Modelling football match results and testing the efficiency of the betting market

Researched by

Declan Sheridan

Submitted to

National University of Ireland, Maynooth

The Department of Economics, Finance and Accounting

August 2012

This thesis is submitted in partial fulfilment of the requirements for a

Master of Literature in Economics

<u>Head of Department:</u> **Prof. Rowena Pecchenino**

<u>Supervisor:</u> **Prof. Donal O'Neill**

Contents

Acknowledgements

This thesis owes its greatest debt of gratitude to Professor Donal O'Neill, for his tireless guidance and expertise throughout the project. Without his commitment, energy and enthusiasm, none of this could have been possible. Throughout the year, he provided me with encouragement, sound advice and lots of great ideas. I would have been lost without him.

This research project would not have been possible without the aid of Dr. Tom Flavin and Professor David Madden. It was abundantly helpful to get such detailed feedback and guidance. As a result, I was able to eliminate some of the most obvious omissions of the paper.

I would also like to thank all of my family and friends for helping me each step of the way. I would especially like to thank my parents. They raised me, supported me and helped me in so many ways.

Lastly and most importantly, I would like to dedicate this thesis to my daughter, Ailbhe Lilly Sheridan and my girlfriend, Hazel O'Boyle.

Summary.

The research models football results using an ordered probit regression. The football market differs from that of horse racing in that it is typically fixed odds in nature. The betting prices generally remain unchanged in relation to bettor demand. Created by the bookmakers, this added risk exposure generates ample opportunities to uncover inefficiencies in the market. A dataset consisting of information from the most recent 11 years from six European countries has been used. Evidence is found of departures from weak-form efficiency. The betting odds on offer do not reflect their true probabilities and evidence is found showing favourite long shot bias consistent with past research. Also, the betting odds available for favourites tend to be overpriced. The evidence shows that a strategy of betting on home teams offers better value for the money than betting on draws or away teams. The bookmaker odds on offer include a premium charged to compensate risk exposure and include economic rents charged. As a result, the researcher was not able to capitalise on these betting inaccuracies because of this over-round mechanism. The researcher's ordered probit model suggests that there is available information not reflected in bookmaker prices. The research uses this information to create strategies capable of exploiting betting inefficiencies. Evidence shows that a strategy of betting on favourites and home teams that are overpriced provide positive returns.

1. Introduction

The bookmaking industry is a significant employer in Ireland with approximately 5,000 staff employed in retail shops and an additional 1,000 staff employed in head office operations[1]. It is a multi-million euro industry that has undergone drastic change in recent times predominantly because of the emergence of online betting. Licensing for online gambling only came into force in Ireland in 2003. The Irish Bookmaking Association (IBA) reported that between 2008 and

[1] http://www.theirishfield.ie/site/images/articlefiles/1315578040IBA-Budget-Submission-2012.pdf

2009, the retail betting market suffered a decline of 15.5% while non retail betting (online) increased by 13.5%. With the increase in mobile applications and the subsequent ease of availability, mobile phone betting grew by over 300% in 2010. Punters are choosing to bet online, as opposed to betting in local retail shops, as they can easily make use of the best betting odds online. The IBA, the representative body of the bookmaking industry, has stated its concerns in relation to the growth of online betting facilities in Ireland. It claims that, due to tax differences, online facilities have an unfair advantage over retail outlets. Currently, retail outlets face a 1% tax on revenues, whereas telephone and online betting are not subject to betting taxation. The IBA suggests that failure to address this issue will result in job losses to retail outlets in Ireland. However, the net job change could go either way as a result of an increase in online betting.

Online betting can have a significant impact on profits. Although it is widely accepted that bookmakers are profitable, it is not uncommon for bookmakers to face substantial individual losses from time to time. On November 19th, 2011, bookmakers lost an estimated £20 million resulting from a succession of favourites winning[2]. Ladbrokes posted a 12.3% decline in operating profits for the year ending 2011, directly as an outcome of football results. Conversely, Paddy Power achieved record results during 2011 directly as a consequence of the expansion into new markets including Canada and Bulgaria[3]. Paddy Power posted a 28% increase in group revenues so far this year and opened 17 new betting shops in the U.K. They increased the number of online active customers by 40% highlighting the changing landscape of the betting industry and the need to use online facilities.

The advent of online betting facilities and increased competition amongst bookmakers, combined with the abolition of a tax, make it easier for punters to exploit betting inaccuracies.[4] Armed with a greater wealth of knowledge and information, punters can now place bets from the comfort of their home. Today bookmakers entice punters to join with introductory free bets

[2] http://www.ft.com/intl/cms/s/0/eca84c9c-62ff-11e1-b837-00144feabdc0.html#axzz20suoIq00
[3] http://www.ft.com/intl/cms/s/0/c502acd6-a018-11e1-90f3-00144feabdc0.html#axzz20suoIq00
[4] In late 2001 the betting tax was abolished in the U.K.; it was previously 9% of the stake or of the winnings. The punter had the choice of paying a small certain amount or an uncertain larger amount.

and other promotional offers. Only bookmakers offering competitive odds will survive. Odds comparison sites have enabled bettors to receive the best odds available.
A key feature of football match betting is that odds tend to be fixed odds with the odds for games set days in advance. It is rare that fixed odds change in response to bettor demand, unlike in horse racing, where bookmakers continually balance their books according to bettor demand. Therefore, fixed odds betting present a greater risk exposure for bookmakers to inside information. They face an exposure to new information after the setting of odds.

This paper aims to exploit departures from efficiency by creating strategies capable of generating positive returns. The research models football results using an ordered probit regression based on the most recent 11 years of data from six European countries. Evidence is provided that the model is capable of accurately describing the data available. The research provides details in relation to the betting industry and discusses bettor demand. It assumes that, where available, bettors use the best odds available to them from 11 bookmaker odds. The findings point to the favourite long-shot bias. The betting odds available for favourites are too generous relative to their probabilities. Similarly, the betting odds available for home teams are overpriced. Bettors are unable to exploit these inefficiencies because of the existence of margins charged. Using the model, the research builds upon these betting inaccuracies. The model suggests that there is available information not reflected in bookmaker odds. According to the model, a strategy of betting on favourites and home teams that are overpriced provide positive returns.

Section 2 discusses the empirical approaches used to model football outcomes. It outlines the ordered probit model used. It details the betting industry and discusses past research into tests of weak-form efficiency. Section 3 describes the dataset used and the creation of variables. It analyses the bookmaker odds available to punters. Section 4 reports the results of the ordered probit model and the weak-form efficiency tests. Section 5 creates betting strategies designed to generate positive profits. Section 6 gives a summary and conclusion of the paper's key findings.

2. Literature Review

2.1. Modelling football outcomes

The modelling of football matches involves creating a model capable of accurately describing a past dataset. Dixon and Coles (1997) state that a good model should reflect different team abilities, factor for the home team advantage and account for recent team form. Although team strength can never be known with certainty, each team has some inherent qualities. Football teams that possess favourable characteristics such as high goal scoring and low goal conceding records should be more likely to win than teams with poor characteristics. There is a distinct home advantage found in football results. Home teams tend to win more games than away teams when accounting for heterogeneity of teams. Clarke and Norman (1995) use least squares to model individual match results in the English League. They estimate home ground advantage for each team in addition to individual team ratings. Home advantages for all teams in the English Football League from the 1981-82 season to the 1990-91 season are calculated. The dataset consists of 20,306 matches. They find that home advantage is quite variable across teams and from year to year. In some years, some teams have a negative home advantage. However, on average the home ground advantage was estimated to be worth just over 0.5 of a goal. They found that there was no division effect. They found that home advantage may have a greater effect on winning than on goal difference. Whatever produces the home advantage phenomenon tends to operate more efficiently in determining winners rather than winning margins. They found that clubs with special facilities had significantly higher home advantages and that London clubs had less than average home advantages.

Pollard and Pollard (2005) analysed various European leagues for evidence of a home advantage. They found that in home advantages, there was a wide regional variation. In addition, and perhaps more importantly, they found that crowd size had no impact on the level of home advantage. Home advantages recorded in the top two divisions in the English, French, German, Italian and Spanish leagues were similar despite higher average attendance being recorded in Division one games. Their research revealed that home advantage in the premier league was 60.7% with an average attendance of 31,009 whereas the home advantage calculated in Division one was 61.2% with an average attendance of 14,160. Finally, they found that home advantage decreased from 67.9% for the period from 1888 to 1900 to 60.6% for the recent period from 1992 to 2002. It was speculated that this may be due to a decrease in travel costs resulting from improved infrastructure and transport for away team supporters (Pollard and Pollard, 2005).

Neville, Balmer and Williams (2002) investigate the phenomenon of home advantage further by examining crowd noise. They suggest that crowd noise may offer an explanation for the home advantage phenomenon and also examine the influence of crowd noise on referee's decisions. The presence of crowd noise influences referees when assessing tackles or challenges recorded on videotape. A binary logistic regression is used to assess the effect of crowd noise and years of experience on referees' decisions. It was found that those viewing the challenges with background noise were more uncertain in decision making and awarded fewer fouls (15.5%) against home teams compared to those watching in silence. The avoidance of potential crowd displeasure is stated as a possible factor when decisions are made in favour of the home team.

Forty qualified referees were asked to assess the legality of 47 challenges and incidents recorded during an English Premier League game between Liverpool (home) and Leicester (away) from the 1998/1999 season. The referees were randomly allocated to either a noise group featuring crowd noise with no commentary or a silent condition group. Twenty-two referees were exposed to crowd noise and 18 referees were exposed to a silent condition group. The participants were asked to give their opinion on whether the 47 challenges were either legal (no foul) or illegal (foul). The question was posed: If the challenge was illegal, was it a home foul or an away foul? An uncertain option was also given, which would be taken as no foul in a live game. This gave four categorical response variables (1) home foul (2) away foul (3) no foul (4) uncertain. The analysis estimated probabilities or odds associated with the four categories and how these probabilities would vary due to differences in the independent variable. A binary logistic regression was used to assess the effect of the independent variables (crowd noise and years of experience) on each outcome variable, separately. They found that the silent group were more certain with fewer uncertain responses and awarded a greater number of fouls against home players. The group of referees watching the video with crowd noise awarded 15.5% fewer fouls against home players. The number of years of referee experience was found to have a significantly positive effect on the number of fouls awarded by referees. It was found that rather than penalising away players more, the dominant effect of crowd noise influenced referees to penalise home players less.

Team ability should be measured dynamically as opposed to being measured by a stagnant measure. Teams displaying good recent form should be more likely to win than teams displaying poor recent form. Team ability may vary with time as a result of many factors. Managerial changes, new transfers or depreciation of current team fundamentals offer likely explanations to why team ability may vary with time.

Audas et al., (2002) measured the impact of managerial change on team performance using an ordered probit regression. They included a managerial change dummy variable among their covariates. The empirical analysis focused on ‘within season’ changes. On average, a ‘within season’ change of manager tended to have an adverse effect on the results of matches played during the remainder of that season. They found that, on average, teams that changed their manager ‘within season’ under performed over the following three months of games. Given this adverse effect, they questioned why ‘within season’ managerial change was so common. They found that 70% of all managerial changes between 1972 and 1997 took place within season. In order to deliver improvement in the medium or long term, individual team owners may have been prepared to tolerate reduced team performance in the short run.

It is speculated that new managers may have preferred styles of play inconsistent with current playing staff. It may be that these short-term effects may in fact reverse in the long run. A change in manager is common sense if a team is underperforming and there is a potentially more effective manager as a replacement. A disadvantage is that a change in leadership could be disruptive, making matters worse in the short run. Audas et al., (2002) are careful to control for mean-reversion: Teams that experience poor runs of results eventually recover, whether they change their manager or not. Deciding to remove a manager in the middle of a series of poor results and attributing subsequent returns to form from this decision can represent a form of identification error. After controlling for mean-reversion, teams that changed managers within a season subsequently tended to perform worse than those that did not.

In addition to the above classifications of a good model, it is argued that the inclusion of a measure of teams' past form against a particular opponent is of great importance. It is possible that some teams may struggle against a particular opponent. Football experts often analyse past encounters between opponents in an effort to predict match outcomes. These factors should be incorporated into a model of football outcomes.

There have been two distinct approaches used in the modelling of match results: the direct and indirect methods. The direct method uses ordinal regression models such as the probit and logit, to predict the ordered response outcome, win/draw/loss. The indirect method examines the exact scores of games. It models the distributions of goals scored by each team, either independently or dependently. It assigns probabilities to particular scores and inferences are made to the most likely outcome. Moroney (1956) and Reep et al., (1971) were among the first to model the outcomes of football games. They used an indirect method of modelling match results using the negative binomial and Poisson distributions to model the number of goals scored in matches. If the model predicts home team i to score 1 goal and away team j to score 1 goal with greatest probability, the model predicts a draw indirectly. However, their analysis failed to account for varying team abilities and recent form.

