothesis 2b predicts that the coefficient *B* in (6b) will be close to zero. Alternatively, if the theory of Levitt (2004) is correct, and bookmakers are the most accurate forecasters of sports events, there should be little convergence from position-taking bookmakers to the odds offered by book-balancers. In this case, the coefficient B is expected to be close to zero in Equation (6a).

Early-to-Late Odds Movements between Bookmakers

In testing hypothesis 2c, that the closing prices of the position-taking bookmaker (Ladbrokes) will be significantly related to both their own early prices *and* the early prices of the book-balancer (SBOBet), whereas the closing prices of the book-balancing bookmaker will be related to their own early prices, but not the early prices of the position-taking bookmaker, , we utilize a fixed-effects approach, incorporating normalized odds-implied probabilities for all three outcomes along one dimension of the panel, and all games along the other. This technique is used in order to examine the influence of the position-taking bookmaker's early odds on the closing odds of the book-balancing bookmaker and vice-versa. First, both bookmakers' payoffs ($X_{i,t,k,o}$) one day before kickoff (for ease of notation, time t_0) and at the time one second before kickoff (terminal time T) are converted to normalised probabilities ($Y_{i,t,k,o}$). This is achieved by dividing bookmaker$k$'s payoff reciprocal for outcome o by the gross over-round for the game i at time t from Equation (1):

$$Y_{i,t,k,o} = \frac{1}{X_{i,t,k,o}(1+\rho_{i,t,k})} \tag{6}$$

The fixed-effects models used to test the informativeness of odds in hypothesis 3a take the following form:

$$Y_{i,T,LAD,o} = c + \beta_1 Y_{i,t_0,SBO,o} + \beta_2 Y_{i,t_0,LAD,o} + \alpha_i + u_{io} \tag{7a}$$

$$Y_{i,T,SBO,o} = c + \beta_1 Y_{i,t_0,SBO,o} + \beta_2 Y_{i,t_0,LAD,o} + \alpha_i + u_{io} \tag{7b}$$

where $Y_{i,T,LAD,o}$ is the normalized probability implied in the terminal odds of Ladbrokes for outcome o in game i, $Y_{i,t_0,LAD,o}$ is the normalized probability implied in the day-ahead odds of Ladbrokes for outcome o in game i, $Y_{i,t_0,SBO,o}$ is the normalized probability implied in the day-ahead odds of SBOBet for outcome o in game i, $Y_{i,T,SBO,o}$ is the normalized probability implied in the terminal odds of SBOBet for outcome o in game i, α_i is the unobserved game-specific effect from the fixed-effects model, and u_{io} is the i.i.d. white noise error term.

The models in (7a) and (7b) test the influence of each bookmakers' odds on those of the other, as described in hypothesis 2c. In (7a) we test the influence of the early SBOBet odds on the terminal odds of Ladbrokes. Our expectation is that the coefficient β_2 is significant and positive, if the influence of sophisticated gambling money at the book-balancer drives the odds in the position-taker's market. In (7b) we expect that the position-taking bookmaker's odds do not influence those of the book-balancer, and therefore the coefficient β_1 should be close to 1, while the coefficient β_2 should be insignificant and close to zero.

Due to the expected high correlation between the explanatory variables in these models, we also test whether the nested model in (7c) excluding $Y_{i,t_0,SBO,o}$ provides a better fit than the unrestricted version in (7a), evaluated using the Bayesian Information Criterion in (5):

$$Y_{i,T,LAD,o} = c + \beta_1 Y_{i,t_0,LAD,o} + \alpha_i + u_{io} \tag{7c}$$

3.4.3.3. The Efficiency of Odds-Based Estimates

According to hypotheses 3a and 3b, closing odds provided by the bookmakers are expected to be more efficient predictors of actual game outcomes compared to early odds and book-balancers' forecasts (as implied by their odds) are expected to be more accurate than those of position-takers'. We run conditional logit models (McFadden, 1974) to test the relative efficacy of both the early and late odds at both bookmakers. Conditional logit modelling has been used to test the efficiency of odds in many previous betting studies (e.g. Figlewski, 1979; Asch *et al*, 1984; Bolton and Chapman, 1986; Benter, 1994; Sung and Johnson, 2010). The conditional logit model is used without any other explanatory variables, and takes the following form, where $E_{i,t,k,o}$ is the probability of the outcome $E \in \{H, D, A\}$ occurring in game i at time t with odds on outcome o from bookmaker k:

$$\mathrm{P}\left(E_{i,t,k,o} = 1\right) = \frac{e^{Z_{i,t,k,o}}}{\sum_{o=1}^{3} e^{Z_{i,t,k,o}}} \qquad (8a)$$

where

$$Z_{i,t,k,o} = b \ \ln\left(Y_{i,t,k,o}\right) \qquad (8b)$$

Model (8a) is estimated for each bookmaker at times t_0 one day prior to kickoff and T one second before kickoff. The model fit is evaluated using McFadden's (1974) pseudo-R^2 statistic; a higher pseudo-R^2 implying a superior model fit (relative to the naive prediction of each outcome having equal probability), and hence more explanatory power from bookmaker$k$'s odds at time t. Hypothesis 3a predicts that earlier odds should have less explanatory power than later odds, so the model fit at time T should be greater than the model fit at time t_0 for both bookmakers. Hypothesis 3b predicts that the position-taking bookmaker (Ladbrokes) provides less efficient odds than the book-balancing bookmaker (SBOBet). As such, we expect that the model fit should be higher for SBOBet than Ladbrokes at both times t_0 and T.

3.5. Results

3.5.1. Differences between Bookmakers

In order to provide some illustration of odds movements in the betting markets from a couple of days prior to the game until its kick-off, we examine the movement of odds through time, for 4 randomly selected games from our sample, for both bookmakers under investigation. These are presented in Figures 3.2 – 3.5. Casual observation of Figures 3.2 – 3.5 shows clearly that SBOBet has a greater frequency of odds changes than Ladbrokes. This is confirmed by analysis of the complete dataset. Table 3.2 presents an analysis of the proportion of occasions when price changes occurred at both Ladbrokes and SBOBet at different time intervals from the kick off of matches and descriptive statistics relating to price changes undertaken by the two bookmakers are provided in Table 3.2. It is clear from the information presented in Table 3.2 that there are far more occasions when SBOBet change their odds in a given time interval prior to a kick-off (cf. Ladbrokes) and the mean number of odds changes across time intervals during the last 24 hours prior to the game is 5.4 per match for SBOBet out of a possible maximum of 7 such changes, compared with only 0.8 changes per match for Ladbrokes (Table 3.3). A t-test is applied and it is confirmed that the difference in the frequency of changes is very unlikely to be random (p-value= 0.000). This confirms the first component of hypothesis 1: the book-balancer indeed changed their odds far more frequently than the position-taking bookmaker. We take this to imply that they are constantly responding to the large flow of money from bettors. This also provides evidence that the position-taking bookmaker is less likely to move prices in the lead-up to games, as suggested by Levitt (2004).

Figure 3.2: The evolution of Sbobet's and Ladbrokes' odds for the game between Sunderland and Stoke played on 6/5/2013 for the English Premierleague.

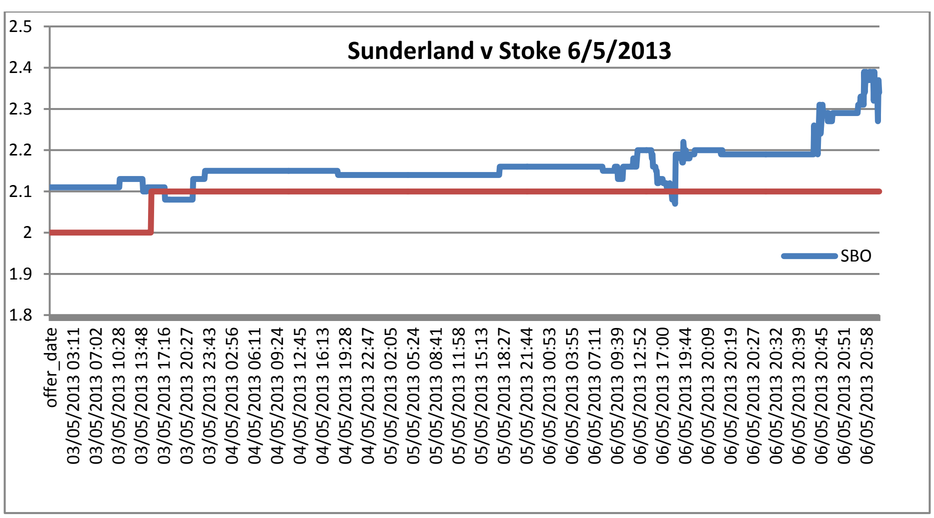

Figure 3.3: The evolution of Sbobet's and Ladbrokes' odds for the game between Dortmund and Freiburg played on 16/3/2013 for the German Bundesleague**.**

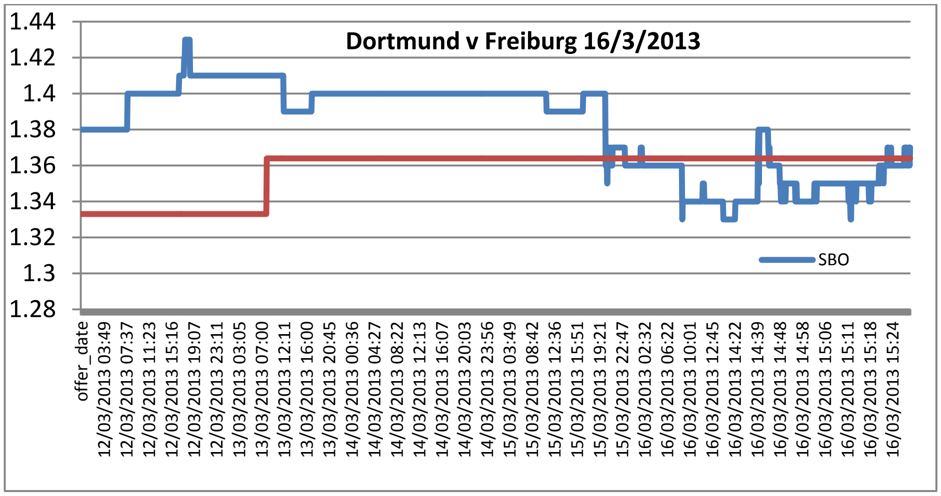

Figure 3.4: The evolution of Sbobet's and Ladbrokes' odds for the game between Tottenham and Everton played on 7/4/2013 for the English Premierleague**.**

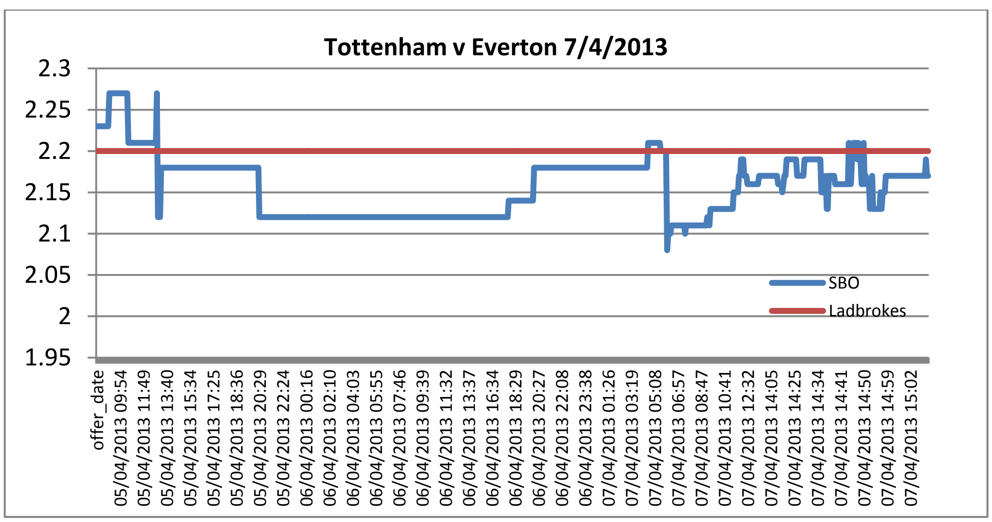

Figure 3.5: The evolution of Sbobet's and Ladbrokes' odds for the game between Fiorentina and Inter played on 17/2/2013 for the Italian Serie A.

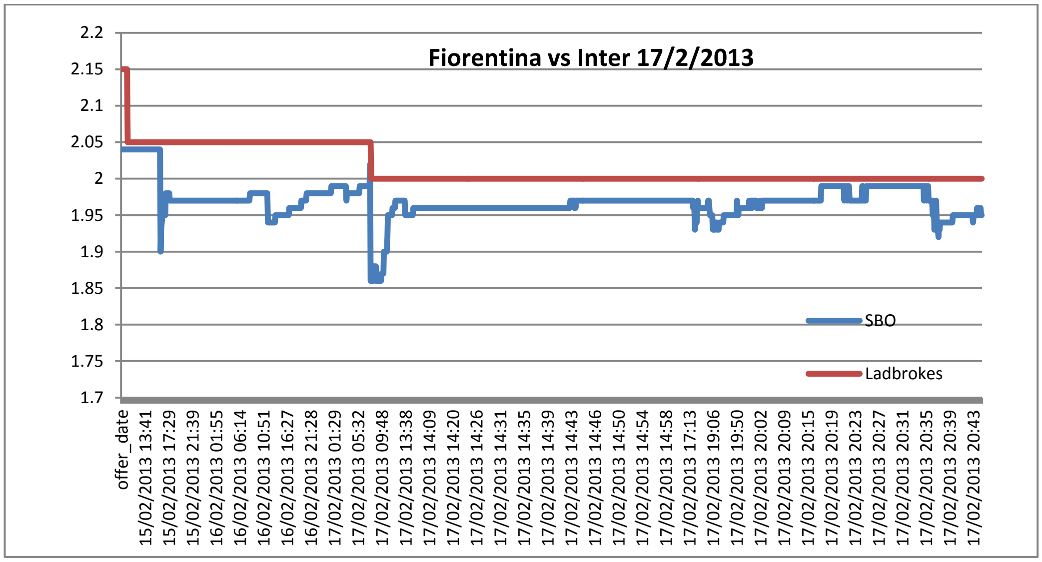

Table 3.2: Proportion of occasions in various time intervals prior to kick off when Ladbrokes and SBOBet changed their odds.

Time Period prior to kick off	**Ladbrokes**	**SBO**
over 1 day-16 hrs	0.09	0.64
16 to 8 hrs	0.11	0.75
8hrs to 4 hrs	0.22	0.86
4hrs to 2 hrs	0.15	0.77
2hrs to 1hr	0.09	0.73
1hr to 30 mins	0.09	0.78
(5 mins to 1 sec)	0.14	0.87

Table 3.3: Descriptive statistics regarding the number of price changes per match.

	Mean	St. Deviation	Min	Max
Ladbrokes	0.795	1.24	0	6
SBO	5.360	1.03	0	7

In Table 3.4, we present data on the bookmakers' margins in the form of their over-round (as calculated using Equation (1)). It is clear that SBOBet operates with a lower over-round at all points in time than Ladbrokes, and that the difference in over-round actually increases as the kickoff approaches. Applying a t-test confirms that the difference in over-round between the bookmakers is significant (p-value = 0.000), in each of the eight time periods. This confirms the second part of Hypothesis 1, namely that the book-balancer (SBOBet) does indeed operate at lower margins than the position-taking bookmaker (Ladbrokes). A low-margin operation is expected to lead to higher volumes, as more sophisticated traders are expected to prefer it to a bookmaker that offers consistently lower odds compared[26].

Table 3.4: Average over-round at Ladbrokes and SBO in lead up to kick-off.

Time Period prior to kick off	**Ladbrokes**	**SBO**
> 1 day	0.077	0.073
1day to 16hrs	0.077	0.070
16hrs to 8hrs	0.077	0.070
8hrs to 4 hrs	0.077	0.065
4hrs to 2 hrs	0.077	0.064
2hrs to 1 hr	0.077	0.064
1hr to 30 min	0.077	0.064
30min to 1 sec	0.077	0.064

[26] It has to be clarified that the main market for *Sbobet* is the Asian Handicap market, where the over-round is about 2%. On th eopther hand,for Ladbrokes, the over-round is at similar levels for binary markets as it is for 1x2. Hence, the presented differences on Table 3, significantly understate the difference in transaction costs across the two markets. However, it is interesting to point out that even in the 1x2 market (which makes up a far greater degree of business for Ladbrokes than it is does for *Sbobet*) the over-round of SBOBet is significantly lower.

3.5.2. Transmission of Information from Bettors to Bookmakers

The results of the Fisher type unit-root test, presented in Table 3.5, show that the first differences in the odds are stationary for both bookmakers. Therefore, we proceed to modelling such differences using Equations (4a) and (4b).

Table 3.5: Fisher-Type Unit-Root test statistics. The table presents the statistics for the four stationarity tests described by Choi (2001). Ho: The panels contain a unit-root.

Test	**Ladbrokes**	**SBO**
Inverse chi-squared	4976.5 (0.000)	12900 (0.000)
Inverse normal	-37.5 (0.000)	-55.7 (0.000)
Inverse logit	-45.1 (0.000)	-69.6 (0.000)
Modified inv. chi-squared	14.4 (0.000)	100.1 (0.000)

The Bayesian Information Criterion is applied in order to identify the number of lags that constitute the optimal trade-off between fit and complexity. The results are presented in Table 3.6. For SBOBet (Panel A of Table 4), the model producing the lowest BIC value (-9,075.40) is that incorporating the constant term only, and thus Equation (4b) is best modelled using $\Delta X_{i,t,SBO} = c$ .The model incorporating the lagged odds terms from both SBO and LAD, $\Delta X_{i,t,SBO} = c + b_{11} X_{i,t-1,LAD} + b_{12} X_{i,t-1,SBO}$, produces a significantly worse BIC (-9,062.76, p-value 0.002). This implies that odds movements in SBOBet prices are not related to previous odds movements in either SBO prices or Ladbrokes prices, or at least there is no improvement in forecasting power from adding lagged prices from either of the bookmakers.

However, the results are very different when we analyse differences in Ladbrokes' odds from one time period to another prior to the kick off. When a one-lag model is employed, i.e $\Delta X_{i,t,LAD} = c + b_{11} X_{i,t-1,LAD} + b_{12} X_{i,t-1,SBO}$, it produces a significantly lower BIC (-53,973.26) than the constant only model's BIC (-52,822.13), with an R^2 of 8.75% and p-value for the differences in BIC of 0.000.

Table 3.6: Results related to model selection for odds changes at SBOBet. This table reports the Bayesian Information Criterion (BIC) for models of the form in Equation (3b), selected with various lags of bookmaker odds. The first column shows the right-hand side of the non-nested model under consideration, with one, two, and three lags of each bookmakers' odds used, respectively. The second and third columns report the BIC and R^2 of the non-nested model, respectively. The fourth column shows the nested model under consideration, which in each case is the nested model having produced the lowest current BIC value. The fifth column reports the BIC of the nested model. The final column reports the significance of the difference in BIC values, calculated as P-value = $\exp\{(\mathrm{BIC}_{\mathrm{lower}} - \mathrm{BIC}_{\mathrm{higher}})/2\}$.

Panel A: Model Selection for Odds Changes at SBOBet.

