Beating the Odds: Arbitrage and Wining Strategies in the Football Betting Market

NIKOLAOS VLASTAKIS, GEORGE DOTSIS and RAPHAEL N. MARKELLOS*

ABSTRACT

We examine the potential for generating positive returns from wagering on football matches. To this end, arbitrage and a simple betting strategy based on a logit regression forecasting model are employed. The analysis suggests that the differences in the odds quoted by bookmakers can lead to profitable arbitrage opportunities. We show that a betting strategy based solely on the information embedded in the bookmakers' odds can yield positive expected return. This fact undermines the validity of the efficient markets hypothesis for the football betting market.

INTRODUCTION

Over the past decades economists have invested a considerable amount of effort to the study of wagering markets. The interest of researchers has spawned from the fact that betting has become a multi-billion euro industry that operates in well-organized markets. Betting markets are similar in many ways to other financial markets, like the stock market. This resemblance makes betting markets suitable for the examination of market mechanisms, as pointed out by Smith (1971). This paper examines two ways of extracting profits from the football betting market: The exploitation of arbitrage opportunities and the implementation of a betting strategy. The results of this search have important implications for the efficiency of the football betting market.

Wagering on the outcome of sports events has a long history and is probably as old as society. However, reports on organized forms of betting in England, for example, date as early as the 19^{th} century. In the 1840s, for instance, there were over 400 "list houses" that accepted bets on the outcomes of horse and greyhound races, at prices posted publicly (Jones, Clarke-Hill and Hillier, 2000). For a large part of the 20th century betting was illegal in many European countries, including the United Kingdom, where it became legal in 1961. Nowadays, betting shops are a common characteristic of the retail geography of most European countries.

The United Kingdom is by far the largest betting market in Europe, with an annual betting turnover of £2.5 billion in the year 2003. The UK betting market is dominated by a few large betting firms (bookmakers), like William Hill, Ladbrokes and Corals. The situation is similar in other European countries, whereas in some of them, like Greece, betting is a state monopoly. Most European bookmakers operate at a margin of around 12%, with the exception of state-owned monopolistic betting firms that operate at larger margins.

---

* Financial Engineering Research Center, Department of Management Science and Technology, Athens University of Economics and Business, Greece.

The emergence of the Internet and e-business has not passed unnoticed by the betting industry. The first company to launch a betting website was Sportingbet, a small company based in the Channel Islands. The example of Sportingbet was soon followed by others, including large bookmakers like William Hill. The main advantage of web-based betting is that internet bookmakers are based offshore, which allows for punters to avoid their domestic betting taxes. In response to this threat, the UK has abolished the gambling tax and made online betting legal. This move has lead to the relocation of major online bookmaking firms back to the UK and is the main reason for the uncontested leadership of the UK in the European betting industry.

In light of these new developments the market has changed. Traditional barriers of entry into the market are not applicable in web betting. Competition has become intense, a fact that pushes margins down for the benefit of punters. Furthermore, bookmakers seek out new sports on which to accept bets and new types of bets, in search for a niche in a highly competitive environment.

The intensity of competition brought about by Internet betting makes the betting market a very good candidate for exploration of profitable opportunities. We examine two different ways of extracting profits from betting markets. The first concerns risk-free arbitrage profits and the second the formulation of a betting strategy that can yield positive expected return.

The existence of arbitrage opportunities is a well known characteristic of most financial markets. Arbitrage exists when the same asset is traded at different prices in two markets at the same time. When this happens, traders whose only concern is to search and exploit such opportunities, called arbitrageurs, step up to buy the asset at the low price and sell it at the high price almost instantaneously, thus making a risk-free profit. By doing so, the prices in both markets adjust and the anomaly disappears. This is why arbitrageurs are considered to play a balancing role. In fact, arbitrage opportunities last for no more than a few seconds.

In the betting market, arbitrage may exist when two or more bookmakers set different odds for the same event. In this case, bookmakers play the role of different markets and odds play the role of prices. Differences in the odds reported by bookmakers occur very often, but this does not mean that all of these are cases of arbitrage. In reality, arbitrage opportunities in the betting market are, as in all financial markets, quite rare. Nevertheless, when an arbitrage opportunity does occur, a punter can make a risk-free profit by placing a combined bet, a bet with two or more bookmakers, on all outcomes of an event. The analysis of our data supports the existence of such opportunities and the claim that they can be exploited, as the profits that can be made are considerable. Although the existence of arbitrage opportunities in the football betting market is well known among bettors – in fact, there are several Internet sites devoted to betting arbitrage – to our knowledge, this subject has not been examined by anyone in the academic literature.