Maher (1982) incorporated team specific form using strength indicators to model the outcomes of individual matches. He created a model in which the home and away goals scored follow independent Poisson distributions; their means reflected the attacking and defensive strengths of the two teams. The mean of the Poisson varied according to the quality of each team. Unfortunately, there was a problem of interdependence since the goal scoring/conceding distributions of each team follow independent distributions. Interdependence refers to the number of goals scored by the home team being dependent on the number of goals that the away team scores. There is a correlation between the home and away team scores. The match is akin to 2 separate games in opposite ends of the pitch, but to the teams concerned, the result is all that is important.

Maher (1982) compared observed and expected frequencies of goals scored and stated that the independent Poisson model gave a reasonably accurate description of football scores. However, he concluded that a Bivariate Poisson model with correlation of about 0.2 would give an improved description. One drawback to the model discussed is that, although team strength is allowed to vary for each team, it is held constant throughout a particular season. It is not allowed to vary over time and it is unlikely that team strength would remain constant throughout a season. As discussed, a team's inherent ability may change as a result of transfers, depreciation of player capital, managerial changes or any unforeseen circumstances. It is also likely that teams playing well may have added momentum going into future games.

One of the first papers to model the outcomes of football matches using discrete choice specifications was by Kuk (1995). He used a linear paired model in which n players are compared with each other in pairs. The method is based on matching the number of home wins; home draws, away wins and away draws with their expected values. Using a dataset for the 1993-1994 Premier League, with only a table for the number of wins, draws and loses for each team both at home and away he estimates individual team strengths. It was not possible to estimate the parameters by the maximum likelihood estimation as the results of every match are not available. He stated that a common home advantage model is inadequate as it fails to account for teams that did worse at home than away.

Konning (2000) used an ordered probit model including the home ground advantage, although each team's strength indicator is assumed constant, and independent of the opponent. This paper argued that team strength indicators should vary over time and, as a result, football teams displaying poor recent form would be less likely to win than teams playing well. Graham and Scott (2008) used an ordered probit model to predict English football matches and compared predictions with odds of William Hill. Their model was based on Konning's approach and they allowed for changes in team strength by down weighting past results in the maximum likelihood equation. Their dataset consisted of 11,000 English League matches in the four divisions from 11th August 2001 to 26th November 2006. They found that the William Hill implied probabilities outperformed the dynamic probit results model. This result suggests that if bookmakers are prone to bias and irrationality when setting odds, the extra information they possess more than makes up for it. The information they possess may reflect knowledge of player purchases, sales, injuries or managerial changes.

Dobson and Goddard (2003) performed diagnostic tests for normality, heteroskedasticity and structural stability for the ordered probit model applied to English league soccer results. Using a dataset consisting of 60,932 matches played in the English professional leagues between 1970 and 1999, they found that the unsystematic component in match results was normal and homoskedastic, but that the ordered probit model parameters reflecting team strengths were found to change significantly within the soccer season.

Dobson and Goddard (2003) tested for persistence in match results. They asked whether a sequence of good results tended to build morale and increase the probability that the next result was good, or did it breed complacency resulting in a lower probability of a favourable result. They also inquired whether a poor sequence of results lowered morale and increased the probability of a poor result or did it inspire a greater effort resulting in a lower probability of a poor result. Controlling for team heterogeneity, they found that sequences of consecutive results are subject to a negative persistence effect.

2.2. Ordered Probit Model

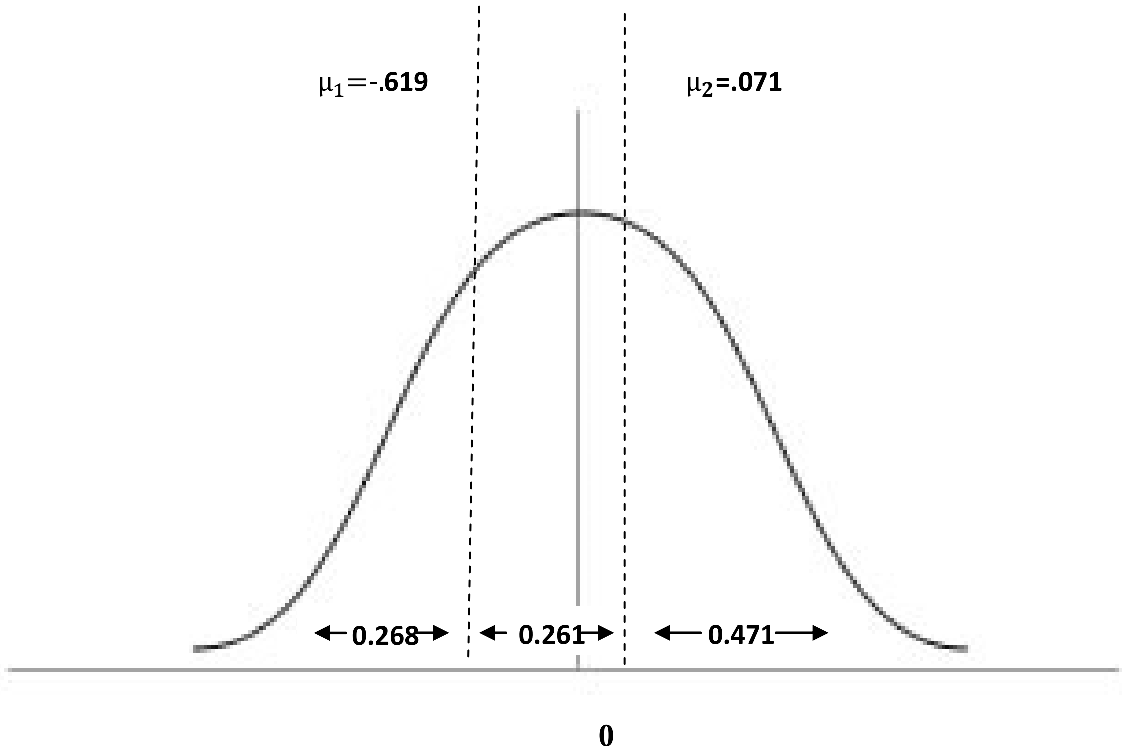
This research uses an ordered probit model to estimate football outcomes in six European countries over an 11-year period. The ordered probit model estimates the probability of a home, draw and away outcome for each observation. Although, individual team strength can never be known with certainty, it can be estimated using past match details. It is these estimates of team ability that alter the probability of observing a particular outcome. It is clear that teams of high ability should be more likely to win than teams of low ability. Where there is no information about relative team strengths the hypothesis could be that all three outcomes (home, draw and away) are equally likely to occur. However, Table 1 below suggests that this would be a naive form of estimation given that there is clear evidence of a home advantage phenomenon. It is clear that home outcomes are more likely to occur than a draw or away outcome. For example, the historic percentage of home wins was 47.1% in the English Premier League, whereas the percentage of draw and away wins was 26.1% and 26.8% respectively.

Table 1. Historic proportions of outcome (2000-2001 Season to 2010-2011 Season)

	Outcome Probabilities		
	Home (%)	Draw (%)	Away (%)
English	47.1%	26.1%	26.8%
French	46.5%	29.7%	23.8%
German	47.3%	24.6%	28.1%
Italian	46.3%	28.5%	25.2%
Scottish	44.3%	23.2%	32.5%
Spanish	48.2%	25.0%	26.8%

Fig 1 models the distribution of outcomes in the English Premier League as standard normal. Since we assumed the distribution to be standard normal, the probability of drawing a value within a chosen range can be measured precisely. We need to divide the normal distribution into regions that define the probabilities 0.268, 0.261 and 0.471. Then we start with a standard normal density and find the thresholds that will divide this density into regions with areas 0.268, 0.261 and 0.471. We need to solve for Φ^{-1}(0.268), the inverse of the standard normal density function, in order to get the value of μ_1, threshold one. In order to find μ_2, we need to solve for Φ^{-1}(0.268+0.261). The probability of observing a value less than μ_1=-0.619 is 0.268; the probability of observing a value greater than μ_2=0.071 is 0.471. The probability of observing a value between μ_1 and μ_2 is 0.261. This is equivalent to the probability of an away win, a home win and a draw in the historic data, respectively. This would be the expected estimated threshold from a model with no team specific factors. The thresholds simply divide the standard normal into appropriate proportions of each category.

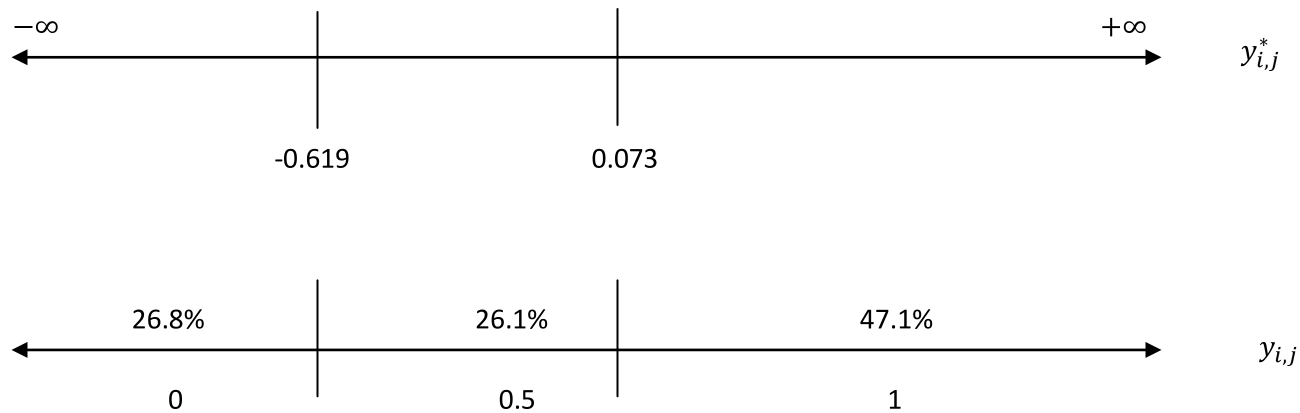
Fig 1. **Standard Normal Distribution for the English Premier League**



Suppose now, that we have information about the teams contesting the match. Available information on relative team strength should change the probabilities of home wins, draws and away wins accordingly. A latent variable is one that is unobservable directly. A team's inherent quality is unknown and is therefore a latent variable. However, it may be estimated indirectly using observable variables. The latent continuous variable $y^*_{i,j}$ is the relative team strength of home team *i* versus away team *j*. It is a linear combination of some predictors, $x_{i,j}$ where $y^*_{i,j} = x_{i,j}'\beta_{i,j}$.

Fig 2 outlines how the continuous variable $y^*_{i,j}$ is mapped into observed ordinal outcomes $y_{i,j}$. It shows the corresponding threshold levels that account for each proportion of outcomes

Fig 2. Mapping latent continuous variable into observed ordered outcomes.



This research uses predictors that reflect home and away team strength. These predictors change the value of $y^*_{i,j}$. Teams of higher home team strength will increase the value of $y^*_{i,j}$, increasing the probability of a home win. Conversely, greater away team strength will decrease the value of $y^*_{i,j}$ resulting in a smaller probability of a home win and a larger probability of an away win. In addition to relative team strength, the probability of observing a particular observation $y_{i,j}$ will also depend on a random factor denoted $\varepsilon_{i,j}$[5]. $y_{i,j}$, the observed ordinal response is a function of the unobserved latent variable $y^*_{i,j}$ and a random factor denoted $\varepsilon_{i,j}$. $y^*_{i,j}$ may assume an infinite range of values, whereas $y_{i,j}$ takes on values 0 for an away win, 0.5 for a draw and 1 for a home win. Table 2 shows how changes in $y^*_{i,j}$, relative team strengths of i and j , alter the probability of observing each particular outcome $y_{i,j}$. Note that when $y^*_{i,j} = 0$, there is no information available on team ability.

[5] Greene (2000) provides a detailed discussion on ordered probit models.

Table 2. Varying the values of the latent variable and the probability of observing each ordinal outcomes.