Dependent Variable: Change in Prices at SBO ($\Delta X_{i,t,SBO}$)

Non-Nested Model	BIC	Model R^2	Nested Model	BIC	P-value (Lower BIC model better than Higher BIC model)
$c + b_{11}X_{i,t-1,LAD} + b_{12}X_{i,t-1,SBO}$	-9,062.76	0.20%	c	-9,075.40	0.002
$c + b_{11}X_{i,t-1,LAD} + b_{12}X_{i,t-1,SBO} + b_{21}X_{i,t-2,LAD} + b_{22}X_{i,t-2,SBO}$	-6,068.26	0.30%	c	-6,082.35	0.001
$c + b_{11}X_{i,t-1,LAD} + b_{12}X_{i,t-1,SBO} + b_{21}X_{i,t-2,LAD} + b_{22}X_{i,t-2,SBO} + b_{31}X_{i,t-3,LAD} + b_{32}X_{i,t-3,SBO}$	-3,509.50	0.55%	c	-3,510.97	0.479

Panel B: Model Selection for Odds Changes at Ladbrokes.

Dependent Variable: Change in Prices at LAD ($\Delta X_{i,t,LAD}$)

Non-Nested Model	BIC	Model R^2	Nested Model	BIC	P-value (Lower BIC model better than Higher BIC model)
$c + b_{11}X_{i,t-1,LAD} + b_{12}X_{i,t-1,SBO}$	-53,973.26	8.75%	c	-52,822.13	0.000
$c + b_{11}X_{i,t-1,LAD} + b_{12}X_{i,t-1,SBO} + b_{21}X_{i,t-2,LAD} + b_{22}X_{i,t-2,SBO}$	-45,102.13	9.81%	$c + b_{11}X_{i,t-1,LAD} + b_{12}X_{i,t-1,SBO}$	-45,051.12	0.000
$c + b_{11}X_{i,t-1,LAD} + b_{12}X_{i,t-1,SBO} + b_{21}X_{i,t-2,LAD} + b_{22}X_{i,t-2,SBO} + b_{31}X_{i,t-3,LAD} + b_{32}X_{i,t-3,SBO}$	-37,326.64	10.08 %	$c + b_{11}X_{i,t-1,LAD} + b_{12}X_{i,t-1,SBO}$	-37,242.05	0.000

Examination of the BICs of the models estimated for changes in Ladbrokes odds shows that there may be marginal improvement over the single-lag model by adding in second- and third-lags of both Ladbrokes and SBOBet odds. For example, the second row of Panel B of Table 3.6 shows that the BIC of the model with two lags is -45,102.13 compared with the BIC of the model with a single lag of -45,051.12. However, the marginal improvement in model fit, although significant (p-value of 0.000) reduces the size of our data set significantly (each additional lag reduces the size of our sample by 12.5%). Consequently, we retain the single lag specification for modelling the changes in Ladbrokes odds. This choice does not affect the assertions drawn for the purpose of this study regarding the influence of book-balancers odds on position-takers’ odds. It will be interesting though for a future study to investigate odds moves on a larger number of points in time per game, in order to analyse potential information signalled by a higher ordered model.

Table 3.7 presents the results of estimating the random-effects model in Equation (4a), using one time period lagged odds from both bookmakers as the independent variables. Upon observation, the coefficients for the lagged odds for the two bookmakers are nearly identical in magnitude, but with opposing signs. Rearranging the model we would end up with:

$$X_{i,t,LAD} - X_{i,t-1,LAD} = c + b_{11}X_{i,t-1,LAD} + b_{12}X_{i,t-1,SBO} + U_i + e_{i,t} \quad (9a)$$

$$X_{i,t,LAD} = c + (1 + b_{11})X_{i,t-1,LAD} + b_{12}X_{i,t-1,SBO} + U_i + e_{i,t} \quad (9b)$$

Now, if $b_{12} \approx -b_{11}$

$$X_{i,t,LAD} = c + (1 - b_{12})X_{i,t-1,LAD} + b_{12}X_{i,t-1,SBO} + U_i + e_{i,t} \quad (9c)$$

Our model therefore predicts that at any time in the lead-up to kickoff, for the matches in our data set, around 9.5% of the odds at Ladbrokes can be explained by the lagged odds at SBOBet, while the remaining 90.5% of the odds are explained by the lagged odds at Ladbrokes. The 9.5% component of Ladbrokes odds due to SBOBets' lagged odds is positive and significant at the 1% level. Hence, we can conclude that Ladbrokes' odds converge towards the odds of SBOBet, which is in line with hypothesis 2a.

This model produces a relatively low R^2 value of 8.74%, partially due to the fact that in most cases $\Delta X_{i,t,LAD} = 0$, as noted earlier. This leads us to undertake further analysis of the cases in which Ladbrokes' odds have moved, that is, in the situations for which $\Delta X_{i,t,LAD} \neq 0$.

Table 3.7: Results of estimating the random effects panel model for full sample of 1,868 European football matches. This table reports the coefficients resulting from estimating the following regression model (equation 4a with one lag): $\Delta X_{i,t,LAD} = c + b_{11}X_{i,t-1,LAD} + b_{12}X_{i,t-1,SBO} + U_i + e_{i,t}$, where $X_{i,t,k}$ indicates the gross odds X offered on game i at time t by bookmaker k, on the favoured outcome. The coefficient and corresponding standard errors are reported in the second column, the third column shows the significance level of each coefficient in the regression: (*), (**), and (***) denote significance at the 10%, 5%, and 1% level, respectively. The fourth column shows the p-value of the Z- statistic. The standard deviation due to the random-effects design is σ_U and the standard deviation due to the white noise error term is σ_E.

	Coefficient (Std. Error)	Sig.	Z-Stat. (P-value)	
$X_{i,t-1,LAD}$	-0.09515 (0.0029)	(***)	-33.32 (0.0000)	
$X_{i,t-1,SBO}$	0.09150 (0.0027)	(***)	34.44 (0.0000)	
c	0.002145 (0.0011)	(**)	1.91 (0.0560)	
σ_U	0.0059			
σ_e	0.0266			
ρ (Fraction of Variance due to U_i)	0.0474		R^2	
n. observations	12,785		Within	0.0933
n. groups	1,868		Between	0.1690
Wald $X^2(2)$	1,192.03	(***)	Overall	0.0874

The result of the estimation is presented in Table 3.8a. It is noted that the number of data points has decreased from 12,785 in the full sample of 1,868 matches, to 1,572 odds movements at Ladbrokes, across 966 different games. The increase in the magnitude of the coefficients when examining cases for which Ladbrokes experiences a move is substantial; nearly 50% of the variation in Ladbrokes odds moves can be explained by the deviation of Ladbrokes' lagged odds from those of SBOBet. The R^2 of the model has increased significantly, from 8.74% to 49.89%. Isolating the cases where Ladbrokes (the position-taking bookmaker) experiences a move clearly shows that a key driver of their decision to change odds is the dispersion of their odds from those of the book-balancing bookmaker.

Overall, the results of the model estimations presented in Table 3.7 and Table 3.8a, suggest that from time period to time period prior to kick-off Ladbrokes is expected to move its odds about 9% on average in the direction of SBOBet's odds in the previous period. However, this 9% average movement is unevenly distributed. The 12,785 possible periods for which Ladbrokes could have experienced a move can be split into the 1,572 (12.3%) periods of time for which Ladbrokes odds did move, and the remaining 87.7 % of cases in which Ladbrokes odds do not change. Conditional upon a Ladbrokes price movement, about 51% of this odds movement is explained by the lagged SBOBet odds, which suggests strongly that Ladbrokes' odds are converging towards the lagged SBOBet odds.

Table 3.8a: Results of estimating the random effects panel model for sample of 966 European Football Matches at time points for which Ladbrokes odds moved, $\Delta X_{i,t,LAD} \neq 0$. This table reports the coefficients from estimating the following regression model (equation 4a with one lag): $\Delta X_{i,t,LAD} = c + b_{11} X_{i,t-1,LAD} + b_{12} X_{i,t-1,SBO} + U_i + e_{i,t}$, where $X_{i,t,k}$ indicates the gross odds X offered on game i at time t by bookmaker k, on the favoured outcome. The coefficient and corresponding standard errors are reported in the second column, the third column shows the significance level of each coefficient in the regression: (*), (**), and (***) denote significance at the 10%, 5%, and 1% level, respectively. The fourth column shows the p-value of the statistic. The standard deviation due to the random-effects design is σ_U and the standard deviation due to the white noise error term is σ_e.

	Coefficient (Std. Error)	Sig.	Z-Stat. (P-value)	
$X_{i,t-1,LAD}$	0.5146 (0.0136)	(***)	-35.19 (0.0000)	
$X_{i,t-1,SBO}$	-0.5225 (0.0148)	(***)	37.95 (0.0000)	
c	-0.0088 (0.0083)		-1.06 (0.2880)	
σ_U	0.0422			
σ_e	0.0442			
ρ (Fraction of Variance due to U_i)	0.4767		R^2	
n. observations	1,572		Within	0.3280
n. groups	966		Between	0.5819
Wald $X^2(2)$	1,458.69	(***)	Overall	0.4989

As an additional specification, to further test Hypothesis 2a, we model the change in Ladbrokes odds, for the restricted sample of 1,572 cases in which their odds have moved, against the lagged difference in SBOBet and Ladbrokes prices ($X_{i,t-1,SBO} - X_{i,t-1,LAD}$) directly. The results of estimating this model are presented in Table 6b. The single-factor specification shows a similar magnitude of coefficients; around 50% of the movement in Ladbrokes' odds (conditional upon a move in Ladbrokes odds) can be explained by the lagged difference in SBOBet and Ladbrokes odds.

Table 3.8b: Results of estimating the random effects panel model for the sample of 966 European Football Matches at time points for which Ladbrokes odds did move, $\Delta X_{i,t,LAD} \neq 0$. This table reports the coefficients from the following regression model (equation 4a with one lag): $\Delta X_{i,t,LAD} = c + B(X_{i,t-1,SBO} - X_{i,t-1,LAD}) + U_i + e_{i,t}$, where$X_{i,t,k}$ indicates the gross odds X offered on game i at time t by bookmaker k, on the favoured outcome. The coefficient and corresponding standard errors are reported in the second column, the third column shows the significance level of each coefficient in the regression: (*), (**), and (***) denote significance at the 10%, 5%, and 1% level, respectively. The fourth column shows the p-value of the Z-statistic. The standard deviation due to the random-effects design is σ_U and the standard deviation due to the white noise error term is σ_E.

	Coefficient (Std. Error)	Sig.	Z-Stat. (P-value)	
$X_{i,t-1,SBO} - X_{i,t-1,LAD}$	0.5038 (0.0129)	(***)	38.91 (0.0000)	
c	-0.0234 (0.0018)	(***)	-12.80 (0.0000)	
σ_U	0.0367			
σ_e	0.0503			
ρ (Fraction of Variance due to U_i)	0.3476		R^2	
n. observations	1,572		Within	0.3210
n. groups	966		Between	0.5828
Wald $X^2(2)$	1,513.77	(***)	Overall	0.5000

Interestingly, Table 6b shows a significantly negative sign for the constant on the random-effects regression. This can be explained by the fact that Ladbrokes payoffs are typically slightly lower than SBOBets' corresponding payoffs for the same favoured outcome (noting the differences in over-round presented earlier). Thus, if

Ladbrokes prices were below (above) the SBOBet odds in the previous time period, we expect less (more) than half of the difference to be made up by the change in odds at Ladbrokes. For example, if at time $t-1$, SBOBet and Ladbrokes were offering gross odds of \$1.90 and \$1.70, respectively, the model in Table 6b predicts that Ladbrokes' prices at time t would be \$1.70 + 0.5038(\$1.90 - \$1.70) – 0.0234 ≈ \$1.78. The predicted price would also be \$1.78 at time t if Ladbrokes were offering \$1.90 and SBOBet were offering \$1.70 at time $t-1$. Hence, in accordance with hypothesis 2a, when a position-taking bookmaker, such as Ladbrokes, move their prices, they do so towards the direction implied by book-balancers' odds. This indicates that information arising from skilled bettors is indirectly transmitted to position-taking bookmakers, by them following the trends in book-balancers' prices.

3.5.3. The Relative Efficiency of Odds-Based Forecasts

Hypothesis 3c predicts that the early odds at the book-balancing bookmaker (SBOBet) will provide information to the market, which is reflected in the final prices of the position-taking bookmaker (Ladbrokes), but the converse will not be true. Tables 7 and 8 present the results of the estimation of Equations 7a and 7b, respectively, using last-minute odds of Ladbrokes (Table 3.9) and SBOBet (Table 3.10) as the dependent variables.

Table 3.9: Results of estimating the fixed effects regression model for full sample of 1,700 European football matches, modelling the closing odds-implied probabilities of Ladbrokes $Y_{i,T,LAD,o}$ against the day-ahead odds-implied probabilities of Ladbrokes and the day-ahead odds-implied probabilities of SBOBet. This table reports the coefficients from the following regression model (equation 7a): $Y_{i,T,LAD,o} = c + \beta_1 Y_{i,t_0,LAD,o} + \beta_2 Y_{i,t_0,SBO,o} + \alpha_i + u_{io}$, where $Y_{i,t,k,o}$ indicates the odds-implied probability Y offered on game i at time t (t_0 being the time one day before kickoff, T being the time one second before kickoff) by bookmaker k, on all match outcomes o. The coefficient and corresponding standard errors are reported in the second column, the third column shows the significance level of each coefficient in the regression: (*), (**), and (***) denote significance at the 10%, 5%, and 1% level, respectively. The fourth column shows the p-value of the statistic. The standard deviation due to the fixed-effects design is σ_A and the standard deviation due to the white noise error term is σ_E.

Bookmaker's Odds-Implied Probabilities One Day Before Kickoff	Coefficient (Std. Error)	Sig.	P-value	
$Y_{i,t_0,SBO,o}$	0.4618 (0.0252)	(***)	(0.0000)	
$Y_{i,t_0,LAD,o}$	0.5547 (0.0248)	(***)	(0.0000)	
c	-0.0040 (0.0010)	(***)	(0.0000)	
σ_A	0.0060			
σ_e	0.0236			
ρ (Fraction of Variance due to α_i)	0.0612		R^2	
n. observations	3,397		Within	0.9921
n. groups	1,700		Between	0.9426
F(2, 1695)	105,769.98	(***)	Overall	0.9912

Table 3.10: Results of estimating the fixed effects regression model for full sample of 1,700 European football matches, modelling the closing odds-implied probabilities of SBOBet $Y_{i,T,SBO,o}$ against the day-ahead odds-implied probabilities of Ladbrokes and the day-ahead odds-implied probabilities of SBOBet. This table reports the coefficients from the following regression model (equation 7b): $Y_{i,T,SBO,o} = c + \beta_1 Y_{i,t_0,SBO,o} + \beta_2 Y_{i,t_0,LAD,o} + \alpha_i + u_{io}$, where $Y_{i,t,k,o}$ indicates the odds-implied probability Y offered on game i at time t (t_0 being the time one day before kickoff, T being the time one second before kickoff) by bookmaker k, on all match outcomes o. The coefficient and corresponding standard errors are reported in the second column, the third column shows the significance level of each coefficient in the regression: (*), (**), and (***) denote significance at the 10%, 5%, and 1% level, respectively. The fourth column shows the p-value of the statistic. The standard deviation due to the fixed-effects design is σ_A and the standard deviation due to the white noise error term is σ_e.

Bookmaker's Odds-Implied Probabilities One Day Before Kickoff	Coefficient (Std. Error)	Sig.	P-value	
$Y_{i,t_0,SBO,o}$	1.0111 (0.0506)	(***)	(0.0000)	
$Y_{i,t_0,LAD,o}$	-0.0074 (0.0498)		(0.8810)	
c	-0.0034 (0.0020)	(*)	(0.0810)	
σ_A	0.0140			
σ_e	0.0475			
ρ (Fraction of Variance due to α_i)	0.0796		R^2	
n. observations	3,397		Within	0.9673
n. groups	1,700		Between	0.7399
F(2, 1695)	25,069.49	(***)	Overall	0.9624

It is evident that early odds posted by SBOBet do exhibit significant forecasting power in predicting closing Ladbrokes odds, whereas the converse is not true. Table 3.9 shows that the early odds-implied probability of match outcomes from SBOBet, $Y_{i,t_0,SBO,o}$, is highly significant in the model predicting the closing Ladbrokes odds implied probability, $Y_{i,T,LAD,o}$. The magnitude of the coefficient, 0.4618, indicates that the late Ladbrokes odds converge, on average, nearly halfway towards the early SBOBet odds.

The results displayed in Table 3.10 show that the late odds of SBOBet are unrelated to the early odds of Ladbrokes. Ladbrokes price adjustment seems to be influenced by the prices in the book-balancing market, whereas SBOBet's odds moves are not predictable, which confirms the prediction of Hypothesis 2c.

Table 3.11: Results of estimating the fixed effects regression model for full sample of 1,700 European Football Matches, modelling the closing odds-implied probabilities of Ladbrokes $Y_{i,T,LAD,o}$ against the day-ahead odds-implied probabilities of Ladbrokes. This table reports the coefficients from the following regression model (equation 7b): $Y_{i,T,LAD,o} = c + \beta_1 Y_{i,t_0,LAD,o} + \alpha_i + u_{io}$, where $Y_{i,t,k,o}$ indicates the odds-implied probability Y offered on game i at time t (t_0 being the time one day before kickoff, T being the time one second before kickoff) by bookmaker k, on all match outcomes o. The coefficient and corresponding standard errors are reported in the second column, the third column shows the significance level of each coefficient in the regression: (*), (**), and (***) denote significance at the 10%, 5%, and 1% level, respectively. The fourth column shows the p-value of the statistic. The standard deviation due to the fixed-effects design is σ_A and the standard deviation due to the white noise error term is σ_E.

Bookmaker's Odds-Implied Probabilities One Day Before Kickoff	Coefficient (Std. Error)	Sig.	P-value	
$Y_{i,t_0,LAD,o}$	1.0076 (0.0024)	(***)	(0.0000)	
c	-0.0026 (0.0011)	(**)	(0.0140)	
σ_A	0.0057			
σ_e	0.0258			
ρ (Fraction of Variance due to α_i)	0.0469		R^2	
n. observations	3,397		Within	0.9905
n. groups	1,700		Between	0.9486
F(2, 1695)	176,286.85	(***)	Overall	0.9897

As a further robustness check, we present in Table 3.11 the results of estimating a fixed effects model of the closing Ladbrokes odds-implied probabilities at kick-off as a function of the early Ladbrokes' odds implied probabilities only. As one would expect, the coefficient of the early Ladbrokes probabilities is very close to 1, when no other variables are added to the model. However, when comparing this model to that presented in Table 3.9,the improvement in forecasting power by adding the SBOBet early odds-implied probabilities is clearly shown A likelihood-ratio test as to whether the model in Table 3.9 nests the model in Table 3.11, is emphatically rejected (test statistic $\chi^2(1)$ = 615.87, p-value = 0.000). Hence, we conclude that SBOBet odds are a significant determinant of the odds posted by Ladbrokes at kick-off.

Our tests of Hypotheses 3a and 3b are carried out using the conditional logit model described in the methodology section. We present, in Table 3.12, the results of estimating, using the early and late logged odds-implied probabilities from the two bookmakers Equations (8a) and (8b). These estimations are conducted in order to test the forecasting power of these odds for predicting the outcome of football games, In the conditional logit specification employed here, the magnitude of the coefficient is expected to be close to 1 if the bookmakers' probabilities exhibit no favourite longshot bias (e.g. Bacon-Shone et al, 1992). The magnitude of the coefficients (around 1.10) and their standard errors (0.50) indicates that each of the odds sets imply a small bias towards pricing favourites, though this coefficient is only marginally significantly different to 1.

Although the odds-implied forecasts are highly correlated, we can clearly observe in the results presented in Table 3.12 that the Psuedo-R^2 is higher and the BIC lower, for later odds (bottom half of the table) than earlier odds (top half of the table), at both SBOBet and Ladbrokes. The BICs can be formally compared using the formula:

$$\text{P-value} = \exp\{\frac{BIC_{Lower} - BIC_{Higher}}{2}\} \qquad (10)$$

where BIC_{Lower} is the model with the lower observed BIC value, and BIC_{Higher} is the model with the higher observed BIC value. The improvement in the BIC value is significant going from early odds for both SBOBet (p-value =0.000) and Ladbrokes (p-value = 0.020). This confirms hypothesis 3a; later odds are indeed more efficient in the football betting markets for both types of bookmaker.