Varying Latent Variable and Probability of observing outcomes									
	$y_{i,j}^*$ = -1	$y_{i,j}^*$ = -0.75	$y_{i,j}^*$ = -0.5	$y_{i,j}^*$ = -0.25	$y_{i,j}^*$ = 0	$y_{i,j}^*$ = 0.25	$y_{i,j}^*$ = 0.5	$y_{i,j}^*$ = 0.75	$y_{i,j}^*$ = 1
P($y_{i,j}$ = 1) Home	0.142	0.205	0.283	0.373	0.471	0.570	0.665	0.751	0.823
P($y_{i,j}$ = 0.5) Draw	0.21	0.243	0.264	0.271	0.261	0.237	0.203	0.164	0.124
P($y_{i,j}$ = 0) Away	0.648	0.552	0.453	0.356	0.268	0.192	0.132	0.085	0.053

Increasing the value of $y_{i,j}^*$ increases the probability of observing a home win and decreases the probability of an away win. Higher values of $y_{i,j}^*$ correspond to increased levels of home team ability relative to away team ability, therefore, low values of $y_{i,j}^*$ correspond to an increased probability of an away win and a decreased probability of a home win. Equations (1) (2) and (3) highlight the link between the observed ordinal response and the latent variable metric. When the latent variable metric, $y_{i,j}^* + \varepsilon_{i,j}$ is above the threshold μ_2, the model predicts a home win. When $y_{i,j}^* + \varepsilon_{i,j}$ falls between the values of μ_1 andμ_2, the model predicts a draw. When the value of $y_{i,j}^* + \varepsilon_{i,j}$ is less than the cut off point μ_1 the model will predict an away win.

Home win $(y_{i,j} = 1)$ if $\mu_2 < y^*_{i,j} + \varepsilon_{i,j}$ (1)

Draw $(y_{i,j} = 0.5)$ if $\mu_1 < y^*_{i,j} + \varepsilon_{i,j} < \mu_2$ (2)

Away win $(y_{i,j} = 0)$ if $y^*_{i,j} + \varepsilon_{i,j} < \mu_1$ (3)

The probabilities of each ordered outcome will depend on the amount of area under the curve that each outcome has, with Standard Normal Distribution, $\Phi_{i,j}$. Instead of using a Standard Normal Distribution of mean 0 and standard deviation 1, we use the value of $y^*_{i,j}$ as the mean. Increasing the value of $y^*_{i,j}$, the relative team strength of *i* versus *j* will increase the probability of a home win and will decrease the probability of an away win. Increasing the value of $y^*_{i,j}$ will move area out of the away category and move area into the home category. Low values of $y^*_{i,j}$ decrease the probability of home wins and increase the probability of away wins. Adopting Dobson and Goddard's (2001) notation, the probability of a home win, draw and away win are given below.

Home Win

$$P[y_{i,j} = 1] = P(\varepsilon_{i,j} > \mu_2 - y^*_{i,j})$$

$$= 1 - \Phi_{i,j}(\mu_2 - y^*_{i,j}) \quad (4)$$

Draw

$$P[y_{i,j} = 0.5] = P(\mu_1 - y^*_{i,j} < \varepsilon_{i,j} < \mu_2 - y^*_{i,j})$$

$$= \Phi_{i,j}(\mu_2 - y^*_{i,j}) - \Phi_{i,j}(\mu_1 - y^*_{i,j}) \quad (5)$$

Away win

$$P[y_{i,j} = 0] = P(\varepsilon_{i,j} < \mu_1 - y^*_{i,j})$$

$$= \Phi_{i,j} \left(\mu_1 \text{-} y^*_{i,j} \right) \qquad (6)$$

Under the assumption of independent observations, the sample likelihood is just the product of these probabilities. Note that the probability of an observation depends on which category of *y* it falls into. To write the joint likelihood function of the sample, it is necessary to write it so that it associates the correct probability for each observation. An easy way to do this is to define M dummy variables for each of the M categories. In the case of football results there are three categories, home, draw and away, so $d_{im=1}$ if observation i falls into category M and $d_{im=0}$ otherwise. The likelihood is then the product over all M and all i of the probabilities raised to d_{im}. This picks out the correct probabilities associated with each observation. The likelihood of the sample is simply the product of the individual likelihoods. The product is at a maximum when the most likely set of *p's* is used. The maximum of L is solved by differentiating the function with respect to each of the betas and setting the partial derivatives equal to zero, or the values of β_1, ……, β_K that provide the maximum of L . Because the likelihood function is between 0 and 1, the log likelihood function is negative. The maximum to the log-likelihood function, therefore, is the smallest negative value of the log likelihood function given the data and specified probability functions.

$$\mathrm{L} = \prod_{m=1}^{M} \prod_{i=1}^{N} \mathrm{P}(y = m)^{d_{im}}$$

$$= \prod_{i=1}^{N} \mathrm{P}(y = 1)^{d_{i1}} \times \prod_{i=1}^{N} \mathrm{P}(y = 1)^{d_{i2}} \times \ldots\ldots \prod_{i=1}^{N} \mathrm{P}(y = M)^{d_{iM}} \qquad (7)$$

Thus the log likelihood is

$$\mathrm{LnL} = \sum_{m=1}^{M} \sum_{i=1}^{N} \ln\left[P(y = m)^{d_{im}}\right]$$

$$= \sum_{m=1}^{M} \sum_{i=1}^{N} d_{im} \ln[\Phi(\mu_m - x_{i\beta}) - \Phi\left(\mu_{m-1} - x_{i\beta}\right)] \qquad (8)$$

For the purpose of football outcomes, it can be written as

$$\ln L = \sum_{y=0} \ln[\Phi(\mu_1 - x_{i\beta})]$$

$$+ \sum_{y=0.5} \ln[\Phi(\mu_2 - x_{i\beta}) - \Phi(\mu_1 - x_{i\beta})]$$

$$+ \sum_{y=1} \ln[1 - \Phi(\mu_2 - x_{i\beta})] \qquad (9)$$

This forms the basis for our modelling of football results and allows us to predict outcomes from a set of characteristics of *x* such as average team shots on target.

2.3. Gambling Motives.

Bloch (1951) stated that gambling may depend on skill or on chance. Gambling has an ancient history, with artefacts relating to various games of chance found in the archaeological remains of the Chinese, Egyptian and Sumerain cultures. Frey (1984) stated that there exists a popularity of gambling among all class, racial and ethnic lines. Sauer (1998) stated that there may be a pleasure of betting. Small bets that are not life changing may give the bettor some sort of satisfaction or utility of gambling. This may explain why bettors engage in betting activity in spite of the negative expected returns exhibited in betting markets. Smith and Preston (1984) summarised some of the leading motives for gambling existing in the past literature. They found that gambling may provide bettors with a means to rise above their present economic class through monetary gain. Gambling may also provide prestige or influential status from winning. Bettors may also exhibit recreational benefits from gambling outside of monetary or status levels as there may be a sociability of informal gatherings and friendships through gambling. They highlight that gambling may exist as a learned role in which gambling may be taught by others who view this behaviour as traditional. Gambling may also afford the bettor the opportunity to use his or her intellect in making decisions. This intellectual exercise (figuring out the odds and then making the appropriate wagers) may provide an explanation for gambling behaviour. Smith and Preston (1984) also assert that the bettor may have a genuine belief that he or she is especially lucky and that they are an inevitable winner because they are clever or possess some different inherent characteristic. Contrastingly, bettors may find masochism or guilt in losing. The bettor may find pleasure in losing.

Levitt (2004) compared gambling markets to financial markets. In both settings, individuals have heterogeneous beliefs and information and seek to profit through trading. However, for financial markets, prices change frequently in accordance with demand while bookmakers simply announce a price, after which the adjustments made tend to be typically small and infrequent. If the bookmaker sets the wrong odds on a sporting event, they could lose money, even in the long run.

2.4. Market Efficiencies in Sports

Sauer (1998) discussed betting markets for major sports with particular reference to horse racing. The pari-mutuel system is exclusively used by racetracks in North America, France, Hong Kong and Japan and coexists with the bookmaking market in Australia, Great Britain and Ireland. The pari-mutuel system operates as follows: A predetermined percentage is taken out of the betting pool to cover market maker costs and the remainder is returned to winning bettors in proportion to individual stakes. For example, the odds will reflect the bettor demand of the event – the syndicate bears no risk exposure as it takes a predetermined percentage of the overall stake regardless of the outcome. Those bettors who successfully predict the outcome receive a return proportionate to their stake. Conversely, bookmakers offer a set of payoffs conditional on the outcome of a given event. The payoffs offered may change during the betting period but the payoff to each bet is determined at the time each bet is placed. The return of the bet is known at the time of the wager unlike in the pari-mutuel system. Therefore, individuals who make bets large enough to affect the odds naturally prefer to bet with bookmakers.

Hurley and McDonough (1995) investigated market efficiency in horse racing. They identified two types of risk neutral bettors: informed and uninformed bettors. Informed bettors know the true probability of occurrence and uninformed bettors are unable to differentiate with the same precision.

For example, where there are two horses in a race, horse one has a true probability of 0.7 (favourite) and horse two has a true probability of 0.3 (long shot). The uninformed bettor will not be able to distinguish between the horses and will have a subjective probability of 0.5 for each horse. Suppose the number of uninformed bettors is very substantial relative to the number or betting volume of informed bettors. In this case, the odds offered by bookmakers may approximate the opinions of the uniformed bettors. There may exist opportunities for informed bettors to make profits at the expense of uniformed bettors.

Favourites are defined as the outcomes that are most likely to occur according to betting odds. Long shots are defined as the outcomes that occur least likely according to betting odds. The favourite-long shot bias is a phenomenon that has been attributed to both horse racing and soccer betting markets. It states that the betting odds on offer for favourites are too generous relative to their probability of occurring. Hurley and McDonough (1995) provided evidence of this favourite-long shot bias and offered several explanations for its existence. Firstly, bettors may overestimate the chances of long shots winning. Secondly, some bettors may enjoy calling long shots winning or may pick horses for irrational reasons such as the horse's name.

Vergin and Sosik (1999) analysed home field advantage in the NFL for the period 1981-1996. Throughout that period home wins accounted for 58% of the observations. This could be a result of a learning process or 'familiarity with stadium' effect or a travel factor due to physical or mental fatigue. It could also be due to a crowd factor because of social support. Unlike betting on soccer, betting on NFL follows a point spread line for which there is an estimated point spread difference between the teams.

Vikings (-5pts) versus Bears (+5pts)

A bet on Vikings pays off if the Vikings win by six or more points. A bet on Bears only pays off if the Bears lose by four or fewer points. A Viking win by five points is called a push, in which the original stake is returned. The opening spread is set in Las Vegas with the consensus of a small number of expert analysts.

The bookmaking commission is estimated by the 11 for 10 rule. A bettor must pay $11 to win $10. Therefore a bettor winning 50% of their bets would lose money. A win rate of 52.4% is needed to break even. They found evidence of the reverse favourite-long shot bias.

A strategy of betting on long shots outperforms favourites unlike in horse racing and soccer markets. Vergin and Sosik, (1999) found that betting on home teams beats the spread 49.9% of the time and that a strategy of betting on home underdogs (52.5% win rate) outperformed home favourites (48.6% win rate). They furthered the discussion by analysing Monday night games and Play-off games separately. These are typically games of higher importance as the best teams play on Monday nights due to television rights. Similarly, the best teams take part in Play-off games. They found that home underdogs had a win rate of 63% and home favourites had a win rate of 57.1% in Monday night games. In addition, home teams had a win rate of 59.6% on Play-off games. The study highlights that the point spreads on average are efficient but that the home field advantage increases with intensity of interest in the game. Home teams tend to react to the big games better than away teams, possibly due to abnormal intensity or crowd noise.

Gray and Gray (1997) found that there was evidence of overconfidence in favourites in the NFL market. Using a dataset of 4,219 games from 1976-1994, they found that the favourite won by less than expected (according to the spread). On average the favourite gave up a 5.63 points spread but only won by 5.2 points. They also found that home teams and underdogs were more likely to beat the spread than away teams and favourites. A strategy of betting on home underdogs provided a positive return of 4% in excess of commission. Similarly Woodland and Woodland (2001) found that bettors were inclined to over bet favourites in the NHL market. A strategy of betting on teams that were heavy underdogs provided profitable returns as high as 11%.

Woodland and Woodland (1994) analysed the efficiency of the Major League Baseball market from 1979-1989. This market differs from that of the race track as there is no uncertainty of payoff as in the pari-mutuel market. Furthermore commission charged (vigorish or juice) tended to be a lot lower than that on the race track. Unlike the NFL, the margin of victory is not critical. Baseball betting is only concerned with the eventual winner. An example of an odds wager would be -190, +180. A gambler could wager $1.90 on a favourite to win $1.00 while they could wager $1.00 to win $1.80 on the underdog. When the books are in balance the bookmaker receives $0.10, the differential. When the books are not in balance the bookie is an active participant in the gambling process. Their receipts are dependent on the final outcome of the game. Woodland and Woodland (1994) found evidence of the reverse favourite-long shot bias evident in the data. A strategy of betting on the underdog yields expected losses significantly lower than those implied by market efficiency.

Using NHL data from 2005-2008, Rodney and Weinbach (2012) investigated the betting market. They used actual betting percentages by bettors on favourites and underdogs taken from four real online bookmakers. Combined data from four online bookmakers show the percentage of bets placed on favourites and underdogs by subscribers to online betting channels. They found that bookmakers did not price their odds to balance the book. There were significant imbalances of actual bets compared to expected bets proportional to the odds set by bookmakers. The odds were not set by bookmakers based on the percentage wagered by bettors. It was found that bettors seemed to overestimate the odds of away favourites winning. Bookmakers did not appear to set prices to exploit these biases. Prices were set closer to their true probability of outcome. They found that a contrarian betting strategy provided positive returns. In situations in which the public over-bet on a favourite, compared to projected percentages based on odds, a strategy of betting against the public will provide positive returns. However, Rodney and Weinbach (2012) state that any betting strategies in the long run would not be profitable as bookmakers assigned probabilities in line with true probability and not with bettor biases. This could be to discourage informed bettors. Alternatively, bookmakers may like to earn consistent small returns from bettors rather than a large once off return that might discourage uninformed bettors or the general public.

Gray and Gray (1997) found that teams that were performing well over their previous four games were less likely to beat the spread in the NFL. This suggests that the market overreacts to recent form. Similarly, using a dataset of professional basketball games from 1983 to 1986, Camerer (1989) suggested that bettors overestimated recent form. Teams with winning streaks would have point spreads that were too high. This is referred to as the 'hot hand fallacy'.

Using a dataset of the opening and closing betting lines and the match outcomes of 9,940 games for the 1985-1994 seasons in the NBA, Gandar et al., (1998) found evidence that closing betting lines forecast game outcomes better than opening betting lines. This evidence suggests that there are biases in opening lines and that trading appears to remove them. This would suggest that informed traders are influential in the market. Informed bettors are able to identify teams whose chances of winning against the initial betting line are undervalued. However, it is unlikely that bookmakers, using public information, misprice opening lines so frequently, therefore, the betting public may possess either private information or are superior to bookmakers in processing available information in the market.

2.5. Weak-form efficiency in Football

The bookmaking industry presents many challenges. How do bookmakers compute odds? What are punters' attitudes and behaviour to risk? How do betting odds vary with expected bettor demand? The football market typically uses fixed odds betting. This differs from horse and dog racetrack betting, which uses starting prices. Pope and Peel (1989) claim that betting in fixed odds markets differs in at least two important respects regarding starting prices. Firstly, fixed odds are not determined jointly by buyers and sellers; they are chosen by the bookmaker days before. Secondly, once posted, the odds offered will be invariant to the volume of bets placed or new information that could alter the probabilities. As betting odds are fixed, the bettor has more information concerning form than bookmakers have. This constitutes a greater risk exposure for bookmakers to inside information.

Hodges and Lin (2009) discussed markets in which odds were not fixed such as horse racing. In this setting, bookmakers are able to revise odds to mitigate risk. Bookmakers choose to rebalance their books over time. Bookmakers may rebalance their books by changing the odds available for each outcome as bettor demand changes. If a bookmaker has a high liability on a particular outcome, they can lower the demand for that outcome by lowering the odds. Since football odds are set days in advance, bookmakers face a greater possibility of an unbalanced book.

Kuypers (2000) defined bookmakers as profit maximising agents. His model assumed perfect knowledge of information by bookmakers and punters. It assumed that bookmakers do not have private information outside the domain of punters. Private information may include player injuries that teams are keeping quiet. It is assumed that the bookmaker knows the punters' reaction function and will set odds to maximise their profit. This suggests that bookmakers already know the bettor demand of punters prior to setting odds. Unlike the assumptions of Pope and Peel (1989), there is no uncertainty of bettor demand in this model. Since bettor demand is known, and bookmakers are profit maximising agents, the bookmakers' odds may not reflect the best estimates of probability. This may explain possible inefficiencies found in past betting data. Milliner, White and Webber (2008) claimed that punters may not act rationally and that bookmakers may take this into account when pricing outcomes. For example, the punter may engage in sentimental betting−an activity of betting on teams they support irrespective of odds or probabilities. Although, it is difficult to assess bookmaker and bettor equilibrium, it is possible to construct probabilities associated with the odds quoted by bookmakers. Table 3 explains how price implied probabilities can be derived from bookmakers' odds.

Table 3. Constructing price implied probabilities from fixed-odds betting

1. Betting odds in form $\frac{a}{b}, \frac{c}{d}, \frac{e}{f}$ for Home, Draw, Away	
Inter Milan	$\frac{1}{1}$
Draw	$\frac{6}{4}$

Juventus	$\frac{3}{1}$
2. Chance underlined by betting market evaluated by $\frac{b}{a+b}$, $\frac{d}{c+d}$, $\frac{f}{e+f}$ for Home, Draw and Away. This includes profit margin charged by bookmaker.	
Inter Milan	0.5
Draw	0.4
Juventus	0.25
3. Price implied Probabilities = $\frac{chance}{\lambda}$ $where$ $\lambda = \frac{b}{a+b} + \frac{d}{c+d} + \frac{f}{e+f}$	
Inter Milan	0.4348
Draw	0.3478
Juventus	0.2174

Home win odds= $\frac{a}{b}$, Draw odds= $\frac{c}{d}$, Away win odds= $\frac{e}{f}$.[6] If I bet "b" on an Inter Milan win, I will receive my stake of "b" with a profit of "a "should Inter win. If Inter Milan loses the match or it is a draw, I will lose my initial stake of "b".

The chance underlined by the bookmakers' odds is calculated as follows: Home chance = $\frac{b}{a+b}$, Draw chance = $\frac{d}{c+d}$, Away chance = $\frac{f}{e+f}$

The chance of an Inter Milan victory is 0.5; the chance of a draw is 0.4 and the chance of a Juventus win is 0.25. In the example provided, the summation of these chances do not sum to 1; it sums to 1.15. This is because of the existence of a margin, a premium that bookmakers charge[7]. Fingleton and Waldron (1996) stated that the margin covers operating costs such as licence fees, labour and equipment expenses. The margin also compensates for risk exposure,

---

[6] Note that the English equivalent term Evens = $\frac{1}{1}$. In Europe, it is more common to find odds given in decimal form where $\frac{6}{4}$=2.5. The decimal notation includes the stake bet in its calculation.

[7] Makropoulou and Markellos (2007) stated that if the punter bet on all outcomes (home, draw and away win) they would lose a percentage of their initial stake.

bourn by bookmakers. Finally, bookmakers may earn monopoly rents because of collusion amongst themselves to avoid direct competition.

To find price implied probabilities. We constrain our probabilities to sum to 1 by dividing by the constant λ . Note that the margin is equal to $\lambda - 1$. In the example, the price implied probabilities of an Inter win, draw and Juventus win as 0.4348, 0.3478 and 0.2174 respectively. It is these probabilities we will use in our analysis of weak-form efficiency.

The margin is assumed constant across each outcome. Pope and Peel (1989) stated that the margins imposed are not uniformly distributed over each outcome. They stated that the burden of the margin tends to be placed on long shots. Long shots are defined as the outcome that is least likely to occur. This is due to either an increased risk exposure or because bettors have a high demand for long shots. This may be because bettors are wanting to gamble small proportions of income when gambling. Grant (2008) found that the margin will generally be between 0.05 and 0.15 for sports betting. He stated that the presence of a margin makes it difficult for bettors to exploit potential inefficiencies in prices.

Weak-form efficiency states that all historical information is captured in the odds of bookmakers.[8] Pope and Peel (1989) performed regression-based tests of betting odds using a linear probability model, and a logit model. The results of both methods provided evidence that the probabilities implied by the odds of bookmakers were not significantly different from their outcome probabilities. Goddard and Asimakopoulos (2004) used the regression-based tests shown in equations (10) and (11).

$$r_{i,j}=\alpha_r+\beta_r\theta_{i,j}^r+\varepsilon_{i,j}$$

$$\text{for } r=\text{Home, Draw and Away} \qquad (10)$$

[8] Note that semi-strong efficiency requires that prices of bookmakers’ odds reflect all public information. Strong form efficiency requires prices to reflect all private and public information available.

$\theta^{r}_{i,j}$ *f*or *r*= Home, Draw and Away are the implied probabilities of bookmaker odds. A weak-form efficiency test for Home odds, $\theta^{H}_{i,j}$, involves regressing r against the home odds, where *r=1* if the home team wins and 0 if the home team doesn't. A necessary weak-form efficiency condition is $\{\alpha_r = 0, \beta_r = 1\}$. The price implied probabilities equal their true values. They found little or no evidence of systematic departure from the weak-form efficiency conditions.

Furthermore, equation (11) investigates if information contained in the ordered probit model is reflected in bookmaker odds.

$$r_{i,j} = \alpha_r + \beta_r \theta^{r}_{i,j} + \gamma_r \left(P^{r}_{i,j} - \theta^{r}_{i,j}\right) + \varepsilon_{i,j}$$

$$\text{for } r = \text{Home, Draw and Away} \qquad (11)$$

$P^{r}_{i,j}$ for *r*=Home, Draw and Away is the probability of each outcome given by the ordered probit model. $P^{r}_{i,j} - \theta^{r}_{i,j}$ is the difference between the model and bookmaker probabilities. The model should not contain any additional information not already reflected in bookmaker odds. $H_0 : \{\alpha_r, \beta_r, \gamma_r\} = \{0, 1, 0\}$. They concluded that their model does not include any additional information not reflected in bookmaker odds. Interestingly, when the test results were applied to this research there is evidence of departures from weak form efficiency. This could be due to having a dataset of best odds and analysing best odds available. It could also be due to the fact that the most recent data available was used.

Although past regression-based tests have shown that odds are weak-form efficient, a number of inefficiencies have been uncovered, including the favourite-long shot bias. Favourite-long shot bias refers to a strategy of betting on all favourites as opposed to long shots in match outcomes. Cain, Law and Peel (2000) found evidence of favourite-long shot bias in the English fixed betting data. They tested outcomes and exact scores from the '91-'92 season using odds from William Hill. Their indirect model outperformed the bookmakers. They concluded that it was likely that bookmakers offered biased odds on unusual scores. Long shot scores of 5-1 or 6-1 will offer punters worse value than bets on favourites. Similarly, Vlastakis et al., (2009) claimed that favourites tended to be too generously priced. Firstly, bookmakers were more exposed to inside information on long shots than they were on favourites because of higher odds available for long shots. Secondly, bookmakers may have adjusted profit margins to tailor perceived bettor demand of the general public. Bookmakers may have maximised profits whilst setting odds that were not weak-form efficient.

Recently, match result forecasting models have been used to investigate if the incorporation of available information, including team strength indicators and in match statistics can outperform the forecasts of bookmakers. If betting odds are weak-form efficient, it should not be possible to gain positive returns on simple betting strategies. Dixon and Coles (1997) implemented an intuitive method of analysing the efficiency of the betting market. They assigned probabilities for each outcome in the indirect model in the normal manner and used a simple betting strategy. The strategy involved betting on a particular outcome of a match when the ratio of the model's probability to bookmaker implied probability for that outcome was greater than some predetermined value. They generated positive abnormal profits from betting on overpriced outcomes. This provided evidence against the weak-form efficiency of the market.

Kuypers (2000) used an ordered probit model to test the weak-form efficiency. The data comprised of the four English leagues throughout the 1993-1994 and 1994-1995 seasons. The model included lagged in-match statistics from performances of each specific season. Betting odds were also incorporated in his model as an explanatory variable. He investigated the profitability of the model's predictions using a simple betting strategy. The strategy bets one pound on an outcome of a particular game when the ratio of the model's predicted probability to the bookmaker implied probability for that outcome is greater than some pre-specified level,

X. Using his sample of results from 1993 to 1995, the strategy generated profits as high as 44% pre tax and 33% post tax.[9]

Past research has primarily dealt with an analysis of English data, whereas this paper adds to the literature by creating betting strategies for six top European divisions. A simple model is created to analyse 11 years of data. This model is capable of accurately describing the data. Strong evidence is found of departures from weak-form efficiency in betting odds. The favourite-long shot bias is evidenced in each of the six markets analysed. Similarly, home odds tend to be overpriced. The model used contains information not reflected in bookmaker odds. A strategy of betting on home teams that are overpriced generated positive returns in five of the six markets. Similarly, betting on home favourites that are overpriced provided positive returns in four of the six markets. This paper found systematic biases evident in betting data and uncovered betting strategies capable of exploiting them.