Table 3.12: Results of comparing the efficiency of bookmaker's odds using conditional logit models. This table reports the results of four separate conditional logit models, using all outcomes of 1,700 European football matches. The modelling is conducted using the method of Equations (8a) and (8b): $\mathrm{P}\left(E_{i,t,k,o}=1\right)=\frac{e^{Z_{i,t,k,o}}}{\sum_{o=1}^{3} e^{Z_{i,t,k,o}}}$, where $E_{i,t,k,o}$ is the event home win, draw, or away win (one of which is 1 for each game i), with odds from bookmaker k at time t for outcome o. $Z_{i,t,k,o} = b\ \ln\left(Y_{i,t,k,o}\right)$ are the logged odds-implied probabilities for each game i, with odds from bookmaker k at time t for outcome o.

Bookmaker (Time Before Kickoff)	Coefficient (Std. Error)	P-value	Pseudo-R^2	Bookmaker (Time Before Kickoff)	Coefficient (Std. Error)	P-value	Pseudo-R^2
$Y_{i,t_0,SBO,o}$ SBO (-1 day)	1.1283 (0.0588)	(0.000)	0.1087	$Y_{i,t_0,LAD,o}$ LAD (- 1 day)	1.0996 (0.0571)	(0.000)	0.1072
BIC	3,623.7			BIC	3,629.7		
$Y_{i,T,SBO,o}$ SBO (-1 second)	1.1111 (0.0572)	(0.000)	0.1123	$Y_{i,T,LAD,o}$ LAD (-1 second)	1.0963 (0.0565)	(0.000)	0.1092
BIC	3,609.3			BIC	3,621.9		

The second comparison we make, to test hypothesis 3b, is of the efficiency of the late odds of Ladbrokes against SBOBet. The results of estimating the conditional logit models including, respectively, Ladbrokes against SBOBet odds shows that there is a relatively greater increase in efficiency of the odds from early to late (the pseudo R^2 of 0.1087 using early odds increases to 0.1123 using late odds) at SBOBet, compared with the increase in efficiency at Ladbrokes (the pseudo R^2 increases from 0.1072 to 0.1092). The difference in the efficiency of the late odds of these two bookmakers is is substantial. Again using the BIC comparison test from Equation (10), the BIC of SBOBet's kick-off odds is significantly lower than that of Ladbrokes' kick-off odds (p-value = 0.000). Thus, Hypothesis 3b, that late odds at SBOBet (the book-balancing bookmaker) are more efficient in predicting game outcomes than the late odds at Ladbrokes (the position-taking bookmaker) is confirmed. Consequently, the results presented which support hypotheses 3a and 3b, demonstrate that prices in both markets move towards efficiency. This can seen because SBObet seems to be improving the accuracy of its odds by responding to information arising from informed traders, and Ladbrokes wisely reacts to trends in the book-balancers' prices, improving the predictive ability of its odds in turn.

3.6. Discussion

This study has explored the new, globalised world of football betting markets, exploring odds movements and efficiency at bookmakers we classify as position-takers and book-balancers. We analyse a unique data set of these bookmaker prices collected for home, draw, and away outcomes in European football at nine points in time prior to kickoff.

The contributions of this research can be summarised as follows. We provide evidence that bookmakers are heterogeneous in type, and can be classified as position-taking bookmakers and book-balancing bookmakers. Position-taking bookmakers, mainly based in Europe, aim to attract unsophisticated clientele, and are willing to lose money in the short-run by either providing incentives to bet or promotional odds, to earn profits against its customer base as a whole in the long-run. They actively maintain a book of unsophisticated clients by restricting or excluding those bettors who are believed to be superior traders. The position-taking bookmaker hence operates against relatively uninformed clientele, who place small bets at high margins. Book-balancing bookmakers, on the other hand seem happy to attract the bets of sophisticated clients by operating under a regime of high-turnover and low margins. Staking limits are reported to be orders of magnitude higher. Prices change almost continuously in response to the volume of bets made by clients as the bookmaker moves its prices in order to achieve a low, risk-free margin from a high volume of stakes. These bookmakers operate almost exclusively online, and are believed to be the top-level of a vast pyramid structure of betting through regionalized bookmakers in the Asian market.

The results in this paper confirm that the position-taking bookmaker (Ladbrokes) charges higher transaction costs in the form of the over-round, and moves prices less frequently than the book-balancing bookmaker (SBOBet). When Ladbrokes prices do move, they tend to converge to SBOBet's odds; lagged SBOBet prices can explain around 50% of Ladbrokes prices movements. This is consistent with sophisticated traders moving the book-balancers' prices, and this information slowly diffusing to the odds of position-taking bookmakers. The fact that book-balancers seem to encourage access to sophisticated traders results in significant information regarding the outcomes of events to be transmitted from the betting public to their odds earlier compared to the odds. This practice is in contrast to that of position-taking, who discourage the activity of informed traders. However, the latter group of bookmakers tend to be influenced by trends in the book-balancers' odds, attempting to effectively benefit (indirectly) from the flow of *smart money* in order to increase the efficiency of their odds. As a consequence, they potentially gain a higher margin from their casual clients.

Taken as a whole, our results suggest a flaw in the statements of Levitt (2004), that bookmakers (the supply side) set the market and are superior forecasters compared to bettors (the demand side). According to Levitt's proposition, which conforms with the operation of position-taking bookmakers, such operators should have no incentive to react to price moves arising from a demand-driven market, as the betting public should not be able to improve on the bookmaker's expert estimates. Consequently, even when these bookmakers do move their odds as a result of betting volume, Levitt would argue that this can only be the result of risk-aversion rather than maximization of expected profit. As a result, this reaction should lead to less accurate estimations, as experts' forecasts would now have been adjusted to the opinions of noise-traders. However, we show that the supply-side of the market is actually following a demand driven market (the book-balancers), potentially shaped by the stakes of informed traders. In other words, prices in these markets are effectively processing information from trading volumes, leading market prices towards efficiency. Thus, we conclude that the group of bettors overall does not solely consist of noise-traders and in fact bettors as whole are capable of improving the accuracy of the prices set by expert-forecasters.

Overall, therefore, our results suggest that a highly liquid, globalized market such as the football betting market is capable of efficiently processing information arising from diverse global sources. This finding has broader implications. For example, it strongly supports the value of prediction markets, since we show that as long as liquidity is present, volume weighted average prices constitute more accurate estimations of unknown true probabilities, compared to those of expert forecasters. Moreover, inferences could be drawn based on our study, regarding the pricing mechanism in Over-The-Counter financial markets, since it could be claimed that bookmakers are effectively acting as dealers in such markets. In that context, it would be interesting to investigate the impact on the efficiency of market prices of the dealer in these markets adopting a position-taking as opposed to a book-balancing approach. Finally, and importnatly, the outcome of our study suggests that setting barriers to trade, such as the discrimination against skilled players, is likely to lead markets away from efficiency.

References:

Asch, P., Malkiel, B. G., and Quandt R. E. (1984). Market Efficiency in Racetrack Betting. *Journal of Business*, 57, 165-75.

Bacon-Shone, J., Lo, V. S. Y., and Busche K. (1992). Logistic Analyses for Complicated Bets. *Research Report 11,* Department of Statistics, University of Hong Kong.

Baye, M. R., Morgan, J., and Scholten, P. A. (2004). Price Dispersion in the Small and in the Large: Evidence from an Internet Price Comparison Site. *Journal of Industrial Economics*, 52, 463-496.

Baylis, K., and Perloff, J.M. (2002). Price dispersion on the internet: good firms and bad firms. *Review of Industrial Organization*, 21, 305-324.

Benter, W. (1994). Computer Based Horse Race Handicapping and Wagering Systems: A Report. In D. B. Hausch, V. S. Y. Lo, and W. T. Ziemba (Eds.), *Efficiency of Racetrack Betting Markets* (pp. 193-198). London: Academic Press.

Bolton, R. N., and Chapman, R. G. (1986). Searching for Positive Returns at the Track: A Multinomial Logit Model for Handicapping Horseraces. *Management Science*, 32, 1040-1060.

Bryman, A., and Bell, E. (2003). Business Research Methods. Oxford:Univerity Press.

Cain, M., Law, D., and Lindley, D. (2000). The Construction of a Simple Book. *Journal of Risk and Uncertainty*, 20, 119-140.

Choi, I. (2001). Unit Root Tests for Panel Data. *Journal of International Money and Finance*, 20, 249-272.

Duffie, D. (Ed.) (2012). *Dark Markets: Asset Pricing and Information Transmission in Over-the-Counter Markets*. Princeton University Press.

Edelman, D., and O'Brian, N. (2004). Tote Arbitrage and Lock Opportunities in Racetrack Betting. *European Journal of Finance*, 10, 370-378.

Figlewski, S. (1979). Subjective Information and Market Efficiency in Betting Markets. *Journal of Political Economy*, 87, 75-89.

Fingleton, J., and Waldron, P. (1999). Optimal Determination of Bookmakers' Betting Odds: Theory and Tests. *Working Paper, Trinity College Dublin*.

Forrest, D. (2006). Sport And Gambling. In W. Andreff, and S. Szymanski (Eds.), *The Edward Elgar Companion to Sports Economics* (pp. 40-48). Edward Elgar.

Forrest, D. (2012). The Threat to Football from Betting-Related Corruption. *International Journal of Sports Finance*, 7, 99-116.

Forrest, D., and Simmons, R. (2008). Sentiment in the Betting Market on Spanish Football. *Applied Economics*, 40, 119-126.

Franck, E., Verbeek, E., and Nüesch, S. (2010). Prediction Accuracy of Different Market Structures – Bookmakers Versus a Betting Exchange. *International Journal of Forecasting*, 26, 448-459.

Franck, E., Verbeek, E., and S. Nüesch, E. (2013). Inter-Market Arbitrage In Betting. *Economica*, 80, 300-325.

Gainsbury, S., Sadeque, S., Mizerski, D., and Blaszczynski, A. (2012). Wagering in Australia: A Retrospective Behavioural Analysis of Betting Patterns Based on Player Account Data. *Journal of Gambling Business and Economics*, 6, 50-68.

Goddard, J. A. (2005). Regression Models for Forecasting Goals and Match Results in Association Football. *International Journal of Forecasting*, 21, 331-340.

Ho, T. S. Y., and Stoll, H. R. (1981). Optimal Dealer Pricing Under Transactions and Return Uncertainty. *Journal of Financial Economics*, 9, 47-73.

Ho, T. S. Y., and Stoll, H. R. (1983). The Dynamics of Dealer Markets Under Competition. *Journal of Finance,* 38, 1053-1074.

Hodges, S., and Lin, H. (2009). Fixed Odds Bookmaking with Stochastic Betting Demand. *Working Paper, Cass Business School, City College London*.

Humphreys, B. R. (2010). Prices, Point Spreads and Profits: Evidence from the National Football League. *Working Paper, University of Alberta*.

Institute of International and Strategic Relations (IRIS) Report (2012). Sports Betting and Corruption: How to Preserve the Integrity of Sport. Paris: L'Institutde Rélations Internationales et Stratégiques.

Kuypers, T. (2000). Information and Efficiency: An Empirical Study of a Fixed-Odds Betting Market. *Applied Economics,* 32, 1353-1363.

Levitt, S. D. (2004). Why are Gambling Markets Organised So Differently From Financial Markets? *The Economic Journal*, 114, 223-246.

Magee, S. (Ed.) (1990). *The Channel Four Book of Racing*. London: Sidgwick and Jackson Ltd.

Marshall, B. R. (2009). How Quickly is Temporary Market Inefficiency Removed? *Quarterly Review of Economics and Finance*, 49, 917-930.

McFadden, D. (1974). Conditional Logit Analysis of Qualitative Choice Behaviour. In P. Zaremba (Ed.), *Frontiers in Econometrics* (pp. 105-142). Academic Press.

Paul, R. J., and Weinbach, A. P. (2007). Does Sportsbook.com Set Pointspreads to Maximize Profit? Tests of the Levitt Model of Sportsbook Behavior. *Journal of Prediction Markets*, 1, 209-218.

Paul, R. J., and Weinbach, A. P. (2008). Price Setting in the NBA Gambling Market: Tests of the Levitt Model of Sportsbook Behavior. *International Journal of Sports Finance*, 3, 2-18.

Pope, P. F., and Peel, D. A. (1989). Prices and Efficiency in a Fixed-Odds Betting Market. *Economica*, 56, 223-241.

Salop, S., and Stiglitz, J. E. (1977). Bargains and Rip-offs: A Model of Monopolistically Competitive Price Dispersion. *Review of Economic Studies*, 44, 493-510.

Shin, H. S. (1991). Optimal Betting Odds Against Insider Traders. *The Economic Journal*, 101, 1179-1185.

Shin, H.S. (1992). Prices of State Contingent Claims with Insider Traders, and the Favourite-Longshot Bias. *The Economic Journal*, 102, 426-435.

Shin, H.S. (1993). Measuring the Incidence of Insider Trading in a Market for State-Contingent Claims. *The Economic Journal*, 103, 1141-1153.

Sidney, C. (Ed.) (2003). *The Art of Legging*. London: Rotex Publishing.

Smith, M., Paton, D., and Vaughan Williams, L. (2006). Market Efficiency in Person-to-Person Betting. *Economica*, 73, 673-689.

Smith, M., Paton, D., and Vaughan Williams, L. (2009). Do Bookmakers Possess Superior Skills to Bettors in Predicting Outcomes? *Journal of Economic Behavior and Organization*, 71, 539-549.

Stoll, H. R. (1978). The Supply of Dealer Services in Securities Markets. *Journal of Finance*, 33, 1133-1151.

Sung, M.-C., and Johnson, J. E. V. (2010). Revealing Weak-Form Inefficiency in a Market for State-Contingent Claims: The Importance of Market Ecology, Modelling Procedures, and Investment Strategies. *Economica*, 77, 128-147.

Thorp, E. O. (2000). The Kelly Criterion in Blackjack, Sports Betting, and the Stock Market. In O. Vancura, J. A. Cornelius, and W. R. Eadington (Eds.), *Finding the Edge: Mathematical Analysis of Casino Games* (pp. 163-213). Institute for the Study of Gambling and Commercial Gaming.

Varian, H., (1980). A Model of Sales. *American Economic Review*, 70, 651-659.

Veitch, P. (2009). *Enemy Number One: The Secrets of UK's Most Feared Professional Punter* (pp. 226-234). UK: CPI William Clowes Beccles NR34 7TL.

Woodland, B. M., and Woodland, L. M. (1991). The Effects of Risk-Aversion on Wagering: Point Spread Versus Odds. *Journal of Political Economy*, 99, 638-653.

Chapter 4

Arbitrage mirage: Exploring the extent to which apparent inefficiency in betting markets is an illusion

Abstract

This chapter explores the claim made in previous studies that the football betting market is weak form inefficient, to the extent that prices offered by competing bookmakers enable arbitrage to take place. To examine these claims, odds from a range of bookmakers are examined. In particular, the odds offered by bookmakers with different operating policies are explored, namely those of book-balancing and position-taking bookmakers. The odds for different outcomes of each of 2,132 matches in the six major European football leagues in the season 2012-13 are examined to identify opportunities for forming arbitrage portfolios by betting at the best available odds from different bookmakers. A linear optimisation program on standard Home/Away/Draw betting along with Asian Handicaps is employed to identify 545 arbitrage opportunities. Most of these arise from betting on favoured teams with the position-taking bookmaker, and hedging with the book-balancing bookmaker. These strategies are seemingly profitable, suggesting that the market is weak form inefficient. However, it is argued that these opportunities are not likely to be exploitable in practice because sophisticated traders, who are the most likely to be aware of such opportunities, could face restrictions from position-taking bookmakers. This arises because these bookmakers appear to deliberately set prices which intentionally deviate from efficient odds but they have in place a range of operating strategies which prevent systematic exploitation of the arbitrage opportunities which arise. As a result, we argue that conclusions regarding market efficiency in betting markets (and in wider financial markets) are risky without a full understanding of the operating policies of institutions in the market.

4.1. Introduction

Several studies have explored the efficiency of betting markets by examining evidence for arbitrage opportunities which may arise from disparate prices in alternate trading venues. This typically leads to the construction of a synthetic 'Dutch Book,' wherein a bettor can lock in a risk-free profit, regardless of the outcome of the event, by betting at the highest individual-outcome odds across a panel of bookmakers. As betting markets are increasingly accessible online, this would seemingly suggest an inefficient market. However, as shown in this chapter, arbitrage opportunities in betting markets may be illusory because they usually require bets with position-taking bookmakers. These bookmakers are inaccessible to most sophisticated gamblers (see for example, chapter 3) and actively manage their portfolio of clients by restricting these supposed informed bettors (Franck, Verbeek and Nüesch, 2013). The remaining counterparties for the position-taking bookmaker bet smaller amounts with higher margins, and are actively targeted by advertising and sign-up benefits including overly-generous odds. One method that position-taking bookmakers use to identify sophisticated gamblers is examining the relative pricing of their bets; bettors who mainly stake when the odds at the position-taker are relatively high can be easily identified and excluded.

Utilizing a sample of 2,132 games played in the 2012-13 season in six major European leagues, with odds drawn at intervals from two hours prior to kick-off from a panel of seven bookmakers, we demonstrate that cross-bookmaker arbitrage opportunities arise in 545 (or around 25% of all) matches. Three bookmakers are identified as book-balancers, who allow large trades and operate a low-margin, high-turnover strategy. The remaining four bookmakers in our sample are identified as position-takers. We construct risk-free portfolios from thirteen available offers in football betting markets (standard Home, Draw, and Away team betting (three offers), plus Asian Handicap bets (ten offers) with payoffs that can be constructed in terms of match outcomes) and show that 84% of the arbitrage opportunities require the bettor to place a stake with a position-taking bookmaker and a book-balancing bookmaker. It is shown that in the arbitrage portfolio, bets with the position-taking bookmaker are usually on the favoured outcome as opposed to the longshot. It is argued that this arises because position-taking bookmakers are likely to face greater competition for bets related to favoured teams (which are more popular among the gambling public), and may be willing to accept losses by holding positions against such propositions in order to maintain an active portfolio of clients. The bets on the favoured outcomes in the arbitrage portfolio tend to be profitable; the hedge on such bets with the book-balancing bookmaker realises average losses equivalent to the transaction cost. Consistent with this finding, we find that the odds from book-balancing bookmakers exhibit a lower degree of favourite-longshot bias than those of the position-taking bookmakers.

This study contributes to existing literature in a number of ways. It is the first to analyse fixed-odds arbitrage betting in an operationally realistic manner; the offered prices from bookmakers being measured contemporaneously and near to kick-off when markets have enough depth to provide meaningful economic returns to bettors. By contrast, previous studies (e.g. Pope and Peel, 1989; Deschamps and Gergaud, 2007) have assumed that odds from multiple bookmakers would have been available simultaneously and do not provide insight into market liquidity. We also show that the use of Asian Handicaps allows the bettor a greater number of assets with which to build an arbitrage portfolio, and demonstrate an appropriate linear program for undertaking the necessary calculations to construct an appropriate portfolio. Finally, we extend the findings from the previous chapter to show that position-taking bookmakers are more likely than the book-balancers to set inefficient odds. We provide a strong caveat, however, that the seemingly profitable opportunities presented here are unlikely to be exploitable by sophisticated traders due to the client-management strategies of the position-taking bookmakers.