3. Data

The research data consists of the most recent 11 years of English, French, German, Italian, Scottish and Spanish football premier divisions.[10] The data comprise the 2000-2001 to 2010-2011 seasons. Each country has varying numbers of teams, games, bookmaker odds and in-match details available. Section 3.1 describes the data from each European market. Section 3.2 outlines how the variables of interest were created. Section 3.3 summarises the betting odds available in each European market.

3.1. Data

The English Premier League dataset consists of 4,180 observations. There are 20 teams competing any one season, each playing 19 home and 19 away games against their rivals. In any given season, there are 380 games. Due to the mechanism of promotion and relegation, there

[9] Apart from the abolition of betting tax since 1995, during the period 1993-1995, single bets on football games was not permitted. Single bets on football games were only allowed when the game was broadcast on television.

[10] All of the data and betting odds were obtained from the website www.footballdata.co.uk.

are 36 teams contained in this research dataset.[11] Seven teams competed all 11 seasons whilst eight teams only played one season. Teams competing for only one season were omitted from the ordered probit results because the recent match variable needed past encounters between opposing teams. The English dataset contained a full set of in-match statistics for each season. This includes home shots, home shots on target, away shots, away shots on target and match outcome against each opponent.

The French Ligue 1 contains results from 4,032 games. In the seasons 2000-2001 and 2001-2002, 18 teams competed. Since 2002-2003, 20 teams have competed. The dataset contains 33 teams of which eight competed each season and four competed only one season. For the seasons 2000-2001 to 2004-2005, there are no in-match details. A full dataset exists for the seasons 2005-2006 to 2010-2011.

The German Bundesliga 1 contains 3,366 results. The league consists of 18 teams playing 17 home and 17 away games against their rival. In any given season there are 306 games contested. The dataset contains 28 teams, eight teams played each season and two teams played only one season. For the seasons 2000-2001 to 2001-2002 there is a full dataset. 2002-2003 contains no in-match details. Seasons 2003-2004 to 2005-2006 have no information on home shots on target or away shots on target. Seasons 2006-2007 to 2010-2011 contain a full dataset.

The Italian Serie A has 3,504 observations. In seasons 2000-2001 to 2003-2004 there were 18 teams. Twenty teams have competed since the 2004-2005 season. The dataset contains 37 teams of which five teams played each season and six teams played only one season. The seasons 2000-2001 to 2004-2005 contain no in-match details. The seasons 2005-2006 to 2010-2011 contain all match details.

The Scottish Premier League contains 2,508 games. There are 12 teams in any given season. Each team plays 38 games. Unlike other European leagues, teams may play an uneven number of home and away games against rivals. The format of the Scottish Premier League is a little

[11] Each European league promotes and relegates three teams in any given season, except for the Scottish Premier League.

different from that of other leagues. Firstly each team plays three games against rivals totalling 33 games− one home and two away games or vice-versa. Then the 12 teams are split into groups of six based on standing in table. The top six teams play five games against each other. The bottom six teams play five games against each other. Only two teams have ever won the Scottish Premier League since its inauguration in 1998. The dataset contains 18 teams of which eight have competed each season and one has only played one season.

The Spanish La Liga 1 has 4,180 observations. There are 20 teams in any given season. Each season consists of 380 games. The dataset contains 35 teams of which nine have played every season and five teams have competed only one season. The dataset contains no in-match details for seasons 2000-2001 to 2004-2005. Full datasets have been obtained for the seasons 2005-2006 to 2010-2011.

3.2. Variable Creation

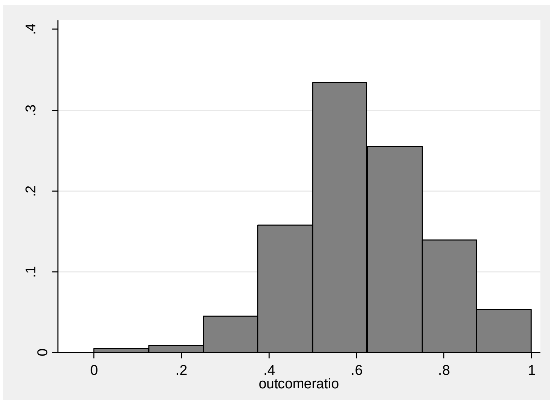
Using available match details, this paper creates simple home and away strength indicators capable of measuring team ability. It is anticipated that teams with high ability are more likely to win games than teams with low ability. The research allows team strength indicators to vary across time, as teams with good recent form should be more likely to win than teams with poor recent form. The research creates a recent match indicator that picks up recent winning and losing streaks against a particular opponent. Winning streaks against a particular opponent are anticipated to continue.

Outcomeratio is defined as the average of home team's most recent 19 home games.[12] As defined by the research's ordered probit model, an away win, draw and home win will correspond to the outcomes 0, 0.5 and 1 in the ordinal scale. A home win corresponds to 1 in the ordinal scale, so values closer to 1 indicate increased home team strength. It is anticipated that outcomeratio will positively affect the probability of a home win and negatively affect the probability of an away win. Fig 3 shows the distribution of outcomeratio. It is clear that the

[12] An outcomeratio for a home team with 18 home wins in the last 19 home games would be $\frac{18}{19}$=0.9473. Note that outcomeratio only contains information on home team's home games due to variability in home and away form.

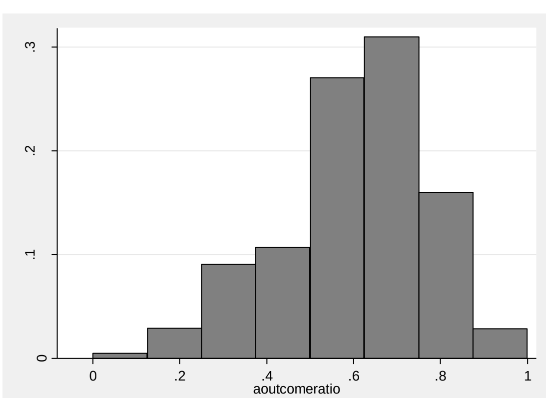
distribution is skewed to the left. Home teams are more likely to have a larger proportion of home wins than the normal distribution would assume. The home team advantage phenomenon could explain this. In the English Premier League, 47.1% of outcomes were home wins.

Fig 3. Histogram of outcomeratio



Aoutcomeratio is defined as the average of away team's most recent 19 away games. An away win corresponds to 0 on the ordinal scale, so a value closer to 0 will indicate a high quality away team. It is anticipated that a decrease in the value of aoutcomeratio will positively affect the probability of an away win and negatively affect the probability of a home win. An increase in away team strength should increase the probability of an away win.

Fig 4. Histogram of aoutcomeratio



Due to the effect that poor refereeing decisions and bad luck have on outcome, the model has developed suitable home and away team strength variables that measure average performance rather than an average outcome previously defined. For example, teams may have had many shots or shots on target throughout a game and been unfortunate in that particular game. Although they may have lost the game, it is suggested that teams that create many chances are more likely to win future games than teams that create few chances.

Avghomeshots are defined as the average number of shots by the home team in their previous 19 games. It is anticipated that an increase in avghomeshots will increase the probability of a home win and decrease the probability of an away win. Avghometarget will capture the average number of shots on target by the home team in their previous 19 games. An increase in avghometarget should increase the probability of a home win whilst decreasing the probability of an away win. Similarly, avgawayshots and avgawaytarget will signify away team strength. An increase in the level of avgawayshots and avgawaytarget should increase the probability of an away win decreasing the probability of a home win.

Fig 5 and Fig 6 show the proportion of full time home goals and full time away goals. Home teams are likely to score more goals than away teams. The percentage of home teams that fail to score is 23.42%, whereas 35.02% of away teams fail to score. Home teams score three or more goals 19.33% of the time whereas away teams score three or more goals only 10.88% of the time.

Fig 5. Full time home goals in the English Premier League

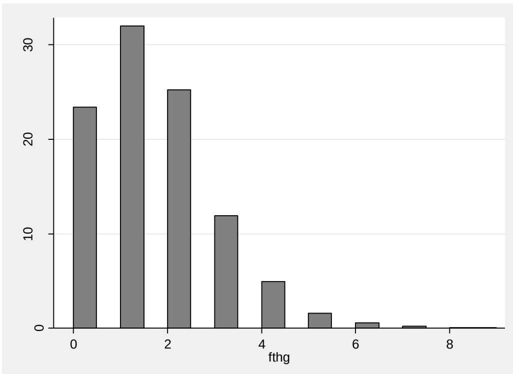
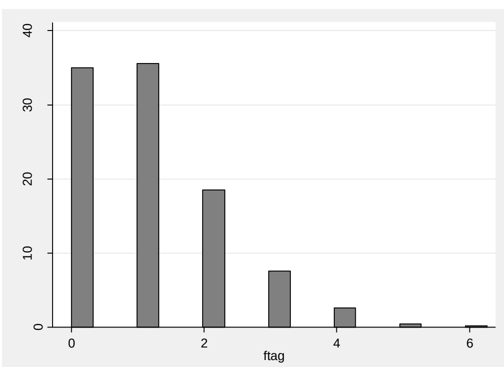


Fig 6. Full time away Goals in the English Premier League



To account for poor form against a particular opponent, the model creates recentmatch, which captures the most recent four home league games against a particular opponent. The results from previous encounters against a particular opponent are an important resource for predicting the result of the next meeting. [13] It captures winning and losing streaks against particular teams in the form of a hoodoo effect. The model expects the value of recentmatch to be positive. A poor recent run of form against a particular team, holding team strength constant would most likely continue into the future.

3.3. Betting odds

So far, an account has been provided of the available dataset used and how each variable used in the ordered probit model was created. The bookmaker odds available to each punter will now be considered. As discussed, the importance of choosing best odds cannot be overlooked. Punters that pick the best odds available to them increase the possible returns available to them while keeping risk constant. It is also possible, although unlikely, for punters to receive risk free returns from picking the best odds available from a variety of bookmakers. Eleven bookmakers are used to compile the best odds available in each country: Bet 365; Blue Square; Bet & Win; Gamebookers; Interwetten; Ladbrokes; Sporting odds; Sportingbet; Stan James; Stanleybet; Victor Chandler and William Hill. It is worth noting that, although eleven bookmakers have been used it was not possible to obtain odds from all of these at any one time. An account has been provided of the characteristics of each bookmaker used and the percentage of time that each bookmaker offered best odds available.

Table 4 analyses the average margin charged by each bookmaker across each European market. The standard deviation of each result is given in italics below. At the outset it should be noted that not only do margins vary by bookmakers, they vary across markets. For example, the average margin charged by Bet365 in the English market is 7.35%, whereas the average margin Sportingbet charged in the same market is 11.62%. Some bookmakers seem to need higher average margin levels. It is suggested that this could be a result of the size of the bookmaking firm. Larger firms may be more competitive and will effectively be able to survive on lower margins. The value of shopping around cannot be underestimated. The average margin charged by bookmakers in different markets is interesting. For example, Bet365 charged an average

[13] A complication arising from the inclusion of recentmatch outcomeratio is the omission of teams with only one season in top flight from our betting strategy. A team with no previous encounters against rival teams will be omitted from the ordered probit model.

margin of 9.37% in the French market, which is significantly higher than their average margin of 7.35% in the English market. This may be because of a lack of information of some bookmakers in certain markets. In order to hedge the uncertainty of information on bettor demand or indeed the outcome, they may charge higher margins.