The remainder of this chapter proceeds as follows: Section 2 reviews the literature on market efficiency in financial markets and contrasts the findings with the literature on betting markets. We specifically discuss arbitrage in both markets, and explain the mechanisms which prevent bettors from exploiting mis-priced outcomes. This discussion is employed to develop testable hypotheses in Section 4.3. The data set and methodology are discussed in Section 4.4., including the linear program used to construct arbitrage portfolios. The results are presented in Section 4.5.and these are discussed in Section 4.6.

4.2. Literature Review

4.2.1. Information Efficiency

Fama (1970) defined an efficient market one in which prices fully reflect all available information and identified markets as weak, semi-strong and strong-form efficient if, respectively, prices fully reflected market prices, all information is publicly available and information is restricted to a small group of investors. As a result, for weak form, semi-strong and strong-form efficient markets, it is not possible for investors to profit simply by utilizing, respectively, market price signals, any publicly available information and any information whatsoever (including private information). The Efficient Market Hypothesis (EMH) has been the object of extensive investigation in finance literature, where several studies suggest that inefficiencies exist (e.g. DeBondt and Thaler, 1985; Lakonishok et al., 1994; Ikenberry et al., 1995; Barberis et al., 1998). However, according to Fama (1998) such evidence is not sufficiently consistent to reject the EMH and, consequently, that there is insufficient evidence to require the creation of a more robust alternative theory to replace it.

The exploration of the EMH has also attracted significant attention in the betting market literature (e.g., see Sauer, 1998 and Vaughan Williams, 2005, for a literature review on information efficiency on betting markets; Oikonomidis and Johnson, 2010, for a review focusing on football). Along the lines set by Fama (1970), Sauer (1998) indicates that a betting market is considered efficient if opportunities for making profit are not possible (strong test) or if it is shown that there is significant systematic asymmetry in the expected returns of alternative strategies (weak test). In general, independent of the definition adopted, the main determinant of market efficiency is the degree to which the market odds reflect the unknown, true probabilities of event outcomes over a large sample. This is analogous to the efficiency in wider financial markets, where efficiency is determined by how representative market prices are of the value of underlying assets market prices. However, as discussed below, studies which explore the EMH should consider the structural idiosyncrasies of the betting market, so that successful inferences concerning market inefficiency are drawn. Levitt (2004), for example, notes that position-taking bookmakers strategically set odds in a manner to profit from the biases of uninformed bettors. Consequently, a test of betting market efficiency is really a joint test of bookmaker price-setting behaviour and bettor biases. This is unlike many financial markets where the trader can both purchase and sell at a given price.

4.2.2. Arbitrage

According to Fama's (1970) definition, weak-form efficiency implies the absence of arbitrage opportunities, namely the realization of guaranteed profits by "the simultaneous purchase and sale of the same, or essentially similar, security in two different markets for advantageously different prices" (Sharpe and Alexander,1990). Arbitrage is a key concept in finance, as it constitutes the force that drives prices to converge to a point where profits from arbitrage are zero after allowing for transaction costs (Jensen, 1978) and, thus where prices are driven towards fundamental values (Sharpe, 1964; Fama, 1965; Ross, 1976). In theory, arbitrage cases provide risk-free opportunities. However, as Shleifer and Vishny (1997) point out, in reality there are risks associated with the exploitation of arbitrage, since an instant price move is likely to eliminate the opportunity for the investor to fully hedge their position in a way that secures profit (i.e. perfectly "simultaneous" purchase and sale is not feasible in the real world). Consequently, when investigating arbitrage opportunities, exploitability issues have to be considered.

The literature concerning arbitrage prospects in betting markets shows mixed results concerning whether effective arbitrage is possible. Several studies examining horserace betting markets find that arbitrage opportunities do not exist. For example, Adams, Rusco, and Walls (2002), who focus on high volume pools and Gramm and Owens (2005), who consider on course and off course betting, find no real opportunities for arbitrage. However, some studies suggest that, arbitrage is possible in horserace betting markets (Willis, 1964; Hausch and Ziemba, 1990; Edelman and O'Brian, 2004). Nevertheless, it has to be noted that arbitrage opportunities in pari-mutuel markets cannot be categorized as risk-free, since the effective odds that the bettor eventually secures cannot be certain at the exact time of the bet, unless that bet is the last one for that race. Thus, it seems fair to classify these cases as quasi-arbitrage (Paton and Vaughan Williams, 2005) or risk-arbitrage opportunities (Lane andZiemba, 2004) where the dispersion of prices might lead to a betting opportunity but they are not certain of a guaranteed profit regardless of the event's outcome.

Quasi-arbitrage has also been investigated in football betting markets. Dechamps (2008) found, when analysing prices from different bookmakers in fixed odds betting, that outlying high odds indicate lower probability for the corresponding outcome to be realized compared to the market average. This contradicts evidence from the spread-betting market provided by Paton and Vaughan Williams (2005), who find that the average mid-point of the quoted spreads from different bookmakers estimates the real outcome more accurately compared to the outlying spread, to the extent that profitable trading against outlying spreads is possible. The possibility of "zero" risk arbitrage has also been explored in the fixed odds football betting market. In particular, studies have examined the extent to which the differences in the odds quoted by different bookmakers are adequate to guarantee profitable, fully hedged positions. For example, Pope and Peel (1989) show that arbitrage opportunities exist due to price dispersion across bookmakers. However, later research suggests that the degree of coordination between bookmakers has increased, possibly due to the emergence of professional arbitrageurs (Dixon and Pope, 2004). The latter finding is in line with further studies (Deschamps and Gergaud, 2007; Luckner and Weinhardt, 2008; Vlastakis, Dotsis and Markellos, 2009; Dechamps, 2008; Franck, Verbeek and Nüesch, 2010), which show that bookmakers' prices are fairly well aligned and as a result cross-bookmaker arbitrage opportunities rarely arise in the major football betting markets. Clearly, this finding is consistent with the EMH. It has to be noted though that such studies mainly utilize data regarding odds offered by traditional European bookmakers and studies that extend the investigation of price dispersion into a less homogeneous population of market operators reach different conclusions. For example, Marshall (2009) analyses data from 50 different bookmakers originating from a variety of jurisdictions, covering a range of sports (including football) between January 2003 and December 2005. He finds that 19,882 arbitrage opportunities existed overall in this period and that they lasted 15.75 minutes on average[27]. Franck et al. (2013) analyse bookmakers' odds in parallel to betting exchange odds and find that cross-market arbitrage opportunities exist for 19.2% of the events.

From the foregoing review it has become clear that conclusions regarding the existence of arbitrage opportunities are subject to the choice of the markets investigated, as the between-markets correlation of offers seems to vary. Therefore, in order to fully explore the extent to which a market is truly inefficient when apparent arbitrage opportunities are discovered, it is important to examine which market operators offer odds at the two extremes of the price spectrum and what type of characteristics they seem to exhibit regarding their price-setting.

4.2.3. The Bookmaker Market

Levitt (2004) argues that betting markets are organized very differently from financial markets. He suggests that, as the main providers of liquidity, bookmakers (unlike financial market makers) take large positions against their customers rather than matching sellers with buyers. This approach to bookmaking is examined in a number of other studies (e.g., Kuypers, 2000; Paul and Weinbach 2007, 2008; Humphreys, 2010). Franck et al. (2013) suggest that bookmakers do not only attempt to maximize their profit by taking positions against their customers, but also effectively choose the bettors against whom they take such positions. They achieve this by monitoring their trades and filtering out the ones profiled as potentially skilled. In this context, bookmakers may occasionally publish inefficient odds that are likely to attract customers knowing that they can always eliminate those who mainly bet on

[27] Given that Marshall obtained such data from a company supplying software that explores arbitrage opportunities, 15.75 minutes seems to be a considerable amount of time for the arbitrage opportunities to take to disappear, as one would expect the automation offered by such software to lead to high orders being immediately placed against outlying odds by its users. This should lead to an almost immediate correction (i.e. convergence of the outlying prices towards the market mean).

such odds. As a result, the harm caused by setting theoretically inefficient odds is minimized and is potentially lower compared to the gain in the size of the customer base. Analysing bookmaker odds in parallel to matched offers from the betting exchange market, the Franck, et al. (2013) find that significant price dispersion exists, which leads to the frequent emergence of inter-market arbitrage opportunities (i.e., taking the best bookmaker offer and hedging on the betting exchange). They also show that betting exchange odds are more accurate predictors of event outcomes compared to bookmaker odds (similar to Smith, Paton and Vaughan Williams 2006, 2009), which suggests that arbitrage opportunities emerge due to the bookmakers' "inefficient" pricing. Thus, they argue that given that it would be easy for bookmakers to align their odds with those offered in betting exchanges, the fact that they tend not to do so indicates that these "inefficiencies" are intentional. This is consistent with their theory that bookmakers do not set their odds in order to maximize the profit per game, but their long term overall profit (arising from a growing customer base).

The theory presented by Franck et al. (2013) is innovative as it links odds-setting with the bookmakers' option to filter their customer base. Such a potential tactic by bookmakers constitutes an interesting distinction between betting and financial markets, which should be accounted for when the efficiency of odds published by bookmakers is investigated. However, it could be claimed that due to the nature of arbitrage, which requires that different prices must be offered at the same time, there are a number of weaknesses in the evidence provided by Franck et al. (2013). First, the bookmaker odds which they analyse bear no timestamp; rather, the time of the offer is *assumed* and odds are *assumed* to be constant for a given time interval. This is not always true, even for position taking bookmakers (see chapter 3), and as a result the matched offers between the betting exchange and the bookmakers are not guaranteed to have coexisted at the same point in time (see Marshall [2009] regarding the duration of arbitrage opportunities). Second, the matched offers between the exchange and the bookmakers can be measured up to 2 days prior to the game's kick-off when the amount of money that can be staked in a betting exchange is very low (see Table 4.1), suggesting that even if arbitrage opportunities exist they do not constitute any significant inefficiency. Third, the betting exchange which is analysed is Betfair, and this exchange levies a very high commission on winning customers (referred to as 'premium charges'), the implication of this for the generation of a risk-free portfolio is ignored[28]. Fourth, even though the volume that one can stake in a betting exchange increases as the kick-off time approaches, it is shown below that there is significant variation in the size of the stakes that can be placed across games. This harms the homogeneity of the sample, as several apparent arbitrage opportunities may bear no economic significance if only a very low amount can be staked.

Consequently, it is important to test the proposition suggested by Franck et al. (2013) on a dataset that does not exhibit the limitations outlined above, in order to explore whether robust empirical evidence supports their theoretical proposition.

Unlike the studies which describe bookmakers as position takers, several researchers suggest that a bookmaker's objective is to balance their books, and as a result, secure profit independently to the event's outcome (Magee, 1990; Woodland and Woodland 1991, Hodges and Lin, 2009). Hence, a bookmaker operating in this manner, changes its odds frequently in order to equalize payouts across each possible outcome. Such bookmakers are in effect acting as market makers and therefore, they are to some extent, setting up a betting exchange type of market in which they guarantee liquidity. However, in betting exchanges, a bettor can only successfully place a bet if that bet is 'offered' by another bettor. A market making bookmaker will, in contrast, accept the bet on the spot and may then decide to move its odds in order to attract an equal order on the alternative event outcomes, and, as a result, balance its book. Franck et al. (2013) note that bookmakers provide the equivalent advantage in betting exchanges that dealers offer in auction markets- namely, guaranteed liquidity (Madhavan, 2000). Moreover, when it comes to market making bookmakers, problems related to discrimination against skilled bettors are less likely to appear, since their model is based on the maximization of volume rather on successful positions (see Forrest, 2012 and chapter 3 of this thesis for a description of the market makers' model).

[28]Betfair may withhold up to 60% out of winning bettors profits (see ttp://www.betfair.com/www/GBR/en/aboutUs/Betfair.Charges/)

In football betting, despite the fact that position taking and market making bookmakers coexist (for a detailed discussion see Chapter 3 of this thesis), the literature has exclusively focused on analysing odds offered by the former. This is surprising given the fact that the economic significance of the latter is probably greater, at least in terms of the size of transactions (Forrest, 2012).

Given the significant differences in how the two bookmaker business models operate, arbitrage opportunities between the two types of bookmakers could be regarded as inter-market opportunities. Hence, it could be argued that the examination of inter-market opportunities between these market makers offer a more suitable environment for the investigation of price dispersion between demand driven and supply oriented markets, than the betting exchange-market maker environment explored by Franck et al. (2013). The advantages of this market making setting being the consistent provision of liquidity and the absence of charges on profits, such as *Betfair's* Premium Charges,.

Consequently, in this paper we explore whether arbitrage opportunities exist between 'book balancing' and 'position taking bookmakers' odds and investigate the roots of any observed price diversion. We test the proposition introduced by Franck et al. (2013) that the existence of cross-market arbitrage opportunities is the product of structural differences between demand and supply driven markets and that it points to intentional "inefficient" pricing by bookmakers. This would indicate that the objective in their odds setting is not necessarily to solely reflect true outcome probabilities rather the odds also reflect the bookmakers' marketing strategies which are designed to acquire and retain customers. However, we argue, in contrast to Franck et al. (2013), that this approach to price setting cannot be generalized to all bookmakers. Rather, it only applies to those falling under the group of position takers. This could also explain the findings of Marshall (2009), who, as opposed to Franck et al. (2013), identifies the presence of arbitrage opportunities within the bookmaker market, but does not investigate the nature of the bookmakers where these opportunities usually arise. The proposition tested in this paper is that arbitrage opportunities are most likely to arise between position takers and market makers and not between two bookmakers of the same type.

In addition, we argue that the investigation of the EMH when using odds provided by position taking bookmakers may lead to biased conclusions. Specifically, as suggested by Franck et al. (2013), such market operators may intentionally set “inefficient” prices as a marketing strategy to attract customers. However, if the operating strategies of these bookmakers ensure that these prices cannot be systematically exploited by informed bettors (because their accounts are closed or the size of their bets are restricted), then any observed arbitrage opportunities are merely an illusion. Equally, this policy would lead to their deliberate ‘inefficient pricing’ strategy leading to a maximization of their profits as it would maximize their customer base and would, effectively, eliminate the bets of informed bettors. The consequence of this is that the cost of accepting well-monitored stakes on a low number of bets against “inefficient” odds is likely to be low compared to the benefit of acquiring and preserving customers, who might consider alternative betting outlets if attractive odds were never offered by the bookmaker. Consequently, given that position- taking bookmakers are likely to encompass such marketing related objectives in their odds setting, it could be claimed that seemingly “inefficient” odds may actually be very efficient if the odds-setters’ objectives are considered. Consequently, any conclusions which are reached regarding market efficiency based solely on the advertised position-taking bookmakers’ odds could be unreliable.

However, such complications in the assessment of market efficiency should not exist when employing the odds of book balancing bookmakers (see chapter 3 of this thesis for more detail), since their prices should be a more accurate reflection of betting volumes staked in the market. Consequently, odds arising from this market constitute more appropriate data for the testing of the EMH in the football betting markets.

4.3. Hypotheses development

We believe that there are two groups of bookmakers simultaneously operating in the betting market, but these two groups pursue different objectives. Consequently, we expect there to be occasions when prices in these markets are sufficiently different for arbitrage to appear possible. Such instances are expected to mainly arise between bookmakers from these different groups rather than between bookmakers from the same group. However, because position taking bookmakers effectively prevent skilled traders from exploiting these opportunities, such arbitrage is effectively non exploitable in the long run. Hence, we refer to these instances as *arbitrage-mirage*.

On the one hand, there are the book-balancing bookmakers, who attempt to maximize the volume traded on their books and move their odds in a way so that their books are fairly balanced; thereby minimizing their risk. The odds offered by these bookmakers are expected to be mainly driven by "smart money" [29] and to be well calibrated, since opportunities in this market are exploitable. Consequently, mis-pricings by these bookmakers are expected to be corrected, according to the EMH. On the other hand the position-taking bookmakers, whose objective seems to be to maximize their customer base (Franck, et al., 2013) rather than the expected profit per game appear to operate policies to deter or prevent skilled bettors from exploiting any mis-pricing(Franck et al., 2013). As a result of this policy the position-taking bookmakers can use their odds as a marketing tool to attract and retain customers (Marshall, 2009; Franck et al., 2013) and therefore, occasionally (even intentionally) take on bets exhibiting a negative expected value. This arises because it may be beneficial to offer these prices to attract more customers knowing that they can prevent skilled traders from exploiting any inaccuracies in their pricing.

The movement of the odds offered by book balancing bookmakers is far higher compared to that of position takers, as the former need to adjust their odds each time a high stake is placed in order to balance their books. By contrast, the position takers rely on their higher over-round in order to allow greater stability in their odds, as the

[29] For a description of a market-maker's model see http://www.pinnaclesports.com/betting-promotions/winners-welcome.aspx?ico=home&icl=box3

higher over-round provides a cushion against (possibly deliberate) inaccurate forecasts of match outcomes. Due to the more informed nature of book balancing bookmakers' customers, the volumes traded on particular outcomes are likely to be informative. As a result, a lagged reaction by position takers to trends in odds offered by the book balancing bookmakers should be expected from time to time, in order to move their odds towards efficiency and increase their edge against their casual customers (see Chapter 3). Hence, any movement of odds of position-taking bookmakers should not necessarily be regarded as the product of arbitrageurs' actions, since those are unlikely to have access to position-taking bookmakers (Franck et al., 2013) (see Chapter 3 of this thesis). Moreover, given the existence of several arbitrage software platforms (Marshall, 2009; Franck et al., 2013) if position takers were moving their odds due to arbitrageurs' actions, *arbitrage-mirage* opportunities should barely be visible.

As a result, in this interactive environment, in which relatively stable and frequently moving prices coexist, instances of significant price dispersion are anticipated. These are likely to occasionally lead to *arbitrage-mirage*. However, given the objectives of the two types of bookmakers, *arbitrage-mirage* is mainly expected to occur between two or more bookmakers of different type. In addition, these apparent arbitrage opportunities will effectively reflect the deviation of position takers' odds from the outcomes' objective probabilities, or at least from the probabilities incorporated in the odds of the book-balancing bookmakers (which are driven by informed traders' actions). Consequently, the consideration of bookmakers whose odds are less efficient predictors of event outcomes (i.e., position takers) is expected to lead to a significantly greater number of instances of *arbitrage-mirage* (cf., when the odds of book balancers are considered). If the empirical evidence supports this proposition then this will reinforce the view that position taking bookmakers are not solely focused on reflecting outcome probabilities in their odds. The important implication of this finding would be that testing propositions related to the EMH employing such odds may lead to biased conclusions.

In order to explore the validity of the proposition suggested above, we test the following three related hypotheses:

H1: There exist instances where the price dispersion in the betting market is adequate to generate *seemingly* risk-free opportunities for bettors to profit by simultaneously betting with different bookmakers on alternative outcomes related to the same event.

Levitt (2004) analysed data regarding bettors' volume derived from the inside of a major operator's book and found that favourites constitute more popular bets compared to longshots, since they seemed to attract significantly higher volumes. In addition, Forrest and Simmons (2008) and Franck et al. (2010) show that bookmakers are likely to inflate the odds for popular bets in order to sustain competition and to build/maintain their customer base. Since all of these studies analyse data from bookmakers that match the profile of position-takers, such bookmakers are anticipated to offer relatively better odds for (popular) favourites compared to (less popular) longshots in order to attract customers. Consequently, we expect that in most cases where apparent arbitrage opportunities exist, the position-taker will be posting the best offer for the favourite and a book balancing bookmaker will be posting the best offer for the longshot. To explore this further we test the following hypothesis:

H2: Apparent arbitrage opportunities most commonly arise between book-balancers and position-takers, by position-takers offering the highest odds for favourites and book-balancers offering the highest odds for the longshots.

In a situation where an apparent arbitrage opportunity exists, it might be that the bookmakers with the two extreme odds may, effectively, share the expected loss that will pay the bettor's certain win or it may be that the expected loss is solely borne by one bookmaker. In the latter case, it could be argued that the arbitrage is created by the bookmaker who offers odds at values which generates an expected loss (expected profit for the bettor). We argue that apparent arbitrage opportunities are most likely to occur when position-takers (who may for marketing reasons set purposely inefficient odds) do not respond fast enough to incorporate price informative trends signalled by informed money traded with book balancing bookmakers. However, it might be argued that simply showing position-takers' odds significantly deviate at times from the odds in the more dynamic book-balancers market does not prove that the arbitrage is arising as a result of their odds setting strategies. Rather, it might be suggested that position-takers are more capable than book balancers of calibrating the probabilities of event outcomes and, as a result, they stick with their estimates even when these differ from those of the book balancing bookmakers (Levitt, 2004). Consequently, it is important to explore whether odds offered by book balancing bookmakers are more representative of the objective probabilities of outcomes. Clearly, if book balancers' odds are more efficient, it could be argued that, if position takers were solely interested in efficient estimation of event probabilities, they could simply adjust their odds to those of the book balancers. However, if it is shown that the position taking bookmakers do not adjust their odds in this way then this is supportive of Franck et al.'s (2013) proposition that marketing considerations forming part of their odds-setting strategy. To examine these issues we test the following three related hypotheses:

H3a: Position-takers suffer losses when an apparent arbitrage opportunity exists. The expected profit of bets placed against such bookmakers is higher compared to that placed against book-balancers, when apparent arbitrage opportunities occur.

H3b: Book balancers' odds constitute more accurate predictors of event outcomes (cf. those of position-taking bookmakers).

Finally, since according to H2, position-takers are more often expected to offer the highest odds for the favourite in the apparent arbitrage opportunities, it is likely that such bookmakers' odds underestimate the favourite's winning probability.

H3c: Book-balancers' odds exhibit lower favourite-longshot bias than position-takers' odds.

4.4. Data and Methodology

4.4.1 Data

As indicated above, studies which find little evidence of arbitrage opportunities in football betting generally analyse a homogeneous sample of bookmakers, all of which could be classified as position-takers (e.g. William-Hill, Ladbrokes, Bet365, Stan-James). On the other hand, studies that analyse data arising from a more heterogeneous sample of market operators (which include book balancers) find evidence of significant price dispersion which generate apparent arbitrage opportunities (Marshall 2009; Franck et al., 2013). However, these studies do not make clear precisely where the opportunities arise from or if they could be confirmed to have existed in practice. For example, Marshall (2009) does not clarify whether the apparent arbitrage opportunities arose in major or minor leagues or in the odds of which type of bookmaker. In addition, it is not clear in Franck et al. (2013) that the offers constituting the apparent arbitrage opportunities actually coexisted and to what extent these were *exploitable* (i.e. what was the possible stake which could be placed at these odds[30]. This is an important consideration, because the staking limits on the betting exchange (*Betfair*) which was employed in this study are considerably smaller than those offered by the typical book balancing bookmaker. For example, Table 4.1 presents summary statistics concerning the amounts that could be staked[31] on *Betfair* and *Sbobet*, a major Asian (book-balancing) bookmaker, for two different points in time: 2 days and 2 hours prior to kick-off, considering the associated transaction costs. It is evident that in all cases, the amounts that can be traded with the bookmaker are higher against lower transaction costs and that the variance of the volume offers is significantly lower. Hence, it can be argued that the economic

---

[30]The bookmaker data used in their study concerns offers posted up to two days prior to kick-off, when the volumes that can be traded in a betting exchange are very low as we show.

[31]Bookmakers often set their limits in terms of payoffs (i.e. max. stake / (odds – 1)), as they are primarily interested in limiting their exposure; the amount that they will have to pay in the event that the bet is successful.

significance of any arbitrage opportunity is greater and more consistent if it largely results from mispricing in the odds related to the bookmaker rather than the betting exchange and for opportunities closer to kick-off.

Table 4.1: The data was collected for all 115 games played from 23/8/2013 until 1/9/2013 for the 6 leagues in our sample, in order to compare transaction costs[32] and the liquidity[33] in the most popular markets for the two operators (i.e. 1x2 for Betfair*vs* Asian Handicap (main line) for *Sbobet*).

Provider	Timing	No of Games	Median Staking Limit	St. Dev. Staking Limit	Median Transaction Cost
Betfair	2 days to K.O.	115	£739	£14,516	2.7%
Betfair	Close to K.O.	115	£11,592	£69,089	2.1%
Sbobet	2 days to K.O.	115	£8,890	£2,997	1.8%
Sbobet	Close to K.O.	115	£26,668	£6,930	1.7%

Note: Once a trader takes all offers on *Betfair's* screen they will have to wait for new offers to be provided by the public, even if they are willing to trade at lower odds. A bookmaker like *Sbobet* though, will immediately place a new offer with (potentially) slightly decreased odds, sustaining the provision of liquidity. As a result, the figures on the table are likely to understate the liquidity differences across the two markets. Moreover, it should be considered that there are a number of book-balancing bookmakers, such as the ones analysed in this chapter, who accept stakes of the same magnitude as *Sbobet*, On the other hand, other than *Betfair*, there is no betting-exchange, where significant volumes can be traded.

As discussed above, this study attempts to explore structural differences between position-taking and book-balancing bookmakers and therefore, the employment of data from major bookmakers representing both groups is essential. In

[32]For Betfair, the transaction cost for each offer is calculated as the over-round (sum of the inverse odds) considering the (net of Betfair's 2% commission on profit) volume-weighted average of odds for Backing (betting in favour of that team) and the volume-weighted average of odds for Laying (betting against that team). For Sbobet, the transaction cost for each game corresponds to the over-round considering the two offers for them main handicap line (e.g. +0.5 handicap team A and -0.5 handicap team B, on A vs B).

[33]As liquidity we consider the total amount that can be staked. Hence for Betfair, this is the sum of money available for all Backing and Laying offers on the screen and for Sbobet, the sum of the maximum stake allowed on both sides.

order to achieve this, we collected data from *Ladbrokes*, *William Hill*, *Bet365* and *Stan James*. *Ladbrokes* and *William Hill* are chosen as representatives of the leading, UK, 'position taking' bookmakers. Ladbrokes and William Hill were established in 1886 and 1934, respectively, and operate retail businesses with thousands of betting shops, mainly in the UK. They also operate online, and their combined aggregate gross revenue exceeds £1 Billion[34]. *Bet365* is a major UK-based online betting company founded in 2001, which in 2011 traded £8.5 Billion and achieved a gross profit of £422 Million[35]. *Stan James* is private company with a well established brand name, operating mainly online, while owning 65 betting shops in the UK[36]. We also collected data from three of the leading book balancer bookmakers: *Sbobet*, *IBCbet*, *Bet188*, and *Pinnaclesports*. The first three of these are the leading Asian bookmakers, handling enormous volumes, allegedly far higher than traded by more traditional European bookmakers (Forrest, 2012). *Pinnaclesports* is also a major online operator, purportedly trading billions of dollars[37]. *Pinnaclesports* are fairly open regarding their operation (unlike the other Asian bookmakers), which they describe on their website[38] as attempting to maximize trading volume while minimizing exposure, using information arising from smart money as a tool to set efficient odds. *Pinnaclesports* state that they are friendly to arbitrageurs, as the expected value of a trade for them should not depend on the motives of the counterparty placing the stake (e.g. if the bettor is an arbitrageur or professional trader). The model described by *Pinnaclesports* also fits the operations of the three other main Asian bookmakers listed above and is fully aligned with the model of book -balancing bookmakers described earlier.

---

[34] See William Hill (2013) "Preliminary Results 2012" and Ladbrokes (2013) "Preliminary Results for the Year Ended December 2012".

[35] Source: http://www.publications.parliament.uk/pa/cm201213/cmselect/cmcumeds/writev/1554/ga104.htm

[36] Source: http://howtobet.net/sportsbook-review/stan-james

[37] See Simon and Schuster (2006) The Smart Money http://www.washingtonpost.com/wp-dyn/content/article/2007/01/16/AR2007011601375_pf.html

[38] See: http://www.pinnaclesports.com/about-us.aspx, http://www.pinnaclesports.com/betting-promotions/arbitrage-friendly and http://www.pinnaclesports.com/betting-promotions/winners-welcome

When it comes to the investigation of arbitrage opportunities, it is vital to ensure that the odds offered by the different bookmakers that compose the arbitrage opportunity coexisted. Thus, we designed the data collection program in such a way that the requests to the bookmakers' websites were conducted simultaneously and forced a time-out of 30 seconds, in order to guarantee that the greater lag between the quickest and slowest response does not exceed this period[39]. All odds were collected in a period within 2 hours from kick-off, when the staking limits reach their peak. We focused on the major European football leagues in order to ensure that the findings of the study carry economic significance, as the volumes traded in leagues of lower status are significantly smaller. We collected data for the whole of the 2011/12 season for the 6 major leagues, namely the English Premier League, the German Bundesliga, the Italian Serie-A, the Spanish La Liga, the French Legue 1 and the Dutch Eredivisie. Overall, this resulted in a sample of 2,132 games.

We collected odds information on the major betting markets, namely, the Asian Handicap and the home-win, draw, away-win (1X2) markets. It should be noted that the former is significantly more popular among book-balancing bookmakers and the latter among position takers. The 1x2 market is one where bets are placed on which of the teams will win the game, with the draw being a possible outcome. In Asian Handicap betting one of the teams (usually the favourite) is handicapped by one, more than one or a fraction of a goal and a bet on this team is only successful if they win the game by a greater margin than this handicap. A bet on the opposing team is successful even if they lose by a margin lower than the handicap. The Asian Handicaps that are of interest for this study are those in the interval of -0.5 to +0.5, since these can generate an arbitrage opportunity with the 1x2 offers.

We denote a team as handicapped by -0.5, when they are deemed to start with a half goal deficit, so if that team does not win the bet loses. On the other hand, a team

---

[39]The cost of this was that several bookmakers occasionally failed to respond within the maximum allowed period. In these cases, we repeated the full request (i.e. for all bookmakers) 3 times in order to obtain a complete sample, but in some cases, due to high load on bookmakers' websites some would still fail to respond. In those cases, the odds of those who failed to respond were not considered, which may lead to a slight underestimation of the frequency of arbitrage opportunities overall. The way round this would be to increase the time-out, however, this would risk the integrity of the results overall, as a higher time interval would increase the chance of odds of the quickest responding bookmaker changing until the response of the slowest bookmaker came back.

which is handicapped 0.5 is deemed to start with a half goal advantage, so a bet on that team wins even if the outcome of the game is a draw. A bet on a team with a 0 handicap is refunded in the event of the draw, whereas a bet on team with a -0.25 (0.25) handicap is considered as a half-bet on 0 handicap and a half-bet on -0.5 (0.5). By retaining this simple approach we have a means of equating Asian handicap bets to 1X2 type bets. In particular, it is easy to see that the outcome of bets on these handicaps is completely determined by whether the team wins, loses or draws, exactly as it is the case with 1x2 bets. This would not be the case if teams with handicaps greater than one were included in the data set, since in this case the success of the Asian handicap bet would depend on the goal difference. Consequently, it is possible that a combination of 1x2 bets and bets on handicaps in the [-0.5, 0.5] region can create a fully hedged position if adequate price dispersion exists in the market.

4.4.2 Methodology

4.4.2.1 Exploring the occurrence of apparent arbitrage opportunities

In order to test hypothesis 1, namely that arbitrage opportunities appear to occur, it is important to develop a means of identifying those cases where sufficient price dispersion exists so that a fully hedged portfolio can be constructed (comprised of diversified bets simultaneously offered by different bookmakers). Investigating an arbitrage opportunity can be formulated as a linear optimization problem, where the objective is to decide the optimal distribution of stakes across different offers that maximizes the return. Arbitrage opportunities will be deemed to exist where such a return is positive and invariant towards all possible outcomes. Obviously, if there is not enough dispersion in the odds across the market to generate an arbitrage opportunity, there will not be a feasible solution to the problem. Providing there is sufficient dispersion, there will be a range of solutions that offer certain positive returns and the linear program will suggest the combination that offers the highest profit. The method below identifies how this problem is formulated and examined for a single game. This process is repeated for all 2,132 games in the sample.

Let $X_{j,k}$ denote the vector of gross odds offered by bookmaker k, where for each game there are $1 \leq j \leq 13$ odds offered on different types of bets (i.e., available markets, such as ‘home win’ ‘home win with a -.05 handicap’ away win with a0.5 handicap etc) as shown in Table 4.2:

Table 4.2: Gross return to a $1 stake for a single bookmaker on each potential match outcome for different types of bet, with odds vector $(X_1, \dots, X_{13})'$ indicating the gross payoff for each corresponding market.

j	Market	Return if Home Win	Return if Draw	Return if Away Win
1	Home Win	X_1	0	0
2	Home Win (-0.5) Handicap	X_2	0	0
3	Home Win (-0.25) Handicap	X_3	0.5	0
4	Home Win (0) Handicap	X_4	1	0
5	Home Win (+0.25) Handicap	X_5	1+0.5($X_5 - 1$)	0
6	Home Win (+0.5) Handicap	X_6	X_6	0
7	Draw	0	X_7	0
8	Away Win	0	0	X_8
9	Away Win (-0.5) Handicap	0	0	X_9
10	Away Win (-0.25) Handicap	0	0.5	X_{10}
11	Away Win (0) Handicap	0	1	X_{11}
12	Away Win (+0.25) Handicap	0	1+0.5($X_{12} - 1$)	X_{12}
13	Away Win (+0.5) Handicap	0	X_{13}	X_{13}

The odds in the Table 4.2 can by multiplied by stake size S to determine non-unit payouts. For example, if a bettor were to stake $5.00 on a Home Win on a +0.25 Handicap bet ($j = 5$) at gross odds of$2.20, and the match is drawn, their payoff (including the initial stake) would be $5 \times \big(1 + 0.5(2.20 - 1)\big) = 5 \times 1.60 =$ $8.00. Alternatively, the $5 bet on the Away Win on a (-0.25) Handicap bet ($j = 10$) would return $2.50, and the corresponding bet on the Away Win (0) Handicap bet ($j = 11$) returns the initial $5 to the bettor.

In seeking a solution to the linear programming problem we first search for the highest odds across the set of $k = 7$ bookmakers in each market. This allows us to search for cases in which a synthetic Dutch book could potentially be constructed. We define the vector X_{max} elementwise as $X_{j,max} = \max_k(X_{j,k})$. In cases where bookmakers are tied for the highest odds, we retain all possible combinations of maximum prices. For example, if both *Ladbrokes* and *William Hill* were offering gross odds of \$1.50 on a home win ($j = 1$) for a particular match, and this price was higher than all the other bookmakers' prices for $j = 1$, we would retain two X_{max} vectors, in order to not lose information for hypothesis H2.

Second, we aim to find the set of stakes that a bettor would place to *best* exploit potential arbitrage opportunities. Let S_j be the bettor's allocated stake for each bet type j. The profit function $Z = \{Z_{homewin}, Z_{draw}, Z_{awaywin}\}$ for each possible match outcome can be defined as:

$$Z_{homewin} = \sum_{j=1}^{6} S_j X_{j,max} - \sum_{j=1}^{13} S_j \quad (1)$$

$$Z_{draw} = S_7 X_{7,max} + 0.5(S_3 + S_{10}) + (S_4 + S_{11}) + 0.5\left(S_5 + S_{12} + S_5 X_{5,max} + S_{12} X_{12,max}\right) + \left(S_6 + S_{13} X_{13,max}\right) - \sum_{j=1}^{13} S_j \quad (2)$$

$$Z_{awaywin} = \sum_{j=8}^{13} S_j X_{j,max} - \sum_{j=1}^{13} S_j \quad (3)$$

To identify the best possible arbitrage opportunity one needs to find the distribution of stakes *S* that maximizes the payoff for any of the three match outcomes, subject to a set of constraints. Hence, the optimization identifies the distribution of stakes *S* that maximizes the payoff for any of the three match outcomes. The optimization can be

$$\textit{Find optimal strategy } S^* \textit{ by varying } S \textit{ that } Z_{homewin} \quad (4)$$

Subject to constraints:

$$Z_{homewin} = Z_{draw} \tag{5}$$

$$Z_{homewin} = Z_{awaywin} \tag{6}$$

$$Z_{homewin} > 0 \tag{7}$$

$$\sum_{j=1}^{13} S_j = 1 \tag{8}$$

$$S_j \geq 0 \; \forall j \tag{9}$$

Due to the linear nature of the problem, the simplex algorithm can be used in order to maximise the objective function (Dantzig, 1951). Constraints (5) and (6) ensure that the selected combination of stakes leads to the same return independently of the outcome. Constraint (7) implies that for the solution to be acceptable, the net return should be positive. Constraint (8) requires that the sum of stakes should equal 1, so that each S_j will represent the fraction of the available capital that should be staked on each bet type. Finally, constraint (9) requires that all stakes are positive. The optimisation will fail to find a feasible solution in the event that arbitrage is not possible for the given set of bet types on a given game. If there is more than one feasible solution per game, we would select the bet with the highest return per outcome.

By way of example, for the game between Mainz and Wolfsburg, played in Mainz's Coface Arena on the 24/08/2013 at 14:30 BST, the bookmaker *PinnacleSports* offered odds $(X_1, X_2, \ldots, X_{13})$ of:

$$X = (2.73, 2.72, 2.36, 1.99, 1.70, 1.54, 3.58, 2.67, 2.67, 2.31, 1.95, 1.68, 1.52)$$

at 8:53 BST on the day of the game. Based solely on the offers of this bookmaker and maximizing $Z_{homewin}$, subject to constraints (5 to (9), no feasible solution is found. This indicates that these offers are internally consistent and no arbitrage opportunities are available (i.e. no Dutch book exists across the odds offered on the various types of bet offered by *PinnacleSports* on this game). However, if a different bookmaker offers odds of X_3= 2.46 for Mainz on a Home Win (-0.25) Handicap ($j = 3$), X_{max} becomes

$$X_{max} = (2.73, 2.72, 2.46, 1.99, 1.7, 1.54, 3.58, 2.67, 2.67, 2.31, 1.95, 1.68, 1.52)$$

and the maximization routine yields an optimal solution of:

$$S^* = (0,0, \mathbf{0.4071}, 0,0,0, \mathbf{0.0794}, 0,0,0, \mathbf{0.5135}, 0,0).$$

In other words betting 40.71% of the bankroll on the Mainz (-0.25) Handicap bet, 51.35% on Wolfsburg (0) Handicap ($j = 11$) bet at odds of $X_{max,11} = 1.95$ and 7.94% of the bankroll on the draw at odds ($j = 7$) of $X_{max,7} = 3.58$, the bettor can secure a profit equal to 0.14% of the total investment, irrespective of the outcome of the game. Details of the return from various bets for this game are presented in Table 4.3:

Table 4.3: Example of Arbitrage Opportunity from Linear Program for game between Mainz and Wolfsburg, played in Mainz's Coface Arena on the 24/08/2013.

j	S^*	X_{max}	Return if Home Win	Return if Draw	Return if Away Win
1	0.0000	2.73	0.0000	0.0000	0.0000
2	0.0000	2.72	0.0000	0.0000	0.0000
3	0.4071	2.46	1.0014	0.2035	0.0000
4	0.0000	1.99	0.0000	0.0000	0.0000
5	0.0000	1.70	0.0000	0.0000	0.0000
6	0.0000	1.54	0.0000	0.0000	0.0000
7	0.0794	3.58	0.0000	0.2843	0.0000
8	0.0000	2.67	0.0000	0.0000	0.0000
9	0.0000	2.67	0.0000	0.0000	0.0000
10	0.0000	2.31	0.0000	0.0000	0.0000
11	0.5135	1.95	0.0000	0.5135	1.0014
12	0.0000	1.68	0.0000	0.0000	0.0000
13	0.0000	1.52	0.0000	0.0000	0.0000
SUM	**1.0000**		**1.0014**	**1.0014**	**1.0014**

In order to test hypothesis 1, that arbitrage opportunities appear to exist, we run this maximization for each game in the sample. For each game we consider the best odds for each offer, in order to uncover whether cross-bookmaker arbitrage opportunities exist in the football betting market close to the games' kick-off.