Table 4. Average betting margins

		English	French	German	Italian	Scottish	Spanish
Bet 365							
	Obs	3420	3419	2752	3230	2050	3414
	Margin	7.35% *0.021*	9.37% *0.026*	8.79% *0.025*	8.55% *0.024*	9.26% *0.023*	8.62% *0.024*
Bet & Win							
	Obs	2660	2659	2142	2644	1595	2660
	Margin	9.54% *0.011*	9.67% *0.012*	9.41% *0.012*	9.59% *0.010*	11.05% *0.004*	9.61% *0.011*
Blue Square							
	Obs	4124	3968	3313	3757	2469	4127
	Margin	10.24% *0.011*	10.74% *0.010*	10.59% *0.011*	10.57% *0.011*	10.71% *0.009*	10.61% *0.011*
Gamebookers							
	Obs	4097	3945	3291	3786	2448	4086
	Margin	8.75% *0.018*	10.34% *0.015*	8.99% *0.017*	9.69% *0.017*	10.84% *0.012*	9.11% *0.017*
Interwetten							
	Obs	4163	3998	3364	3791	2475	4164
	Margin	8.75% *0.018*	12.25% *0.025*	12.39% *0.026*	12.21% *0.025*	13.69% *0.018*	12.35% *0.025*

Ladbrokes							
	Obs	4119	3681	3284	3679	2389	4112
	Margin	11.07% *0.024*	12.35% *0.004*	11.80% *0.018*	11.94% *0.016*	11.11% *0.024*	11.85% *0.016*
Sportingbet							
	Obs	373	238	292	282	225	356
	Margin	11.62% *0.003*	12.44% *0.005*	12.41% *0.004*	12.50% *0.003*	11.62% *0.002*	12.46% *0.003*
Stan James							
	Obs	2279	2277	1834	2261	1368	2278
	Margin	8.06% *0.021*	8.44% *0.018*	8.08% *0.019*	8.23% *0.019*	9.76% *0.012*	8.10% *0.020*
Sportingodds							
	Obs	760	120	588	570	451	746
	Margin	10.67% *0.015*	11.66% *0.006*	11.55% *0.003*	11.54% *0.004*	11.59% *0.003*	11.51% *0.009*
Victor Chandler							
	Obs	2280	2273	1835	2219	1366	2276
	Margin	6.88% *0.022*	9.95% *0.029*	9.67% *0.030*	9.69% *0.029*	9.65% *0.027*	9.45% *0.031*
William Hill							
	Obs	4066	3877	3254	3721	2437	3979
	Margin	10.94 *0.026*	11.62% *0.021*	11.61% *0.020*	11.58% *0.020*	11.77% *0.019*	11.59% *0.020*

Table 5 seeks to analyse which bookmakers give the best odds. The English market is analysed to provide an example. Table 5 calculates the percentage of times that each particular bookmaker offers best odds available. When calculating the best odds, there may be multiple bookmakers offering the best odds. It is important to note that, accordingly, the percentages will not sum to 100%. Due to betting odds being similarly priced, many bookmakers may provide best odds at any one time. However, it is evidenced that some bookmakers offer a larger percentage of best odds available than other rival firms. Bet365 gives the highest percentage of best odds for home outcomes. It offers the best odds 34.21% of the time. Contrastingly, Bet&Win offers the lowest percentage of best odds. They offer best odds only 13.81% of the time. For draw odds, Sportingbet offers the highest percentage of best odds. They offer best odds 67.91% of the time. Interwetten offers the lowest percentage of best odds. They offer best odds only 4.68% of the time. For away odds, Victor Chandler gives the highest percentage of best odds, offering best odds 37.15% of the time. Contrastingly, Blue Square gives the best odds only 15.07% of the time. It would appear that Bet365 and Victor Chandler offer the highest percentage of best odds in the English Premier League. Contrastingly, Bet&Win and Interwetten offer the worst value for the money offering the lowest percentage of best odds.

Table 5. Percentage of bookmaker offers of best odds (English Premier League)

		Home	Draw	Away
Bet365				
	Obs	3420	3420	3420
	% Of Best odds	34.21%	44.94%	35.91%
Bet & Win				
	Obs	2660	2660	2660
	% Of Best odds	13.81%	13.80%	15.23%
Blue Square				
	Obs	4124	4124	4124
	% Of Best odds	13.82%	19.98%	15.07%

Gamebookers				
	Obs	4097	4097	4097
	% Of Best odds	28.14%	22.82%	20.50%
Interwetten				
	Obs	4163	4163	4163
	% Of Best odds	23.56%	4.68%	17.44%
Ladbrokes				
	Obs	4119	4119	4119
	% Of Best odds	19.28%	24.67%	17.14%
Sportingbet				
	Obs	374	374	374
	% Of Best odds	23.86%	67.91%	23.26%
Stan James				
	Obs	2279	2279	2279
	% Of Best odds	23.39%	24.66%	25.14%
Sportingodds				
	Obs	760	760	760
	% Of Best odds	22.37%	45.66%	26.32%
Victor Chandler				
	Obs	2280	2280	2280
	% Of Best odds	31.10%	49.74%	37.15%
William Hill				
	Obs	4066	4066	4066
	% Of Best odds	19.90%	16.33%	15.13%

4. Results

4.1. Ordered Probit Results

Equation (12) outlines the regression used. The dependent variable is outcome. The independent variables used are Outcomeratio, Aoutcomeratio, Avghomeshots, Avghometarget, Avgawayshots, Avgawaytarget and Recentmatch. As discussed in section 2, equations (13), (14) and (15) detail the link between outcome ($y_{i,j}$), the latent variable ($y^*_{i,j}$) , the threshold (μ) and the error term ($\varepsilon_{i,j}$).

$$y^*_{i,j}=\alpha_{i,j}+\beta_{i,j}\text{Outcomeratio} + \beta_{i,j}\text{Aoutcomeratio} + \beta_{i,j}\text{Avghomeshots} + \beta_{i,j}\text{Avghometarget} + \beta_{i,j}\text{Avgawayshots}+\beta_{i,j}\text{Avgawaytarget} +\beta_{i,j}\text{Recentmatch}+\varepsilon_{i,j} \qquad (12)$$

$$\text{Home win} \quad (y_{i,j} = 1) \quad \text{if} \quad \mu_2 < y^*_{i,j}+ \varepsilon_{i,j} \qquad (13)$$

$$\text{Draw} \quad (y_{i,j} = 0.5) \quad \text{if} \quad \mu_1 < y^*_{i,j} +\varepsilon_{i,j} < \mu_2 \qquad (14)$$

$$\text{Away win} \quad (y_{i,j} = 0) \quad \text{if} \quad y^*_{i,j}+ \varepsilon_{i,j} < \mu_1 \qquad (15)$$

Table 6 reports the results of the ordered probit model for each European league. The research indicates that the predictive qualities of the independent variables are relatively similar across each market. It is found that Outcomeratio and Aoutcomeratio are significant at the 1% level in each of the European datasets. For predicting match outcomes, the result is of importance and not the number of opportunities they have in a game. Panel 1 of Table 6 gives the marginal effects of home wins evaluated at the mean value of x, where x indicates the independent variables. Panel 2 of Table 6 gives the marginal effects of draws evaluated at mean level x and Panel 3 gives the marginal effects of away wins evaluated at mean level x.

Table 6. Ordered Probit Results.

	English	French	German	Italian	Scottish	Spanish
Observations	3112	1969	1858	1991	2234	1977
1. Marginal effects of Home win evaluated at mean values of X.						
Outcomeratio	0.444*** (0.073)	0.400*** (0.098)	0.432*** (0.091)	0.482*** (0.093)	0.531*** (0.092)	0.366*** (0.103)
Aoutcomeratio	0.489*** (0.071)	0.524*** (0.097)	0.342*** (0.096)	0.589*** (0.089)	0.518*** (0.090)	0.450*** (0.094)
Avghomeshots	0.015 (0.010)	0.005 (0.007)	0.010 (0.007)	0.015** (0.006)	0.029** (0.013)	0.013* (0.008)
Avghometarget	0.024 (0.015)	0.037** (0.018)	0.011 (0.010)	0.029* (0.016)	0.001 (0.021)	0.033* (0.019)
Avgawayshots	-0.032*** (0.012)	-0.008 (0.008)	-0.021*** (0.007)	-0.010 (0.007)	-0.050*** (0.015)	-0.005 (0.009)
Avgawaytarget	-0.006 (0.019)	-0.007 (0.021)	-0.001 (0.011)	-0.007 (0.018)	0.024 (0.024)	-0.039** (0.020)
Recentmatch	0.036 (0.031)	0.018 (0.043)	0.102*** (0.039)	0.078* (0.041)	0.096** (0.043)	0.016* (0.043)
2 Marginal effects of Draw evaluated at mean values of X.						
Outcomeratio	-0.089*** (0.016)	-0.088*** (0.023)	-0.068*** (0.016)	-0.114*** (0.024)	-0.064*** (0.014)	-0.065*** (0.020)
Aoutcomeratio	-0.098*** (0.016)	-0.115*** (0.024)	-0.053*** (0.016)	-0.140*** (0.024)	-0.062*** (0.014)	-0.079*** (0.018)
Avghomeshots	-0.003 (0.002)	-0.001 (0.001)	-0.002 (0.001)	-0.004** (0.002)	-0.004** (0.002)	-0.002 (0.001)
Avghometarget	-0.005 (0.003)	-0.008** (0.004)	-0.002 (0.002)	-0.007* (0.004)	-0.000 (0.002)	-0.006* (0.003)
Avgawayshots	0.006***	0.002	0.003***	0.002	0.006***	0.001

	(0.002)	(0.002)	(0.001)	(0.002)	(0.002)	(0.002)
Avgawaytarget	0.001 (0.004)	0.002 (0.005)	0.000 (0.002)	0.002 (0.004)	-0.003 (0.003)	0.007* (0.004)
Recentmatch	-0.007 (0.006)	-0.004 (0.009)	-0.016*** (0.006)	-0.019* (0.010)	-0.012** (0.005)	-0.003 (0.008)
3 Marginal effects of Away win evaluated at mean values of X.						
Outcomeratio	-0.355*** (0.059)	-0.312*** (0.077)	-0.365*** (0.077)	-0.367*** (0.071)	-0.467*** (0.081)	-0.301*** (0.085)
Aoutcomeratio	-0.391*** (0.057)	-0.409*** (0.076)	-0.289*** (0.081)	-0.449*** (0.068)	-0.455*** (0.079)	-0.370*** (0.077)
Avghomeshots	-0.012 (0.008)	-0.004 (0.005)	-0.008 (0.006)	-0.011** (0.005)	-0.026** (0.012)	-0.011* (0.007)
Avghometarget	-0.019 (0.012)	-0.029** (0.014)	-0.009 (0.008)	-0.022* (0.013)	-0.001 (0.019)	-0.027* (0.015)
Avgawayshots	0.026*** (0.010)	0.006 (0.006)	0.018*** (0.006)	0.008 (0.006)	0.044*** (0.013)	0.004 (0.008)
Avgawaytarget	0.005 (0.015)	0.005 (0.017)	0.000 (0.009)	0.006 (0.014)	-0.021 (0.021)	0.032** (0.017)
Recentmatch	-0.029 (0.024)	-0.014 (0.034)	-0.086*** (0.033)	-0.060* (0.032)	-0.085** (0.038)	-0.013 (0.035)

Notes: Standard errors of marginal effects are shown in parenthesis. t –tests are for $H_0: \beta_{r=1}$

***=significance at 1% level; **= significance at 5% level and *= significance at 10% level.

Outcomeratio is used as an indicator of home team strength. Since the ordered probit measures home wins as 1, higher values of outcomeratio signify increased home team strength. Outcomeratio is significant at the 1% level in each of the European countries. An increase in the value of outcomeratio will increase the probability of a home win. An increase in the level of outcomeratio, holding all other variables constant, will push area out of the probability of an away win and push area into the category of home win. An increase in the level of outcomeratio at outcome=0.5 draw, will cause a decrease in the probability of draw. It is less likely that home teams will draw as outcomeratio increases. Similarly the probability of an away win decreases as outcomeratio increases. Home teams are more likely to win and less likely to draw or lose games as their outcomeratio increases. The magnitude of the effects that outcomeratio has on outcome are fairly constant across each country. The predictive quality of outcomeratio in predicting outcome is strongest in the Scottish market and weakest in the Spanish market.

Aoutcomeratio measures away team strength since the ordered probit measures away wins as 0. Lower values of aoutcomeratio signify increased away team strength. Aoutcomeratio is significant at the 1% level in each of the European countries. A decrease in the level of Aoutcomeratio at outcome (1), home win, decreases the probability of home win. As away team strength increases, there is a lower probability of a home win. A decrease in the level of Aoutcomeratio at outcome (0.5), draw, will increase the probability of a draw. Increasing away team strength increases the probability of a draw. A decrease in the level of Aoutcomeratio at outcome (0), away win, increases the probability of away win. As aoutcomeratio decreases, away team strength increases. Away teams will be less likely to lose and more likely to either draw or win the game. The predictive quality of aoutcomeratio is strongest in the Italian market and weakest in the German market.

Avghomeshots is significant at the 1% level in the Spanish La Liga. It is significant at the 5% level in the Italian Serie A and Scottish Premier League. An increase in the value of Avghomeshots increases the probability of a home win and decreases the probability of a draw and away win. A team with more previous home shots will reflect a stronger home team resulting in a greater probability of home victory. An increase in home team strength will also decrease the probability of a draw and away win.