4.4.2.2. Exploring the source of apparent arbitrage opportunities

We now explore the methodology employed to test Hypothesis 2, namely that significantly more arbitrage opportunities exist when comparing odds simultaneously offered by book-balancing and position-taking bookmakers rather than when comparing odds offered by two or more bookmakers of the same type (i.e. book-balancing and position-taking, by position-takers usually offering the highest odds on the favourite and book-balancers on the longshot. To achieve this, those games where an arbitrage opportunity was identified are isolated and compare the frequency of instances in which the position taking bookmaker offers the highest odds for the 'favourite' (cf. the 'longshot'). The 'favourite' is considered to be the outcome for which the lowest odds are offered in the 1x2 market, by all the bookmakers. In the rare event that a different outcome was favoured by at least one bookmaker compared to the rest, we eliminate that observation. We then define variable D_f and D_l for each game i as follows:

$D_{fi} = 1$, if the position-taker is offering the highest odds for the favourite of game i.

$D_{fi} = 0$, otherwise.

$D_{li} = 1$, if the position-taker is offering the highest odds for the longshot of game i.

$D_{li} = 0$, otherwise.

Hence, over the sample of n games, in which an arbitrage opportunity arose, we can calculate the relative frequency of cases where the best offer for the favourite was offered by a position taking (cf book balancing) bookmaker as follows:

$$\widehat{p_f} = \frac{\sum_{i=1}^{n} D_{fi}}{n} \quad \text{and} \quad \widehat{p_l} = \frac{\sum_{i=1}^{n} D_{li}}{n} \quad (10)$$

We calculate the following z value, in order to test whether the frequency of the favourites' best odds being offered by position taking bookmaker is random:

$$z = \frac{\widehat{p_f} - \widehat{p_l}}{\sqrt{\hat{p}(1-\hat{p})(\frac{2}{n})}} \quad (11)$$

where

$$\hat{p} = \frac{\sum_{i=1}^{n} D_{fi} + \sum_{i=1}^{n} D_{li}}{2n}$$

The null hypothesis is that price dispersion is random and therefore, there is no systematic tendency from position taking bookmakers setting outlying odds for favourites (i.e. $\widehat{p_f} - \widehat{p_l}$).

4.4.2.3. Exploring which group of bookmakers is more likely to be responsible for the generation of the arbitrage opportunity

According to hypothesis 3a, when the two types of bookmakers offer sufficiently different odds to generate an arbitrage opportunity, the bets placed against position-taking bookmakers exhibit positive expected returns, significantly higher than for the bets against book-balancing bookmakers. In order to test this hypothesis, we conduct a betting simulation, where a unit stake ($1) is placed on each bet that is selected by equation (4) across the total sample of games. Hence for each type of bookmaker, we calculate for each game *i* and potential stake at offer *j* the bettor's profit Z_{ij}, as:

$$Z_{ij} = \sum_{j=1}^{6} S_{ij}X_{ij} + S_{i7}X_{i7} + 0.5(S_{i3} + S_{i10}) + (S_{i4} + S_{i11}) + 0.5(S_{i5} + S_{i12} + S_{i5}X_{i5} + S_{12}X_{i12}) \quad + (S_{i6} + S_{i13}X_{i13}) + \sum_{j=8}^{13} S_{ij}X_{ij} - \sum_{j=1}^{13} S_{ij} \qquad (12)$$

Where:

S_{ij}= 1, if for game *i* there is a bet on offer j (i.e. if S_j> 0 as set by optimization (4))

S_{ij}= 0 if for game *i* there is a bet on offer j (i.e. if S_j= 0 as set by optimization (4))

Xij represents the odds offered for offer *j* of game *i*

As a result, the average profit that the bettor achieves against each type of bookmaker, across the sample of *n* bets can be calculated as:

$$\mu = \sum_{ij=1}^{n} Z_{ij} \qquad (13)$$

Hence, according to hypothesis H3a, it should be that

$$\mu_{position-takers} > \mu_{book-balancers} \qquad (14)$$

It has to be clarified that the objective in this case is to test which bookmaker is more likely to be responsible for the generation of the arbitrage opportunity. Therefore, we are interested in comparing the expected profit that a stake on average exhibits against a book-balancing versus a position-taking bookmaker, when an arbitrage opportunity occurs. Hence, a unit stake is placed for each bet forming a part of the arbitrage portfolio. Thus, for the purpose of this hypothesis, the results of optimization (4) in terms of the capital distribution across different offers are only interesting in identifying those offers that are sufficiently different in order to form an arbitrage portfolio. The size of the stake assigned to each offer by equation (4) though is irrelevant for testing hypothesis H3a (as long as it is greater than 0), since the stakes' allocation in (4) concerns the equalization of payoffs across the possible outcomes.

However, it could be argued that the result of placing a unit stake across each bet is subject to high variance, since the average profit is highly influenced by the outcome of bets on longshots. Therefore, in order to ensure that latter bets are not leading to biased conclusions regarding the expected profit against each bookmaker, we replicate the simulation, where each stake $S_{ij}$is determined by the Kelly Criterion (Kelly, 1956). Here we assume that for each type of bookmaker, all bets bear equal expected profit and, as a result, the application of the Kelly Criterion results in weighting each bet disproportionally to its odds. Hence, for each selected bet j of game i instead of a unit stake, we bet $S_{ij} = \frac{1}{X_{ij}-1}$ (15). As a result the average realized profit against each bookmaker across the sample of n bets is $\mu = \frac{\sum_{ij=1}^{n} S_{ij} Z_{ij}}{\sum_{ij=1}^{n} S_{ij}}$ (16). Consequently, we recalculate the mean for each of the two types of bookmaker and compare them, in order to confirm that the conclusions drawn from the unit-stake simulations are not biased from abnormally positive or negative results on high-odds bets.

In order to test hypotheses 3b and 3c, namely that predictions based on book-balancers' odds are more efficient (3b) and unbiased (3c) predictors of event outcomes compared to predictions based on position-takers' odds, we compare the forecasting accuracy and the favourite longshot bias observed in predictions based on the odds of the two different types of bookmakers. To accomplish this we employ a conditional logistic regression (with the probability of outcome o derived from the odds as the sole independent variable), where the outcome of each game is the dependent variable (i.e. team 1 win, team 2 win or draw), which takes value 1 for the event that occurred and 0 for the events that did not occur. Hence, the probability that outcome o in game i occurs, is given by:

$$P(Y_{io} = 1) = \frac{e^{Z_{io}}}{\sum_{o=1}^{3} e^{Z_{io}}} \qquad (17)$$

Z_{io} is a function of the probability p_{io}, as the latter is the probability of the event outcome implied by the odds for each outcome *o* of game *i* (the superscript *s* here implying the subjective probability based on the bookmaker and bettors combined assessment of the chance of this outcome), such that:

$$Z_{io} = b \times \ln (p_{io}^{s})$$

where p_{io} can be calculated from the odds X_{io} of outcome *o* in game i as

$$p_{io} = 1/X_{io}(1 + \rho_i)$$

and ρ_i is the bookmakers over-round. This can be calculated from the odds offered for all outcomes *o* of game *i*

$$\rho_i = \sum_{i=1}^{3} 1/X_{io} - 1$$

Hence, equation (17) can be written as:

$$P(Y_{io} = 1) = \frac{e^{b \times \ln (p_{io}^{s})}}{\sum_{o=1}^{3} e^{b \times \ln (p_{io}^{s})}} = \frac{(p_{io}^{s})^{b}}{\sum_{o=1}^{3} (p_{io}^{s})^{b}} \qquad (18)$$

Positive favourite longshot bias indicates that the bookmaker odds underestimate the probability of the favoured event occurring. Therefore, if a bookmaker exhibits this bias, the actual winning probability of favourites, as implied by their observed frequency of success, is higher compared to that expected by the odds; whereas for the longshots it is lower. Thus, denoting as p_{io}^{v} the “true” probability (v denoting ‘verifiable’ or objective) of outcome *o* in game *i*, we can infer the following:

$$p_{if}^{s} > p_{il}^{s} \Rightarrow p_{if}^{v}/p_{if}^{v} > p_{il}^{v}/p_{il}^{s} \qquad (19)$$

where f denotes favourite, and l longshot

Subject to equation (18)

$$\frac{p_{if}^{v}}{p_{if}^{s}} > \frac{p_{il}^{v}}{p_{il}^{s}} \Rightarrow \frac{\left(p_{if}^{s}\right)^{b}}{\sum_{o=1}^{3}\left(p_{if}^{s}\right)^{b}} > \frac{\left(p_{il}^{s}\right)^{b}}{\sum_{o=1}^{3}\left(p_{il}^{s}\right)^{b}} \Rightarrow \left(p_{if}^{s}\right)^{b-1} > \left(p_{il}^{s}\right)^{b-1} \quad (20)$$

When $\left(p_{if}^{s}\right) > \left(p_{il}^{s}\right)$, (19) is only valid where $b > 1$. As a consequence, the odds of a given bookmaker underestimates favourites on average, only if b in (6) is significantly greater than 1 and higher values of b indicate higher degree of bias. Maximum likelihood is employed to estimate b and therefore, assess the degree of the bias in the odds offered by each bookmaker. To assess the accuracy of each bookmaker's predictions, we compare the values of McFadden's (1974) pseudo-R^2 statistic that each bookmaker's odds-implied probabilities achieve in the conditional logit model (a higher pseudo-R^2 implies a superior model fit).

4.5. Results

4.5.1. Investigating Arbitrage Opportunities

Table 4.4presents the observed correlation between the bookmakers' *closing* offers in our sample. As expected, the odds offered by diverse market operators are highly correlated, demonstrating that on average the bookmakers' offers are fairly aligned. However, the highest level of correlation is observed among book-balancing bookmakers. This indicates that, despite the fact that these bookmakers move their odds more frequently (adjusting for individual high stakes), such adjustments seem to happen in parallel across the set of book balancing bookmakers group. Among position-takers, *Bet365* seems to be the bookmaker more aligned with the book-balancers (possibly due to the fact that it is the only bookmaker in the group for which online betting is its sole focus), whereas *Ladbrokes* and especially *Stan James* are the operators showing the least correlation with the book-balancing market.

Table 4.4: This table reports the correlation in the final odds for each bookmaker, as well as the average correlation within and across the groups of bookmakers we identify as position takers (European bookmakers) and book balancers (Asian bookmakers).

		European Group				Asian Group		
	Bookmaker	Ladbrokes	William Hill	Bet 365	Stan James	SBO Bet	188Bet	Pinnacle
European Group	Ladbrokes	1.0000						
	William Hill	0.9927	1.0000					
	Bet 365	0.9907	0.9950	1.0000				
	Stan James	0.9928	0.9917	0.9891	1.0000			
Asian Group	SBOBet	0.9884	0.9925	0.9947	0.9849	1.0000		
	188Bet	0.9894	0.9942	0.9970	0.9867	0.9971	1.0000	
	Pinnacle	0.9890	0.9944	0.9971	0.9868	0.9971	0.9987	1.0000

Average Correlation Within and Across Groups

Group 1	Group 2	Average
European	Asian	0.9913
European	European	0.9920
Asian	Asian	0.9976

The optimization process described by equations (4) to (9) reveals the existence of 545 arbitrage opportunities across the 2,132 games, which indicates that 25.6% of the games in the sample offer apparent arbitrage opportunities. The distribution of these opportunities across the different leagues is shown in Table 4.5and the frequency with which each bookmaker's odds feature in the optimized portfolio are shown in Table 4.6. It has to be noted that the optimization can result in the odds of a diverse number of bookmakers featuring in each potential arbitrage opportunity (ranging from 2 to 6 in our sample), in order to achieve the maximum risk-free profit. In cases, where multiple bookmakers post equal odds for the same offer, we consider them all to have been responsible for the arbitrage opportunity. It is clear from Table 4.6 that some bookmakers are more likely than others to be involved in the generation of a theoretically risk-free portfolio[40]. These distinctions between bookmakers seem likely

[40] Removing Stan James from the sample causes the instances of potential arbitrage opportunities to drop to 287, which is indicative of the influence of a bookmaker which applies a policy of offering outlying odds, on the creation of arbitrage instances.

to be related each operator's policy. The chance of these distinctions occurring at random is in fact very small ($X^2(5) = 615.4$, $p = 0.000$).

Consistently with the behaviour suggested by the inter-bookmaker correlation statistics, *Bet365* and *William-Hill* appear to be more aligned in their pricing strategy with the book-balancing bookmakers as they are less frequently part of any risk-free array of bets. On the other hand, all book-balancing bookmakers are often part of an arbitrage portfolio, with *Stan James* or *Ladbrokes* often being on the opposite side (i.e. offering the opportunity to hedge the stake placed against the book-balancers). Amongst the book balancing bookmakers, *Pinnaclesports* is selected most frequently by the optimization programme to be part of an arbitrage portfolio. This may arise because this is the only book-balancing bookmaker which exhibits the same low over-round on 1x2 offers as on the Asian Handicap offers[41]. As a result, *Pinnaclesports* often offers the highest odds on a draw, which is frequently a useful bet in terms of equalizing payoffs across all outcomes[42].

[41] Book-balancers maintain an over-round of about 2% in the Asian Handicap markets, but their over-round is nearer 5%-6% for the 1x2 market. On the other hand, *Pinnaclesports*' over-round is about 2% in the 1x2 market.

[42] Mainly when a positive handicap (i.e. either +0.25 or +0.5) is not selected, the optimization indicates a stake should be placed on the draw so that there is no negative exposure on the draw outcome.

Table 4.5: Number of Matches with Arbitrage Opportunities by League. This table reports the number of matches with an arbitrage opportunity in our sample, broken down by league. If arbitrage opportunities occurred across leagues with the equivalent per-game frequency as the sample of matches in each league (380 in England, Spain, Italy, and France; 306 in Germany and Holland) we should expect to see the same number of matches with arbitrage opportunities in column 2 as the expected number of matches, shown in Column 5. The final column reports the ratio of (Observed – Expected)^2/Expected, and the $\chi^2(5)$ test p-value.

League	Number of Matches with Arbitrage	Frequency	Matches in Sample	Expected Matches with Arbitrage Opps.	(Observed – Expected)^2/ Expected
England	90	16.5%	380	97.139	0.525
Spain	109	20.0%	380	97.139	1.448
Italy	101	18.5%	380	97.139	0.153
Germany	94	17.2%	306	78.222	3.182
France	94	17.2%	380	97.139	0.101
Holland	57	10.5%	306	78.222	5.758
TOTAL	**545**	**100.00%**	**2,132**	**545**	**11.168**
				P-Value from $\chi^2(5)$Test	**0.048**

Table 4.6: Number of times a bookmaker's odds feature in a potential arbitrage portfolios selected by the optimisation programme for the 2,132 league matches played in the major Euroepan leagues in the season 2012-13. For each bookmaker we report the relative frequency with which their odds appeared in the X_{max}vector, and featured in the optimal potential arbitrage portfolio identified for a single game.

	Bookmaker	Number of times their odds featured in potential optimal arbitrage portfolios	Relative Freq.
Position-Takers	Ladbrokes	237	13.27%
	William Hill	59	3.30%
	Bet 365	65	3.64%
	Stan James	452	25.31%
Book-Balancers	SBOBet	248	13.89%
	188Bet	262	14.67%
	Pinnacle	463	25.92%
	TOTAL	1786	100%

If a bettor could have maintained access to all 7 bookmakers, without suffering restrictions in her betting size, the fully hedged strategy would have returned 7.56 times the initial bankroll across the season (assuming no reinvestment). This corresponds to an average risk-free profit of 1.38% per game. In reality of course, in order to properly measure the anticipated yield of such a strategy, one should account for complications relating to the occurrence of overlapping times of matches and the application of staking limits from the bookmakers for the various offers, which are likely to restrict profitability for sizeable bankrolls.

Taken together, the results discussed above show that sufficient price dispersion exists in the market for bettors to create a *seemingly* risk-free portfolio of bets that would guarantee to them profits for about 25% of football games played, if we assume that they could successfully implement this strategy. These results, therefore, serve to support the proposition at the heart of hypothesis 1 that there exist instances where the price dispersion in the betting market is adequate to generate *theoretically* risk-free opportunities for bettors to profit by simultaneously betting with different bookmakers on alternative outcomes related to the same event.

4.5.2. The Parties in the Arbitrage-Mirage

We determine for each apparent arbitrage opportunity the source of the odds which make up the arbitrage portfolio of bets (i.e. from book balancing or position taking bookmakers). In particular, we look at the instances where the odds offered by book balancing or position taking bookmakers on the favourite or the longshot feature in the optimal apparent arbitrage opportunities identified for the games in our sample of matches. These results are displayed in Table 4.7.It is evident from these results that on the vast majority of occasions when apparent arbitrage opportunities exist, one or more position-takers will be at one extreme (i.e. will offer the highest odds on the favourite/longshot) and one or more book-balancers will at the other extreme (i.e. will offer the highest odds on the longshot/favourite). In fact, this situation exists on 84% of occasions when apparent arbitrage opportunities exist and based on this frequency of arbitrage opportunities being created by bookmakers of different types (book-balancers and position-takers), the chance that this phenomenon is random is very low ($Z = 16.12$, $p=0.000$). This finding supports hypothesis 2, namely that most apparent arbitrage opportunities involve bets placed with different types of bookmakers (book-balancers and position-takers).

Given the considerable differences in how the two types of bookmakers operate, this finding is related to that of Franck et al. (2013), who found that the frequency of inter-market arbitrage opportunities was significantly higher (cf. frequency of intra-market arbitrage opportunities). Our results suggest that the intensive mobility of the book-balancers' odds occasionally drives them away from the position -takers' odds. In turn, the position-takers are likely to show a lagged response to such moves (see chapter 3). However, whether or not this occurs depends on their marketing related objectives. As a result, opportunities for *seemingly* risk-free profit arise by combining the odds offered by position -takers and book -balancers on the same event. On the other hand, intra-market competition is likely to lead to high price-coordination within each group (i.e. position- takers or book-balancers), restricting the likelihood of high price-dispersion between two bookmakers in the same group. Consequently, instances of arbitrage opportunities between bookmakers of the same group rarely arise.

The results displayed in table 4.7, also show that position takers are significantly more likely to offer above market odds for the favourite rather than for the longshot. On 58% of occasions, a position-taking bookmaker offered the highest odds for the favourite, as opposed to 40.8% for the longshot. Such a difference is unlikely to be random ($Z = 5.86$, $p = 0.000$). This tendency is more pronounced on stronger favourites (i.e. the more heavily favoured teams). Consequently, since position-taking bookmakers attract higher volumes on the favourites than on the longshots (Levitt, 2004), our finding is consistent with the view that position-taking bookmakers[43] are inclined to inflate odds for popular bets in order to attract customers. As a result, arbitrage opportunities most commonly emerge where a position-taker

[43] Such studies do not distinguish between diverse types of operators, but they assume a type of bookmaker consistent with our position-taker's definition.

offers the highest odds for the favourite and a book-balancer offers the highest odds in the market for the longshot[44]. These results are in line with hypothesis 2.

Table 4.7: This table reports the constituent bets in arbitrage portfolios by type of bookmaker. The first (second) column shows the type of bookmaker for which the optimal arbitrage portfolios contain bets on favourites (longshots). The favourite is identified as the team with the lower odds on the 1x2 betting market, bets on offers 1-6 are considered 'favourite' bets if the home team has lower odds; bets on offers 8-13 are considered 'favourite' bets if the away team has lower odds. All other bets (excluding draw bets) are considered 'longshot' bets. The number and proportion for which arbitrage portfolios are constructed, using bets from each type of bookmaker, are presented in column 3. For example, 42 opportunities (7.5% of the total) were identified for which both the favourite and longshot side of the portfolio were made with position-taking bookmakers. Columns 4, 6, 8, and 10 report the z-statistic and significance of a test (***, **, and * denoting 1%, 5%, and 10% significance, respectively) that the proportion of cases is equal across the four favourite/bookmaker-type combinations. Columns 5 and 7 report similar results to column 3, with the strength of the favourite increasing to $2.00 and $1.70 per dollar bet. Column 9 repeats the results from column 3 with the exclusion of the outlying position-taking bookmaker Stan James.

Best Offer Favourite by:	**Best Offer Longshot by:**	**N (all-sample)**	**z-stat**	**N (fav. < 2.00)**	**z-stat**	**N (fav.< 1.70)**	**z-stat**	**N (Exc. StJames)**	**z-stat**
		prop	***(Sig.)***	***prop***	***(Sig.)***	***prop***	***(Sig.)***	***prop***	***(Sig.)***
Position Taker	**Position Taker**	41	*-15.46*	27	*-13.6*	6	*-6.14*	18	*-13.6*
		0.075	(***)	*0.071*	(***)	*0.073*	(***)	*0.061*	(***)
Position Taker	**Market Maker**	277	*12.06*	183	*9.00*	57	*8.76*	152	*9.07*
		0.508	(***)	*0.480*	(***)	*0.695*	(***)	*0.514*	(***)
Market Maker	**Position Taker**	181	*4.07*	142	*4.95*	14	*-1.91*	78	*0.53*
		0.332	(***)	*0.373*	(***)	*0.171*	(**)	*0.264*	
Market Maker	**Market Maker**	46	*-13.91*	29	*-12.80*	5	*-7.15*	48	*-4.09*
		0.084	(***)	*0.076*	(***)	*0.061*	(***)	*0.162*	(***)
Total		545		381		82		296	
		1		*1*		*1*		*1*	

[44] In general, position-takers, do not offer higher odds for favourites on average compared to book-balancers due to their higher over-round). However, their odds on favourites are closer to those offered by book-balancers than the odds they offer on longshots. This finding suggests that they probably do not distribute their over-round proportionally (as one would expect based on chapter 2).

5.4.3. "Winners" and "Losers" from Apparent Arbitrage Opportunities

In order identify the type of bookmakers which would lose against potential arbitrageurs, should the identified risk-free opportunities be exploitable, we employ the simulation described by equations (12) to (14). Placing $1 on all the 813 outcomes for which the odds posted by position taking bookmakers form part of the optimal fully-hedged portfolio, yields an average profit of $0.16 per bet. Adopting the same strategy of backing all outcomes where the book-balancers' odds feature in the optimal fully-hedged portfolio results in a loss of $0.024 per bet[45]. The profit obtained on the bets placed at the position-takers odds in these cases is significantly higher than the profit (in fact a loss) obtained on the bets placed at the position takers odds(t-stat = 3.37, p = 0.000). The high return against the position-takers though is exaggerated by the fact that a number of high paying longshots won in this particular sample[46]. Thus, adjusting the strategy as described by equation (15), leads to an average profit per bet of $0.04 per $1 stake against position taking bookmakers and an overage loss of -$0.047 per $1 against book-balancers. These returns remain significantly different (t-stat = 2.60, p = 0.005). This result supports hypothesis 3a, namely that Position-takers suffer losses on average when an apparent arbitrage opportunity exists.

Interestingly, the loss incurred by the book-balancing bookmakers on these bets is close to their over-round[47].This suggests that the bets accepted by such bookmakers in the event that these form part of an arbitrage portfolio equal the

---

[45]It has to be clarified that in this case $1 is bet on each offer that falls part of the portfolio, no matter what the fraction of capital allocated from the optimization ((4) to (9)). Therefore, the results of this simulation are not comparable to the results of the fully hedged strategy. By way of example, the fully hedged strategy may assign 90% of the capital to bet A and 10% to bet B and according this, we would bet $0.9 and $0.1 respectively. However, in the unit-stake simulation $1 is staked on bet A and $1 on bet B, since the objective is to identify how the profit is distributed across the two types of bookmakers, rather than to create a hedged position.

[46] 55 out of 813 $1 bets that were placed at odds > 5 had an extremely high profit of $1.3. As a result, this small number of *lucky* bets account for $71.6 out of the $130 won in total by this strategy. Hence it is important to ensure that they do not bias the conclusions. This is achieved by applying the weighting implied by equation (15).

[47] The over-round of such bookmakers is about 2% for Asian Handicap offers and 5%-6% for 1x2 offers, excluding *Pinnaclesports*, whose over-round is also about 2% in the 1x2 market.

expected loss for the bettor should they place a random bet with these bookmakers. In other words, the fact that another bookmaker offers sufficiently different odds to generate (at least in theory) an arbitrage opportunity, does not change the expected value of the bets that they receive. Hence, this result effectively justifies *Pinnaclesports'* statement on their website that the motive for a bet (e.g. intention to arbitrage) should be irrelevant to a book-balancer[48]. Therefore, from the bettor's perspective it seems that a higher return is expected by taking positions against outlying odds of position-takers, rather than by hedging such positions against book-balancers, since in the latter case the average profit drops to $0.013 per $1 bet. Consequently, there is evidence from these simulations to support H3a, namely that the expected loss from an apparent arbitrage opportunity is likely to be suffered by the position-taking bookmakers. This is consistent with the findings of Franck et al. (2013).

The Efficiency of Book Balancers and Position Takers Odds

Table 4.8 presents the results of estimating separate conditional logistic regression models (as described in equations (17) and (18)) based on the odds offered on football matches across the 6 leading European leagues by position taking and book balancing bookmakers, respectively,[49] for period, 2009/10 to 2011/12. As expected, the forecasting accuracy of odds offered by book-balancing bookmakers is higher on average compared to that of position takers' odds. This result is in line hypothesis 3b and is consistent with the evidence provided by Franck et al. (2013) and Smith, Paton and Vaughan Williams (2006, 2009) that demand driven (cf. bookmaker) markets are more efficient predictors of event outcomes. Furthermore, the favourite-

---

[48] *Pinnaclesports* statement is: "all bookmakers shouldn't care about the motivation for placing a bet, but should simply look to balance the bet volume".Source: http://www.pinnaclesports.com/betting-promotions/arbitrage-friendly

[49] We did not have data for Stan James covering the same period. However, we reran the regressions over a one-year sample where we could acquire data for all bookmakers and found that the McFadden-R^2 achieved by the probabilities implied by Stan James' odds was the lowest of all bookmakers in the same sample of games. This is likely to be related to the fact that this bookmakers' odds often lead to theoretical arbitrage.

longshot bias is more pronounced for position-takers than for book-balancers, confirming hypothesis 3c. However, this is not the case for the odds offered by the book balancer *Sbobet,* whose over-round for the 1x2 market is the highest amongst book-balancing bookmakers. On the other hand, the conditional logistic regression based on *Pinnaclesports* odds (whose over-round in the 1x2 market is as low as it is for Asian Handicaps) has a coefficient nearly equal to 1, suggesting no favourite-longshot bias. This variation of the bias is in line with the Vaughan William's (1998) proposition that the level of transaction costs affect the degree of favourite longshot bias. In this case is likely to indicate that book-balancing bookmakers (excluding *Pinnaclesports*) are trying to direct the demand for the longshots to their main Asian Handicap markets, in order to facilitate the balancing of their book in that market. By way of example, regarding the game between Bayer Leverkusen and Wolfsburg, played in the German Bundesliga on 14/09/2013 at 14:30 BST: On the day of the game, at 8am BST the odds on the main handicap markets were almost identical, (Bayer Lev. -0.5, Wolf. +0.5)[50] are (1.77, 2.20)[51] for both *Pinnaclespots* and *Sbobet* and (Bayer Lev. -0.75, Wolf. +0.75) are (2.00, 1.94) for *Pinnaclespots* and (1.99, 1.95) for *Sbobet*. However, on the 1x2 market the offers are (1.77, 4.01, 4.86) for *Pinnaclespots* and (1.77, 3.7, 4.5) for *Sbobet*. Given that both bookmakers accept significantly higher stakes in the Asian Handicap market compared to the 1x2, it is fair to assume that their evaluation of the game's outcomes is very similar. Hence, by discounting the non-favourite odds on the 1x2 market, *Sbobet* seems to be inviting its clients who are interested in betting on Wolfsburg to bet on the +0.5 or +0.75 goal handicap. This is likely to assist *Sbobet* in generating a balanced book. However, if one focuses on the odds related to the 1x2 market alone (as we employed in the conditional logit models) these would indicate that *Sbobet's* odds reflect a higher probability for the longshot (cf. that suggested by the Asian Handicap odds). Consequently, this may lead to inaccurate conclusions being drawn regarding the existence of favourite-longshot bias. In other words, the fallacy is likely to have been

[50] This notation indicates that Bayer is handicapped by 0.5 goals and Wolfsburg assigned a 0.5 goal premium

[51]$1.77 ($0) is returned to $1 stake on Bayer, if they win (don't win) and $2.20 ($0) is returned to Wolfsburg if they win/draw (lose).

created by the operator's policy rather than a pricing inefficiency. To prove this more formally, in a future study, probabilities need to be extracted from Asian Handicap offers[52] and be analysed over a large number of games. These could be used to show that any bias indicated by odds in the 1x2 market is removed when employing odds for the same events sourcing from the Asian Handicap market. Such a complication is not apparent for odds offered by *Pinnaclesports*, as this bookmaker's over-round is consistently low across different markets, leaving little space for skewing policies. Consequently, even though we provide some evidence in support of H3c, some additional analysis based on Asian Handicap offers is required in a future study, in order test the robustness of this evidence.

Table 4.8: This table reports the results of conditional logit modelling (using Equations (17) and (18)) based on the odds offered by six bookmakers on all 6,396 matches in the English, Spanish, Italian, German, French and Dutch leagues over the three seasons 2009-10, 2010-11 and 2011-12 for all match outcomes (Home win, Draw, Away win (19,188 total observations of odds per bookmaker). Bookmakers are classified as either position-takers or book-balancers. The third and fourth columns of the table report the estimated coefficient of the conditional logit model and its standard error, respectively. The fifth column reports the P-value of a z-test to determine whether the true value of the coefficient in column three is equal to 1. The sixth column reports the result of test whether the coefficient is significantly greater than 1 at the 10%, 5%, and 1% levels with the signs (*), (**), and (***), respectively. The final column reports the McFadden Pseudo-R^2 of the conditional logit model.

Group	Bookmaker	Coefficient	Std. Error	Prob. (Coeff. = 1)	Sig.	Pseudo-R^2
Position takers	Ladbrokes	1.0743	0.0309	0.0081	(***)	0.1085
	William Hill	1.0605	0.0304	0.0232	(**)	0.1101
	Bet 365	1.0560	0.0302	0.0318	(**)	0.1106
Book Balancers	SBOBet	1.0784	0.0311	0.0059	(***)	0.1106
	188bet.com	1.0413	0.0299	0.0831	(*)	0.1113
	PinnacleSports	1.0081	0.0289	0.3900		0.1114

[52] This can be done assuming a distribution (such as bivariate Poisson) for the exact number of goals to be scored by the two teams in the game, in order to extract the teams' parameters and then apply them on the distribution to obtain the probability distribution of each exact score. Then, one can sum the probabilities of the exact scores that lead to each outcome, in order to end up with 1x2 probabilities. Incorporating such an analysis in this paper would significantly increase its length without adding to the main thrust of the arguments presented.

4.6. Discussion

In this paper, we analysed a unique data set of 1x2 and Asian Handicap odds for football games played in major European leagues, offered simultaneously by the main bookmakers. All these odds were taken from periods close to the games' kick-offs when the trading volumes are at a maximum. Employing a linear programming methodology, we identified the best combination for each of 545 games where a fully hedged profitable investment appears to be possible. Notwithstanding challenges related to its implementation, such a strategy could, in theory, guarantee a profit of 1.3% per game on average. To some extent, our findings confirm those of Franck et al. (2013). However, we focused our analysis on a sample of *concurrently available* offers, in highly liquid markets, in order to ensure that the odds comprising a theoretically fully-hedged profitable portfolio coexisted. The periods close to the games in which the odds occurred ensured that significant amounts could be placed on such odds, assuming that someone had access to the corresponding operator at the time. We also included Asian Handicap offers in our analysis, which significantly increase the possibility of arbitrage due to the low over-round of such offers in the book-balancers' market. In line with Franck et al (2013), we find that the arbitrage possibilities mainly arise *inter* rather than *intra* market. However, in our case we consider *inter* market to be between book-balancing and position-taking bookmakers, due to the fundamental structural differences in their operations. As a result, we add to the evidence provided by Marshall (2009), according to which (at least theoretically) arbitrage opportunities between bookmakers are not infrequent, We suggest that this finding, which appears to contradict that of other studies (e.g., Dixon and Pope, 2004; Deschamps and Gergaud, 2007; Luckner and Weinhardt, 2008; Deschamps, 2008; Vlastakis, Dotsis and Markellos, 2009; Franck, Verbeek and Nüesch, 2010) arises from the choice of bookmakers under investigation. Thus, studies that consider position-taking or market-making bookmakers only are expected to identify a significantly lower frequency of potential arbitrage instances compared to studies that consider both types of bookmakers together.

Investigating the bookmaker-specific attributes of the parties involved in the apparent arbitrage opportunities leads to the conclusion that such opportunities are likely to be created by position-takers' inefficient pricing. This pricing policy maybe intentional, in order to publish odds that may attract customers, or the result of their prices lagging behind book-balancers (due to the pace at which book-balancers' odds are informatively updated, driven by the flow of "smart money"[53]). As a consequence, given the public availability of odds, it seems fair to assume that if the sole objective of the price setting strategies of position-taking bookmakers' were the efficient calibration of event outcomes they would fully align their odds with those of book-balancers. Hence, the fact they do not do so, probably indicates that marketing related values are also considered. Such considerations may include offering higher odds for popular bets compared to their competitors or keeping their odds stable in order to avoid upsetting casual bettors by continuous changes[54]. As a result, the combination of such bookmakers not balancing their books and being aware of their setting of inefficient prices, leads to the conclusion that they are likely to restrict the activities of bettors who systematically place stakes on offers that are likely to exhibit negative expected value for the bookmaker. In addition, significant anecdotal evidence exists that such bookmakers operate discriminating behaviour, against long-term winning customers, (Veitch, 2009; Franck et al., 2013). Consequently, we would argue that the apparent arbitrage opportunities observed in fixed odds betting markets are very likely to be a mirage.

The implications for bookmakers' setting objectives that are not compatible with their pricing in a manner which reflects the efficient estimation of outcomes' probabilities are very broad. For example, efficient pricing in betting markets, as in wider financial markets, implies prices that accurately reflect the true, underlying value of assets. However, it seems that in the most popular forms of betting

---

[53] *Pinnaclesports* state on their website "This limiting of arbitrage players is a reflection of a bookmaker's short-comings, such as posting 'bad odds', or an inability to move odds fast enough to avoid being the focus of arbitrage players".

[54] Frequent odds moves often lead to bets not going through because the odds have changed, which might be annoying to the casual punter. Also, a casual player may want to spend some time comparing odds across different events and diverse markets, which can be a difficult task if odds move continuously.

bookmakers may consider alternative objectives when setting its odds, which may lead to (even intentional) mispricing. However, as discussed above, such mispricing is unlikely to be systematically exploitable, due to restrictions imposed on potentially skilled bettors. In that sense, the original concept of the EMH, according to which market prices are expected to converge to the fundamentals' true values, subject to the activity of informed traders, may be radically distorted. Consequently, the main conclusion which emerges from this study is that questions concerning the efficiency of markets and inferences concerning behavioural characteristics of the trading population can only be fully answered if the price-setters' objectives are taken into consideration.

References

Adams, B. R., Rusco, F. W., and Walls, W. D. (2002).Professional bettors, odds-arbitrage competition, and betting market equilibrium.*The Singapore Economic Review*, 47, 111-127.

Avery, C., and Chevalier, J. (1999).Identifying investor sentiment from price paths: the case of football betting.*Journal of Business*, 72, 493-521.

Barberis, N., Shleifer, A., and Vishny, R. (1998).A model of investor sentiment.*Journal of Financial Economics*, 49, 307-343.

Dantzig, G. B. (1951). Maximization of a Linear Function of Variables Subject to Linear Inequalities.In T. C. Koopman (Ed.) *Activity Analysis of Production and Allocation* (pp. 339-347). New York: John Wiley and Sons.

Deschamps, B., and Gergaud, O. (2007). Efficiency in Betting Markets: Evidence from English Football. *Journal of Prediction Markets*, 1, 61-73.

Deschamps, B. (2008). Betting Markets Efficiency: Evidence From European Football. The *Journal of Gambling Business and Economics*, 13(2), 66-76.

DeBondt, W., and Thaler, R. (1985). Does the stock market overreact? *Journal of Finance*, 40, 793-805.

Dixon, M. J., and P. F. Pope, P. F. (2004). The Value of Statistical Forecasts in the U. K. Association Football Betting Market. *International Journal of Forecasting*, 20, 697- 711.

Edelman, D., & O'Brian, N. (2004). Tote arbitrage and lock opportunities in racetrack betting. *The European Journal of Finance*, 10, 370–378.

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

Fama, E. (1970). Efficient Capital Markets: A Review Of Theory And Empirical Work. *Journal of Finance*, 25, 383–417.

Fama, E. (1998). Market efficiency, long-term returns, and behavioral finance. *Journal of Financial Economics*, 49(3), 283-306.

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

Forrest, D. (2012). The Threat to Football from Betting-Related Corruption. *International Journal of Sports Finance*, 7, 99-116.

Franck, E. P., Verbeek, E., and Nüesch, S. (2010). Prediction Accuracy of Different Market Structures - Bookmakers Versus a Betting Exchange. *International Journal of Forecasting*, 26(3), 448-459.

Franck, E., Verbeek, E., and S. Nüesch, E. (2013). Inter-Market Arbitrage In Betting. *Economica*, 80, 300-325.

Gramm, M., and Owens, D. H.(2005). Determinants of Betting Market Efficiency. *Applied Economics Letters*, 12(3), 181-185.

Hausch, D. B., and Ziemba, W.T. (1990). Arbitrage Strategies for Cross Track Betting on Major Horseraces. *Journal of Business*, 63(1), 61-78.

Hodges, S., and Lin, H.(2009). Fixed Odds Bookmaking with Stochastic Betting Demand. *Working Paper, Cass Business School, City College London*.

Humphreys, B. R. (2010). Prices, Point Spreads and Profits: Evidence from the National Football League. *Working Paper, University of Alberta*.

Ikenberry, D., Lakonishok, J., and Vermaelen, T. (1995). Market underreaction to open market share repurchases. *Journal of Financial Economics*, 39, 181-208.

Jensen, M. C. (1978). Some anomalous evidence regarding market efficiency. *Journal of Financial Economics*, 6, 95-101.

Kelly, J. L., Jr. (1956). A New Interpretation of Information Rate. *Bell System Technical Journal*, 35, 917-926.

Kuypers, T. (2000). Information and efficiency: an empirical study of a fixed odds market. *Applied Economics*, 32(11), 1353-1363.