Avghometarget is significant at the 5% level in the French Ligue 1. It is significant at the 10% level in the Italian and Spanish divisions. An increase in the number of home shots on target will increase the probability of a home win and decrease the probability of a draw and away win. Home teams producing many goal scoring opportunities in games will be more likely to win and less likely to draw or lose games at home.

Avgavgawayshots is significant at the 1% level in the English, German and Scottish divisions. An increase in average shots by the away team decreases the probability of a home win. Home teams playing against strong away teams, will be less likely to win. An increase in average shots by the away team increases the probability of a draw and increases the probability of an away win. The more away shots, the more likely the away team will either draw or win the game. Increasing away team strength decreases their likelihood of defeat away from home.

Avgavgawaytarget is significant at the 5% level in the Spanish La Liga. Increasing average shots on target by the away team decreases the probability of a home win. Similarly it increases the probability of a draw or an away win. Away teams with more shots on target reflect stronger opponents and decrease the probability of home victory accordingly. Away teams creating many goal scoring chances will be more likely to draw or win away from home.

Recentmatch is the average outcome that home teams get against a particular opponent. Increased values of recentmatch signify that the home team has had favourable past results against a particular opponent. The marginal effects of an increase in the level of x relating to recentmatch, is significant at the 1% level in the German Bundesliga 1. It is significant at the 5% level in the Scottish Premier League and significant at the 10% level in the Italian and Spanish top flights. The effect of recentmatch is found to be positive. Increasing the value of recentmatch will increase the probability of a home win. It will decrease the probability of a draw and an away win. Good recent spells against an opponent at home will continue when accounting for relative team strength. Teams with poor recent form against an opponent will be less likely to get a favourable outcome the next time they meet.

4.2. Weak-form efficiency results

This section discusses the weak-form efficiency results from the two tests outlined in section 2. This paper investigates whether betting odds reflect their true probabilities and if the model contains any information that is not already contained in bookmaker odds. Note that the betting odds tested are the best odds available to punters from 11 bookmakers. Panel 1 of Table 7 gives the results of the regression-based weak-form efficiency tests .

$$r_{i,j}=\alpha_r + \beta_r\theta^r_{i,j} + \varepsilon_{i,j}$$

for r=Home, Draw and Away. (16)

$\theta^r_{i,j}$ for r= Home, Draw and Away are the implied probabilities of bookmaker odds. A weak-form efficiency test for home odds, $\theta^H_{i,j}$, involves regressing r against the home odds, where r =1 if home team wins and 0 if the home team doesn't. A necessary weak-form efficiency condition is $\{\alpha_r = 0, \beta_r = 1\}$. The price implied probabilities equals their true values.

Evidence is found of departures from $\{\alpha_r = 0, \beta_r = 1\}$, for home wins in the English, Italian, Scottish and Spanish divisions at the 1% significant level. Evidence is also found of departures from $\{\alpha_r = 0, \beta_r = 1\}$, for draw odds in the Italian, Scottish and Spanish divisions at the 1% level and at the 5% level in the German division. Away odds are found to be inefficient in the English and Italian divisions at the 1% significance level. There are clear departures from weak-form efficiency conditions in many of the major European leagues.

Panel 2 of Table 7 gives results of the regression-based weak-form efficiency tests;

$$r_{i,j}=\alpha_r + \beta_r\theta^r_{i,j} +\gamma_r(P^r_{i,j}-\theta^r_{i,j}) + \varepsilon_{i,j}$$

for r= Home ,Draw and Away. (17)

$P^{r}_{i,j}$ for r=Home, Draw and Away is the probability of each outcome given by the ordered probit model. $P^{r}_{i,j}$-$\theta^{r}_{i,j}$ is the difference between odds compiled by the bookmaker and odds underlined by this model. A weak-form efficiency condition is that information contained in the ordered probit and is reflected in bookmaker probabilities. If betting odds are weak form efficient, no additional information should be contained in the model. H_0: $\{\alpha_r, \beta_r, \gamma_r\}$=$\{0,1,0\}$.

This paper finds that the model contains information that is not reflected in the home betting odds in the English, Italian and Scottish divisions. Also, information is contained in this model that is not reflected in the draw odds in the German, Scottish and Spanish divisions. In addition, information contained in this model is not reflected in the away odds in the English, French and Italian divisions. These findings provide evidence that there exist inefficiencies in the betting odds in the European leagues. Contrastingly, Goddard and Asimakopoulos (2004) found that at a 5% significance level there seems to be little or no evidence of systematic departure from the weak-form efficiency conditions. This could be attributed to the more recent data used in this study. Additionally, it may also be due to using many different bookmaker odds. It was assumed that punters received best odds available. Increasing the number of bookmaker odds available may decrease the margin imposed on the punter. This increases the probability of outcomes implied by bookmaker odds. Best odds probabilities will be higher than they would be if only one bookmaker's odds were tested. Section 5 seeks to exploit these inefficiencies by creating strategies capable of generating profits.

Table 7 Weak-form efficiency: regression-based tests

	English	French	German	Italian	Scottish	Spanish
Observations	4180	4030	3366	3875	2507	4180
1. TESTS BASED ON $r_{i,j}$=$\alpha_r + \beta_r\theta^{r}_{i,j} + \varepsilon_{i,j}$ $for\ r = Home, Draw\ and\ Away$						
Home wins						

Constant	-0.006 (0.021)	0.028 (0.031)	0.003 (0.028)	-0.063*** (0.022)	-0.058*** (0.022)	-0.005 (0.024)
$\theta^{H}_{i,j}$	1.059 (0.043)	0.962 (0.067)	1.023 (0.058)	1.164*** (0.046)	1.147*** (0.047)	1.057 (0.051)
F1	4.96***	0.78	1.41	7.55***	5.09***	4.71***
Draws						
Constant	-0.030 (0.047)	0.001 (0.074)	-0.029 (0.064)	-0.112*** (0.040)	-0.102** (0.045)	-0.087* (0.048)
$\theta^{D}_{i,j}$	1.099 (0.177)	1.015 (0.252)	1.032 (0.242)	1.401*** (0.139)	1.328* (0.176)	1.255 (0.178)
F1	0.32	0.29	3.74**	4.15***	4.47***	4.99***
Away Wins						
Constant	-0.037*** (0.014)	-0.027 (0.018)	0.015 (0.018)	-0.049*** (0.014)	-0.004 (0.017)	0.006 (0.015)
$\theta^{A}_{i,j}$	1.073* (0.043)	1.043 (0.067)	0.968 (0.061)	1.138*** (0.047)	1.056 (0.048)	0.968 (0.051)
F1	4.85***	2.15	0.55	6.14***	1.99	0.24
2. TESTS BASED ON $r_{i,j}=\alpha_r + \beta_r\theta^{r}_{i,j} + \gamma_r(P^{r}_{i,j}\text{-}\theta^{r}_{i,j}) + \varepsilon_{i,j}$ *for r = Home, Draw and Away*						
Home Wins						
Constant	-0.016 (0.024)	0.035 (0.049)	0.009 (0.045)	-0.048 (0.033)	-0.060*** (0.024)	-0.015 (0.037)
$\theta^{H}_{i,j}$	1.069 (0.050)	0.929 (0.105)	1.008 (0.091)	1.156*** (0.067)	1.148*** (0.050)	1.084 (0.075)
$P^{H}_{i,j}\text{-}\theta^{H}_{i,j}$	0.167* (0.103)	0.192 (0.168)	0.155 (0.158)	0.163 (0.118)	0.193* (0.108)	-0.068 (0.134)
F2	2.79**	1.18	1.07	4.19***	3.62***	0.17
Draws						

Constant	-0.028 (0.055)	0.088 (0.131)	-0.062 (0.113)	-0.132** (0.068)	-0.112** (0.049)	-0.071 (0.075)
$\theta^{D}_{i,j}$	1.102 (0.206)	0.734 (0.435)	1.194 (0.466)	1.445* (0.246)	1.381* (0.204)	1.198 (0.307)
$P^{D}_{i,j}$-$\theta^{D}_{i,j}$	0.405 (0.298)	-0.314 (0.543)	0.551 (0.611)	0.489 (0.301)	0.178 (0.361)	0.176 (0.458)
F2	0.65	0.31	2.22*	1.69	2.89**	2.13*
Away wins						
Constant	-0.048*** (0.016)	-0.060** (0.028)	0.007 (0.029)	-0.038** (0.020)	-0.009 (0.019)	-0.006 (0.024)
$\theta^{A}_{i,j}$	1.123*** (0.051)	1.211** (0.109)	0.992 (0.096)	1.081 (0.070)	1.072 (0.052)	1.028 (0.079)
$P^{A}_{i,j}$-$\theta^{A}_{i,j}$	0.230** (0.107)	0.447** (0.178)	0.169 (0.165)	-0.099 (0.124)	0.011 (0.109)	-0.016 (0.139)
F2	4.23***	2.64**	0.75	2.03*	1.35	0.10

Notes: Standard errors of estimated coefficients are shown in parenthesis.

t-tests on individual coefficients are forH_0: α_r=0; H_0: β_r=1 and H_0: γ_r=0

F1 is an F-test for H_0: { α_r , β_r } ={0,1} and F2 is an F-test for H_0: { $\alpha_r, \beta_r, \gamma_r$}={0,1,0}

***=significant at 1% level; **=significant at 5% level; *=significant at 10% level.

5. Betting Strategy

Section 4 reported the findings of the ordered probit model, which is capable of describing the data in each European league. It found that outcomeratio and aoutcomeratio possessed the best predictive qualities in analysing football outcomes. The section reported the results of the weak-form efficiency tests detailed in Section 2. It provided clear evidence of departures from weak-form efficiency. It found that betting odds did not reflect all the information available. Furthermore, the model provided additional information not reflected in bookmaker odds. Although these findings are important in highlighting departures from weak-form efficiency, the main goal of the paper was to provide a sensible betting strategy capable of generating positive returns. Panel 1 of table 8 shows the returns to betting €1 on each outcome in the different markets. If betting odds are weak-form efficient, the returns to each strategy should be constant. Interestingly, home bets perform best in each market except in the Scottish market. This suggests that home odds are overpriced relative to draw and away odds. It is likely that the burden of margin is not evenly distributed on all outcomes. Panel 2 of Table 8 confirms this suspicion by comparing bookmaker implied probabilities with historic proportions. Bookmaker implied probabilities are lower than historic proportions. Bookmakers are under-estimating the home advantage in each European market. Contrastingly, away odds are under priced in four of the six markets. Therefore, a sensible strategy would be to bet on home outcomes. As an aside, panel 2 offers an explanation of why betting on away outcomes in the Scottish division out performs other European markets. Firstly, the Scottish market sees an unusually high proportion of away outcomes observed. This may be because of a competitive imbalance or because of the league format differing from other major European leagues. Secondly, although bookmakers account for this phenomenon by altering their betting odds, they still overprice away odds relative to their historic proportion.

Table 8. Analysis of home, draw and away returns combined with bookmaker implied probabilities and actual probabilities.

1. Home, Draw and Away returns				
		Home	Draw	Away
English				
	No. Of Bets	4180	4180	4180
	Return %	-0.55%	-6.32%	-14.91%
French				
	No. Of Bets	4032	4032	4032

	Return %	-2.59%	-3.37%	-11.78%		
German						
	No. Of Bets	3366	3366	3366		
	Return %	-1.9%	-12.13%	-2.76%		
Italian						
	No. Of Bets	3884	3884	3884		
	Return %	-5.55%	-6.12%	-14.92%		
Scottish						
	No. Of Bets	2508	2508	2508		
	Return %	-7.26%	-14.68%	-4.01%		
Spanish						
	No. Of Bets	4180	4180	4180		
	Return %	-1.25%	-12.96%	-4.80%		

2. Comparing bookmaker probabilities with historic probabilities

	Bookmaker			Actual		
	$\Phi_{i,j}^{H}$	$\Phi_{i,j}^{D}$	$\Phi_{i,j}^{A}$	H(%)	D(%)	A(%)
English	0.451 *0.167*	0.264 *0.038*	0.285 *0.149*	0.471	0.261	0.268
French	0.455 *0.114*	0.292 *0.028*	0.253 *0.097*	0.465	0.297	0.238
German	0.460 *0.141*	0266 *0.030*	0.274 *0.122*	0.473	0.246	0.281
Italian	0.453 *0.162*	0.285 *0.051*	0.262 *0.137*	0.463	0.285	0.252
Scottish	0.437 *0.190*	0.252 *0.047*	0.311 *0.177*	0.443	0.232	0.325

Notes: The standard deviation of betting odds are given in italics.

Table 9 analyses the returns of a strategy of betting on favourites and long shots. A favourite is found to be the outcome with the highest price implied probability underlined by betting odds. A long shot is found to be the outcome with the lowest price implied probability underlined by betting odds. The findings are consistent with past literature and it is evidenced that favourites outperform long shots in every country. It is suggested that favourites may be overpriced due to bookmakers pricing odds according to punters' expected volume of betting. It is expected that bookmakers will offer less generous odds for long shots where they anticipate that punters possess a preference for placing small wagers on higher-priced long shots (as opposed to larger wagers on favourites). After all, bookmakers are profit maximising agents and seek to maximise expected returns.

Table 9. Favourite-long shot bias.

		Favourites	Long shots
English			
	No. Of Bets	4163	4065
	Return %	-2.88%	-11%
French			
	No. Of Bets	4004	3895
	Return %	-2.45%	-10.74%
German			
	No. Of Bets	3348	3265
	Return %	-2.32%	-5.29%
Italian			
	No. Of Bets	3859	3749
	Return %	-0.06%	-19.72%
Scottish			
	No. Of Bets	2488	2450

	Return %	-3.38%	-7.96%
Spanish			
	No. Of Bets	4160	4076
	Return %	-1.73%	-5.52%

It is clear that a strategy of betting on favourites outperforms long shots. Table 10 further analyses the returns to a strategy of betting on home, draw and away favourites. It is found that, although home favourites perform well, they are unable to generate positive returns. Surprisingly, draw and away favourites actually provide positive return in some markets. Unfortunately, due to the existence of margins, all the betting strategies discussed are unable to exploit betting inefficiencies and generate consistent returns.

Table 10. Analysis of the returns to home, draw and away favourites.

		Favourites		
		Home	Draw	Away
English				
	No. Of Bets	3104	0	1059
	Return %	-0.38%	0%	+3.72%
French				
	No. Of Bets	3399	2	603
	Return %	-2.75%	+25%	-0.82%
German				
	No. Of Bets	2600	0	748
	Return %	-2.01%	0%	-3.38%
Italian				
	No. Of Bets	2912	61	886
	Return %	-0.79%	+12.37%	+1.45%

Scottish				
	No. Of Bets	1780	0	708
	Return %	-4.27%	0%	-1.15%
Spanish				
	No. Of Bets	3314	0	846
	Return %	-1.08%	0%	-4.29%

Using information contained in the model, the research builds upon the inefficiencies present in betting odds. This paper introduces a strategy of betting on outcomes for which the probability underlined by the ordered probit is greater than the bookmaker implied probability. Betting on outcomes where $P_{i,j}^{r} > \theta_{i,j}^{r}$, where $P_{i,j}^{r}$ is the probability underlined by the ordered probit and $\theta_{i,j}^{r}$ is the probability implied by the best odds from bookmakers. This strategy aims to place bets on outcomes that are overpriced according to the model. A strategy of betting on home teams that are overpriced provides positive returns in five of the six European markets. This strategy is capable of generating positive returns as high as 6.65%. This provides clear evidence of inefficiencies in the betting odds. The Scottish market provides a negative return to this strategy. As discussed the Scottish market sees an unusually high proportion of away outcomes observed. Although bookmakers account for this phenomenon by altering their betting odds, they still overprice away odds relative to their historic proportion.

A strategy of betting on away teams that are overpriced is also capable of generating positive returns in two of the six European markets. However, it is clear that betting on overpriced home teams is a sensible strategy and one that may be possibly replicated in the future.

Table 11. Returns to betting on outcomes that are overpriced according to the model.

		Home adv>1	Draw adv>1	Away adv>1
English				
	No. Of Bets	1740	202	1170
	Return %	+2.27%	-14.33%	-15.31%

French				
	No. Of Bets	953	240	776
	Return %	+6.65%	-12.63%	+1.84%
German				
	No. Of Bets	899	0	959
	Return %	+3.71%	0%	+3.25%
Italian				
	No. Of Bets	1129	73	788
	Return %	+0.90%	+43.29%	-16.56%
Scottish				
	No. Of Bets	1131	0	1102
	Return %	-5.68%	0%	-6.29%
Spanish				
	No. Of Bets	1025	5	947
	Return %	+0.71%	-100%	-3.38%

Table 12 furthers the discussion by analysing favourites that are overpriced according to the model. It is found that a strategy of betting on away favourites performs okay, generating positive returns in three of the six markets. The problem is that the results are highly variable across each country. A strategy may only be deemed successful if it performs adequately well in each market. It is for that reason that a strategy of betting on home favourites is preferred to away favourites. It produces positive returns in four of the six divisions. More importantly the returns are much less volatile.

Table 12. Favourites that are over-priced according to the model

		Favourites with Adv>1		
		Home	Away	Total
English				
	No. Of Bets	1199	221	1420
	Return %	+2.98%	-7.85%	+1.29%
French				
	No. Of Bets	712	44	756
	Return %	-3.62%	+24.61%	-1.98%
German				
	No. Of Bets	552	112	664
	Return %	+1.71%	+7.10%	+2.62%
Italian				
	No. Of Bets	766	114	880
	Return %	+7.82%	-9.53%	+5.57%
Scottish				
	No. Of Bets	733	240	973
	Return %	-2.23%	-5.43%	-3.02%
Spanish				
	No. Of Bets	671	77	748
	Return %	+5.14%	+15.15%	+6.17%

This paper has shown that there exists betting inaccuracies in many of the European leagues. Bookmakers tend to overprice betting odds on home teams and favourites. Combining these inefficiencies with the information contained in the ordered probit model, the research found that a simple strategy of betting on home teams that are overpriced generates positive returns in five of the six European markets. The paper has developed a simple betting strategy capable of generating positive returns in the European market. Section 6 gives a brief summary and conclusion of the key findings of this paper and suggests future recommendations for future research in this interesting field.

6. Summary and Conclusions

The betting industry is a multi-million euro outlet that employs thousands of staff in Ireland. However, recently the landscape of the betting industry has changed. Bookmakers now face increased competition and as a direct result the punter may receive better odds. The broad assumptions underlying the football betting market are outlined. Football odds tend to be fixed in nature; they do not interact with bettor demand. The paper examined the interaction of punters and bookmakers. It assumed that punters could receive the best odds available to them from 11 bookmakers. This paper examined how bookmakers formulated their betting odds. It found that betting odds were inefficient. There was evidence of the favourite-long shot bias in each European league. Furthermore, home odds were similarly overpriced. The research was unable to exploit these inefficiencies due to betting margins.

A simple model was created that is capable of predicting football outcomes in six European countries. Using 11 years of data, the research found that the model contained information not already reflected in betting odds. A strategy of betting on outcomes was developed that were overpriced according to the model. A strategy of betting on home teams that were overpriced produced significant positive returns in five of the six European leagues. This would provide concrete evidence of inefficiencies in betting prices. The betting industry is extremely vulnerable at the moment. The degree of competition among bookmakers for custom is significant. The advent of online resources and the current economic climate have reduced operating profits in the industry. This increased competition has aided the punter in his attempts to exploit betting inefficiencies. This paper reinforces past evidence suggesting that the betting industry is indeed inefficient.

The betting industry is fast changing in accordance with technology. An analysis of online bookmakers would further the research. The availability of datasets composed of real bettor demand in football outcomes is needed to better understand the bookmaking industry. Invision betting markets have become available in recent times. This is when bettors have the option to gamble on a game throughout its entirety. The betting odds change as further information is received. An analysis of minute by minute bettor demand in conjunction with the alternating bookmaking odds may offer some explanations into bettor attitudes and ultimately the efficiency of the market.

7. References

Audas, R., Dobson, S.M. and Goddard, J.A. (2002) The impact of managerial change on team performance in professional sports. *Journal of Economics and Business* **54,** 633-650

Bloch, H.A. (1951) The sociology of gambling. *American Journal of Sociology* **57,** 215-221.

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

Camerer, C. (1989) Does the basketball market believe in the 'Hot Hand'? *The American Economic Review* **79,** 1257-1261.

Clarke, S.R and Norman, J.M. (1995) Home ground advantage of individual clubs in English soccer. *The Statistician* **44,** 509-521.

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

Dobson, S. and Goddard, J. (2001) *The economics of football* Cambridge: Cambridge University Press.

Dobson, S. and Goddard, J. (2003) Persistence in sequences of football match results: A Monte-Carlo analysis. *Journal of Operational Research,* **2,** 247-256.

Fingleton, J. and Waldron, P. (1996) Optimal determination of bookmakers betting odds: theory and tests. *Trinity Economic Paper Series, Technical Paper no* ***. 96/9,*** *Dublin: Trinity College, Dublin.*

Frey, J.M. (1984) Gambling: A sociological review. *Annals of the American Academy for Political and Social Sciences,* **474,** 107-121.

Gandar, J.M., Dare, W.H., Brown, C.R. and Zuber, R.A. (1999) Informed traders and price variations in the betting market for professional basketball games. *Journal of Finance,* **1,** 385-401.

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 match results and the efficiency of fixed-odds betting. *Journal of Forecasting,* **23,** 51-66.

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

Grant, A. (2008) Statistical and financial evaluation of subjective probability forecasts: empirical applications in betting markets. *PHD Thesis, Discipline of Finance, University of Sydney.*

Gray, P.K. and Gray, S.F. (1997) Testing market efficiency: Evidence from the NFL sports betting market. *Journal of Finance,* **52,** 1725-1737.

Greene, W.H. (2000) *Econometric Analysis,* 4th Ed*.,* Prentice Hall, Englewood Cliffs.

Hurley, W. and McDonough, L. (1995) Note on the hayek hypothesis and the favourite-long shot bias in pari-mutuel betting. *American Economic Review,* **85,** 949-955.

Levitt, S.D. (2004) Why are gambling markets organized so differently from financial markets? *Economic Journal,* **114(495),** 223-46.

Konning, R.H. (2000) Balance in competition in Dutch soccer. *The Statistician,* **49,** 419-431.

Kuk, A.Y.C. (1995) Modelling paired comparison data with large numbers of draws and large variability of draw percentages among players. *The Statistician,* **44,** 523-528.

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

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

Makropoulos, V. and Markellos, R.N. (2011) Optimal price setting in fixed-odds betting markets under information uncertainty. *Scottish Journal of Political Economy,* **58,** 519-536.

Milliner, I., White, P. and Webber, D.J. (2009) A statistical development of fixed odds betting rules in soccer. *Journal of Gambling Business and Economics, University of Buckingham Press,* **3,** 89-99.

Moroney, M.J.(1956) *Facts from figures,* 3rd ed. Penguin: London.

Neville, A.M., Balmer, N.J. and Williams, A.M. (2002) The influence of crowd noise and experience upon refereeing decisions in football. *Psychology of Sport and Exercise,* **3,** 261-272.

Pollard, R. and Pollard, G. (2005) Home advantage in soccer: A review of its existence and causes. *International Journal of Soccer and Science,* **3(1),** 28-38.

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

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

Rodney, J.P. and Weinbach, A.P. Sports book pricing and the behavioural biases of bettors in the NHL, *Journal of Economics and Finance, forthcoming.*

Sauer, R.D. (1998) The economics of wagering markets. *Journal of Economic Literature,* **201,** 2021-2064.

Smith, R.W and Preston, F.W. (1984) Vocabularies of motives for gambling behavior. *Sociological Perspectives,* **27(3),** 325-348.

Vergin, R.C and Sosik, J.J. (1999) No place like home: an explanation of the home field advantage in gambling strategies in NFL football. *Journal of Economics and Business,* **51,** 21-31.

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.

Woodland, L.M. and Woodland, B.M. (1994) Market efficiency and the favourite-long shot bias: the baseball betting market. *Journal of Finance,* **49,** 269-279.

Woodland, L.M. and Woodland, B.M. (2001) Market efficiency and profitable wagering in the national hockey league: can bettors score on long shots? *Southern Economic Journal,* **67,** 983-995.

8. Online Resources.

Irish Bookmaking Association - Pre-budget submission.

http://www.theirishfield.ie/site/images/articlefiles/1315578040IBA-Budget-Submission-2012.pdf [Accessed on the 5th February 2012]

Financial Times – Bookies hope for football score victory

http://www.ft.com/intl/cms/s/0/eca84c9c-62ff-11e1-b837-00144feabdc0.html#axzz20suoIq00

[Accessed the 12th March 2012]

Financial Times- Online betting surge lifts Paddy Power

http://www.ft.com/intl/cms/s/0/c502acd6-a018-11e1-90f3-00144feabdc0.html#axzz20suoIq00

[Accessed the 20th May 2012]

WWW.FOOTBALL.CO.UK

[Accessed 1st October 2011]