Lakonishok, J., Shleifer, A., and Vishny, R. (1994). Contrarian investment, extrapolation, and risk. *Journal of Finance*, 49, 1541-1578.

Lane, D., and Ziemba, W. (2004). Jai Alai arbitrage strategies. *The European Journal of Finance*, 10, 353-369.

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

Luckner, S., and Weinhardt, C. (2008).Arbitrage Opportunities and Market Making Traders in Prediction Markets. *IEEE Joint Conference on E-Commerce Technology (CEC '08) and Enterprise Computing, E-Commerce and E-Services (EEE '08).*USA: Washington D.C.

Madhavan, A. (2000). Market Microstructure: A Survey. *Journal of Financial Markets*, 3, 205-258.

Magee, S. (1990).*The Channel Four Book of Racing*. London: Sidgwick and Johnson.

Marshall, B. R. (2009). How Quickly is Temporary Market Inefficiency Removed? *Quarterly Review of Economics and Finance*, 49, 917-930.

McFadden, D. (1974). Conditional logit analysis of qualitative choice behavior. In P. Zarembka (Ed.), *Frontiers in econometrics* (pp. 105-142). New York: Academic Press.

Oikonomidis, A., and Johnson, J. E. V. (2011). Who Can Beat the Odds? The Case of Football Betting Reviewed. In L. Vaughan Williams (Ed.) *Prediction Markets: Theory and Applications* (pp. 204-220). Abingdon, U.K.: Routledge.

Paton, D., and Vaughan Williams, L. (2005). Forecasting outcomes in spread betting markets: Can bettors use 'Quarbs' to beat the book? *Journal of Forecasting*, 24, 139-154.

Paul, R. J., and. Weinbach A. P. (2007). Does Sportsbook.com Set Pointspreads to Maximize Profit? Tests of the Levitt Model of Sportsbook Behavior, *Journal of Prediction Markets*, 1, 209-218.

Paul, R. J., and Weinbach, A. P. (2008). Price Setting in the NBA Gambling Market: Tests of the Levitt Model of Sportsbook Behavior. *International Journal of Sports Finance*, 3, 2-18.

Pope, P. F., and Peel, D. (1989). Information, Prices, and Efficiency in a Fixed-Odds Betting Market. *Economica*, 56, 323-341.

Ross, S. (1976). The arbitrage theory of capital asset pricing, *Journal of Economic Theory*, 13, 341-360.

Sauer, R. D. (1998). The Economics of Wagering Markets. *Journal of Economic Literature*, 36 (4), 2021-2064.

Sharpe, W. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *Journal of Finance*, 19, 425-442.

Sharpe, W. F., and Alexander G. (1990).*Investments* (4th ed.). Englewood Cliffs, N. J.: Prentice-Hall.

Shleifer, A., and Vishny, R. W. (1997).*The Limits of Arbitrage*. *Journal of Finance*, 52(1), 35-55.

Smith, M. A., Paton, D., and Vaugham Williams, L. (2006).Market efficiency in person-to person betting. *Economica*, 73(292), 673-689.

Smith, M. A, Paton, D., and Vaughan Williams, L. (2009). Do bookmakers possess superior skills to bettors? *Journal of Economic Behavior and Organization*, 71 (2), 539-549.

Vaughan Williams, L. (1999). Information efficiency in betting markets: a survey. *Bulletin of Economic Research*, 51, 1-39.

Vaughan Williams (Ed.) (2005). *Information Efficiency In Financial Betting Markets*. New York: Cambridge University Press.

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

Vaughan Williams, L., and Paton, D. (1998). Why are some favourite-longshot biases positive and some negative? *Applied Economics*, 30, 1505-1510.

Veitch, P. (2009). *Enemy Number One: The Secrets of UK's Most Feared Professional Punter* (pp. 226-234). UK: CPI William Clowes Beccles NR34 7TL.

Vlastakis, N., G., Dotsis, G., and Markellos, R. N. (2009). How Efficient is the European Football Betting Market: Evidence from Arbitrage and Trading Strategies. *Journal of Forecasting*, 28, 426-444.

Willis, K.E. (1964). Optimum no-risk strategy for win-place pari-mutuel betting. *Management Science*, 10, 574-577.

Woodland, B. M., and Woodland, L. M. (1991). The Effects of Risk-Aversion on Wagering: Point Spread Versus Odds. *Journal of Political Economy*, 99, 638-653.

Conclusion

The overall objective of this thesis is to identify the impact of institutional characteristics on the setting of prices in speculative markets and to associate these with apparent pricing anomalies, in order to assess whether such anomalies constitute evidence against the Efficient Market Hypothesis (EMH). The structural characteristics of operators in bookmaker-based betting markets are examined to identify clear differences with operators in financial exchange-based markets that have been overlooked in previous studies. It is suggested that the EMH in such a market is effectively a joint hypothesis on the degree that market prices significantly deviate from the true, underlying value of assets and whether potential biases can be exploited in the long-term to systematically lead investors to profit. In that context, a range of unique football betting data sets are analysed in order to discover whether consistent and persistent evidence of mispricing exists. In addition, the sources of these anomalies are explored, in order to assess whether the apparent price biases are truly exploitable.

The first section of the thesis reviews the literature that focuses on the investigation of the EMH in football betting markets. Overall, the literature suggests that market odds are relatively accurate predictors of football outcomes, but several studies provide evidence concerning biases in prices to the extent that they do not, apparently, fully incorporate price-related signals, fundamental information regarding the abilities of football teams and incidental information related to football games. However, it is also shown that market odds tend to more efficiently approximate event outcomes through time. As a consequence, this evolving increase in market efficiency is likely to challenge the long-term success of strategies attempting to capitalize on market biases, as they run the risk of a rapid market correction in market prices, before significant returns can be generated. This in turn, reduces the incentive for informed traders to invest the time and resources necessary to develop models to exploit any existing anomalies. Consequently, it is particularly important, to provide evidence regarding the persistence and consistency of the apparent pricing anomalies. In addition, by associating these with the market structure it is possible to examine why they might be expected to last in the long run. Moreover, it is pointed out that the literature provides us with little information regarding the potential exploitability of theoretically profitable betting strategies. We argue that if bookmakers are likely to pose restrictions on potentially successful traders, this is likely to form a *cause and effect relationship.* Specifically, if bookmakers are conscious that they can apply such restrictive policies then this may significantly reduce their motivation to remove any observed (non-exploitable) anomalies. As a result, mispricing in this case should not be considered as inefficiency, as it would probably disappear if the bookmakers had no intention of restricting those who are likely to exploit it. Therefore, the structural characteristics of market operators needs to be investigated, in order to assess whether their apparently inefficient pricing is likely to be exploited by investors to generate significant long-term returns.

The second section of the thesis focuses on the examination of the persistence and consistency of the favourite-longshot bias (FLB) in the football betting market, according to which the expected long-term returns of bets on shorter odds are significantly higher compared to those on longer odds. The existence of this bias is evident in most betting (mainly bookmaker) markets and this section extends the investigation of this phenomenon to football betting markets. In particular, data across an extensive period, across different European leagues is employed to assess the extent of the FLB. It is found that the FLB has been persistent through time, being apparent in a large sample of football games. Consequently, FLB is considered to be a structural effect of the traditional bookmaker [55] market. Competition among bookmakers is identified as a possible cause for the creation of the FLB, as bookmakers are likely inclined to offer attractive odds on popular bets in order to attract and retain customers. The literature suggests that betting on favourites is more popular amongst bettors (cf. to betting on longshots) and therefore, bookmakers are expected to attempt to induce current and potential clients to bet with them in preference to betting with their competitors by offering high odds for the bets that they

[55] We refer to bookmakers characterized as position-takers in Chapter 3. Most of the major book-balancing bookmakers did not exist during most of the 1999-2008 period analysed in this study.

prefer. Even though this policy is likely to decrease the bookmaker's average profit per game, that cost is probably outweighed by the growth of a customer base, consisting of clients who in the long-run lose money[56]. This effect is likely to account for the observed heterogeneity in the magnitude of the bias across leagues, as bookmakers have to pay a higher cost for attracting betting on less competitive leagues, which are dominated by a group of strong teams (cf. more equal strength teams in more competitive leagues).

The FLB is a bias "controlled" by the bookmaker and therefore, it is not in itself sufficient for the generation of exploitable, profitable strategies. It could be argued that the existence of a non-exploitable bias is not of interest for the study of the EMH. However, this thesis does not primarily aim to test the EMH, but to set the appropriate framework with regards to how theoretical mispricing should be interpreted. In that sense the investigation of the FLB is useful due to the insights that it offers concerning the bookmakers' structural characteristics. Hence, the persistence of the FLB and the form of its expression indicates that traditional bookmakers are likely to consider marketing objectives in their price setting, in order to satisfy the excessive demand for favourites. Consequently, their relatively high over-round[57]is not symmetrically distributed across the outcomes of a game rather it offers a premium to bets on favourites. This premium should not exceed, on average, the bookmaker's expected profit in laying bets on favourites. Consequently, estimates arising from bookmakers with lower over-rounds are expected to be less biased[58]. Moreover, given that event probabilities are unknown, this asymmetric distribution of bookmakers' over-round increases the probability that several bets on favourites are offered at an expected loss (profit) for the bookmaker (bettor). However, even if this is occasionally

---

[56]As shown on Chapter 2, such bookmakers are likely to severely limit or prevent bets from skilled bettors, further reducing the cost of their attractive offers.

[57] As shown in chapter 3 the position-takers' (to whom Chapter 2 refers) over-round is significantly higher compared to that of book-balancers.

[58] This is the case, as shown in chapter 4 and as shown in the literature for estimates based on betting-exchange odds.

the case, leading authors suggest[59] that it should not be possible for bettors to benefit from it, because, they argue, bookmakers are superior forecasters (cf. bettors) of event outcomes. If one agrees with this view then it cannot be possible for bettors to systematically identify these opportunities. The validity of this view is examined in Chapter 3, and the results of analysing such theoretically profitable opportunities are presented in Chapter 4, alongside an analysis of the bookmakers' structural idiosyncrasies.

Chapter 2, in the tradition of the existing literature, focused on the investigation of traditional bookmakers. As discussed in Chapter 3, such bookmakers attempt to maximize a customer base of casual clients and set their odds in a way that leaves them exposed to the outcome of single games, but potentially maximizes their long-term returns against their unsophisticated clients. Such bookmakers operate with high over-rounds and are likely to restrict any potentially successful bettor from trading. However, as the research presented in Chapter 3 uncovers, not all bookmakers operate in this manner. The main players in the modern football betting industry act as market makers in financial markets, matching buyers with sellers, intending to secure profit equal to a low fraction of the trading volume. Thus, their profitability is linked with the volume of the stakes traded in their books. As a result, such bookmakers are not averse to accepting bets from skilled bettors because their business model (in contrast to the business model of position takers) is that high volumes lead them to high profits, irrespective of the source of the bets. Consequently, such bookmakers set a market similar to that of betting-exchanges, in which they guarantee liquidity. This, therefore, results into the generation of significantly higher trading activity. Even though such a business model is referred to in the literature as a potential method for operating a sports-book, the research conducted in Chapter 3, is the first to directly compare the behaviour of the two types of bookmakers (referred to as book-balancers and position-takers) and to reveal their fundamental distinctions. To achieve this , a unique data set is employed which matches odds data offered at the same time by two major bookmakers, representative of the two bookmaker types, for a full season of

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

football games played across the main European football leagues. Unlike most studies that view the odds as a static quantity, in this study the odds offered simultaneously by the two bookmakers at different points in time for each game, are analysed in order to explore the interactions between the bookmakers' odds through time. The findings show that book-balancers frequently adjust their odds, to reflect a volume-weighted average of the betting public's perception (potentially shaped by "smart money"). Position-takers, on the other hand, rarely move their prices, possibly partially relying on their higher over-round, which should prevent them from frequently accepting bets at expected loss. However, it is shown that when position-takers adjust their odds, these tend to converge to those of book-balancers, effectively showing an indirect, lagged response to information arising from the betting public. On the other hand, position-takers' odds do not seem to influence those of book-balancers. Finally, the analysis demonstrates that the efficiency of forecasts of game outcomes based on odds increases as the kick-off approaches, demonstrating that the transmission of information from the betting public to bookmakers increases the efficiency of market prices. This is consistent with the EMH and rejects the view that bookmakers are superior forecasters compared to the betting public.

As discussed on Chapter 3, book-balancing bookmakers are key players in the betting market, in terms of the volumes they trade. In addition, the results demonstrate that their odds, driven by the trades of skilled bettors, converge to the underlying probabilities of event outcomes more rapidly and accurately compared to those of position takers. Position-taking bookmakers seem to acknowledge this fact by following trends in book-balancers' odds. This finding arguably supports anecdotal evidence that position taking bookmakers restrict access to informed traders. Since the betting public consists of informed and noise traders, the information that leads to improvements in the market estimates should come from the former group. Hence, the facts that book-balancers' odds become more efficient by their volume-driven adjustments and position-takers tend to follow such adjustments (and not vice-versa) might imply that the latter mainly receive bets from noise traders (i.e. they have little *direct* access to information that could improve their estimates). Hence, they occasionally follow the book-balancers' adjustments, in order to indirectly benefit from information arising from "smart money" accepted into the books of the book-balancers. Therefore, it could be claimed that, unlike position-taking bookmakers, which have exclusively been examined by previous studies, book-balancers constitute a highly liquid[60], barrier-free environment closer to that of exchange-based financial markets. Consequently, it could be argued that the book-balancers markets share the characteristics that shaped the EMH and therefore, they arguably offer a more appropriate setting for the purpose of testing the EMH.

The latter argument is reinforced in Chapter 4. In this chapter, data covering a season of football games played in the major European leagues, from several leading bookmakers of both types, are analysed. Most studies in the football betting literature tend to disregard the possibility of odds moving, which is not a valid assumption (as shown in Chapter 3). Consequently, several studies that examine odds from different sources do not ensure that these offers have actually coexisted. In that sense, the study presented in Chapter 4, is innovative in that data are simultaneously collected from different bookmaker sources, using multi-threading programming. The odds for each game in the sample are collected at times close to its kick-off, when the staking limits and consequently the trading volumes are at a maximum. This ensures that a significant level of information has already been transmitted from the betting public to the bookmakers (as demonstrated in Chapter 3). Given the relative stability of position-takers' prices and the (ever-increasing, as games approach kick-off time) mobility in the odds of book-balancers, combined with the marketing objectives of the position-takers, significant price dispersion is anticipated in the market's closing odds. Applying linear optimization, it is shown that in more than 25% of the games in the sample, the formulation of fully hedged portfolio is possible. In the vast majority of cases, the arbitrage is created between odds offered by position-taking and book-balancing bookmakers, rather than between two bookmakers of the same type. This is to be expected, based on the results presented in Chapter 2, namely, that position-takers usually offer the highest odds for the game's favourite. A simulation reveals

[60] This is their main advantage over betting-exchanges, in which as shown in Chapter 4, the liquidity is highly inconsistent across different games and significantly lower on average overall.

that betting on the extreme[61] odds offered by position-takers is profitable, whereas betting on the extreme odds offered by the book-balancers generates an average loss which is equivalent to that of a random bet placed against such bookmakers. In other words, in cases of significant inter-bookmaker disagreement, position-takers on average offer to take on bets at expected loss. This is reinforced by comparing the accuracy of the two types of bookmakers' odds-implied forecasts over a longer period, which shows that predictions based on the odds of book-balancers are significantly more accurate. Hence, it seems fair to argue that these theoretical arbitrage opportunities are created by position-takers either lagging behind rapid odds adjustments (based on "fresh" information signalled by smart money) or by them sticking to attractive offers for the purpose of acquiring and maintaining customers.

It could be claimed that the findings of Chapter 4, extend those presented on Chapter 2, showing that the asymmetric distribution of the over-round, applied by position-taking bookmakers (potentially for marketing purposes), often cross the line of efficiency, offering profitable betting opportunities to the betting public. In that sense, it could be argued that they attempt to create a balance between occasionally offering attractive odds and not laying bets at expected loss, but often fail to satisfy the latter objective. Assuming that these bookmakers imposed no restrictions on informed traders, they would be expected to receive very high volumes on their outlying offers. Since position-takers are shown to exhibit significant expected loss on such positions, allowing access to informed traders would result in major losses over a large sample the games (i.e. the analysis shows arbitrage opportunities in more than 25% of games). This is likely to be too high a cost to be reimbursed from the acquisition of casual bettors (who trade in mainly small amounts). Thus, position-takers are more likely to optimize their long-run profit by restricting access to skilled bettors. This is consistent with anecdotal evidence and also with the findings discussed in Chapter 3.

Consequently, despite evidence that theoretically, positive returns can be generated by trading against position-takers' published odds, the EMH cannot be rejected, as the successful implementation of such a strategy seems unlikely given the

[61] i.e. the offers that form the fully-hedged portfolio

structural characteristics of such bookmakers. On the other hand, book-balancers, who do not seem to restrict access to potentially successful traders, do not exhibit similar anomalies in their pricing.

The findings of this study as a whole suggest that the structural characteristics of price-setters influence the degree to which prices converge to the underlying value of assets. In particular, market operators that take risks with regards to event outcomes, tend to operate with larger transaction costs and appear to set barriers to trade that result into theoretical pricing anomalies. We argue that such anomalies do not constitute strong evidence against the EMH, as they are unlikely to be exploitable in the long run. On the other hand, operators which act as market makers (book-balancers) do not impose such barriers to trade and guarantee high liquidity and low transaction costs in the market. These bookmakers create a dynamic environment of continuously moving prices that rapidly and efficiently respond to information relevant to event outcomes, driven by the trades of sophisticated investors. As a result, the juxtaposition of the two business models (book-balancing and position-taking) shows that when trading restrictions are removed, the market forces shift prices towards efficiency. Furthermore, more broadly, the efficient incorporation of information in the prices of a demand-driven market could be regarded as encouraging for the prospects of prediction markets, as it shows that, subject to the provision of significant liquidity, estimates reflecting volume-weighted averages of the public's opinion are likely to be more accurate than those supplied by expert-forecasters. Finally, due to the similarities of the bookmaker betting market and Over-The-Counter financial markets, it would be interesting to investigate to what degree structural idiosyncrasies of operators in OTC markets affect the efficiency of prices.

The football betting market literature has exclusively focused on analysing prices from position-taking bookmakers and more recently from betting exchanges. We have shown that position-takers' odds appear to be subject to marketing related objectives, which are likely to lead to non-exploitable anomalies. The EMH on other hand relies on the assumption that informed traders force prices to converge to the assets' fundamental values. Consequently, markets that restrict their trades do not form an appropriate setting for the investigation of theories related to the EMH since potential anomalies are likely arise from trading restrictions on the very individuals who are capable of removing the biases. On the other hand, such barriers do not seem to be imposed by book-balancing bookmakers. Therefore, we suggest that future research should focus on these bookmakers, as they create an environment which is line with that of exchange-based financial markets. More specifically, we encourage researchers to employ close to kick-off odds when analysing such markets, as liquidity is at maximum at that point in time. In that way, researchers will also overcome issues associated with the inconsistency of volumes across games in betting exchange markets.

In that context, it will be interesting to see whether strategies that utilize fundamental information, which are shown to generate (theoretical) positive returns from position-taking bookmakers, achieve similar results when employed using book-balancers' closing odds. Moreover, our research on the theoretical weak form anomalies can be extended, as researchers could investigate more closely under what circumstances, the "arbitrage mirage" arises due to a delayed response by position-takers in incorporating "fresh" information into their odds and in which cases it is the product of position-takers offering "promotional" odds. Finally, it will be of interest to observe whether, in the face of increasing competitive pressure from book-balancers(as the latter attempt to establish their brand name in Europe), position-takers will further decrease their transaction costs and increase the degree to which their prices' converge to those of book-balancers.