The results of this study have direct implications for the efficiency of the betting market. Fama (1970) was the first to define the concept of market efficiency for the stock market. He characterized the stock market as an information market and defined an efficient market as one in which prices fully reflect all available information. He distinguished three forms of market efficiency: Weak form, in which prices reflect the information embedded in past prices· semi-strong form, in which prices reflect the information of past prices and all other publicly available information, and strong form, in which prices reflect the information of past prices, other publicly available information and information over which certain individuals have monopolistic access.

The above definition implies that in an efficient market no individual can make abnormal returns, which means returns greater than the return of the market, without assuming greater risk. For example, if a market is semi-strong efficient, no trading strategy based on past prices and other publicly available information must be able to outperform the market. If such a strategy existed, then the market would not be semi-strong efficient. This is the basis of most tests of market efficiency. If a strategy can yield abnormal return, then the market is not efficient in the form suggested by the information incorporated in the strategy.

In the context of betting markets, efficiency demands that no bettor or bookmaker can achieve greater return than the bookmaker's margin. For bookmakers, this means that no bookmaker can operate at greater margin than the others. For punters, that no one can have expected losses less than the bookmakers' margin. We develop a strategy based on past prices (odds). We show that such a strategy not only reduces expected losses, but that it can yield positive expected return.

The paper is structured as follows: Section 1 makes a brief review of the academic literature. Section 2 discusses the methodology applied. A description of the data is found in section 3. Section 4 presents empirical results followed by a brief discussion. Section 5 concludes.

I. LITERATURE REVIEW

The literature on sports betting markets is quite extensive. This fact is not the result of chance. As Thaler and Ziemba (1988) explain, sports betting markets are better suited for the testing of market efficiency than the stock market. They claim that the main advantage of betting markets over the stock market is that the assets (perceived as the bets) in these markets have a well defined period of life, at the end of which their value becomes certain. This makes the testing of the efficiency of wagering markets much less complicated. In fact, Thaler and Ziemba suggest that the characteristics of betting markets are such, that these markets have better chances of being efficient than other financial markets.

The literature exhibits sport – specific concentration. The majority of publications concerns racetrack betting markets, which accommodate the betting on horse races (for an overview of the literature on the efficiency of racetrack betting markets see, for example, Dowie, 1976, Thaler and Ziemba, 1988, Gabriel and Marsden, 1990). Much fewer papers examine the markets for betting on sports like baseball or basketball. The American National Football League (NFL) has gained significant attention in the literature, as opposed to association football (soccer). Nevertheless, the recent literature on sports betting includes some very interesting papers on the efficiency of association football betting markets.

Pope and Peel (1989) were among the first to examine the efficiency of the football betting market. To this end, they used data from the UK fixed odds betting market (odds quoted by bookmakers) and run a series of tests in an effort to detect biases. To test for market efficiency, they developed a linear probability model for the relationship between the actual probabilities of results occurrence and the ones implied in the odds quoted by bookmakers. The model was used to devise a betting strategy based on the information embedded in the odds (weak form efficiency test). Another betting strategy based on other publicly available information (semi – strong form efficiency test), namely predictions of specialists published in the press, was implemented. They concluded that the market is efficient, as no strategy yielded positive expected after tax return, although they were able to substantially decrease the expected losses, a fact that they explained as evidence that the odds do not meet the criteria of rational expectations.

A large part of the literature concentrates on the development of statistical models to predict the outcome of football matches. Dixon and Coles (1997) developed a parametric model to predict the score of football matches. Their model uses a Poisson distribution for the number of goals scored by each team, with parameters related to past team performances. They found that a betting strategy based on their model can lead to positive expected return, which implies semi-strong inefficiency. More recently, Goddard and Asimakopoulos (2004) developed an ordered probit regression model to forecast English league football results, rather than scores. Their model incorporates information of past match results, but also a number of other explanatory variables, all publicly available. By using their model as a basis for a betting strategy, they found that positive expected return could be achieved, a fact that they maintain to constitute a violation of weak form market efficiency. Since their model does not rely solely upon information embedded in prices (odds), but incorporates other publicly available information, we must comment that their test examines semi – strong form efficiency, rather than weak form.

Cain, Law and Peel (2000) examine the existence of the favorite – longshot bias, observed in racetrack betting (see, for example, Quandt, 1986), in football betting. They analyzed data from the UK football betting market and found that there seems to be a tendency for favorites to be overpriced and longshots to be underpriced by bookmakers. They developed a model in which the goal scoring processes of the home and away teams follow a Poisson and a Negative Binomial distribution, respectively, the expected values being functions of the quoted odds. Although the existence of the favorite – longshot bias has direct implications on market efficiency, their model detected very few profitable betting opportunities.

Kuypers (2000), on the other hand, modeled the bookmakers' odds setting decision, under the assumption that the bookmakers are profit maximizers. He found evidence that inefficient odds could be set by the bookmaker as a result of his effort to maximize expected profit from bets placed by punters with biased estimations. He compared subjective probabilities implied by the odds with *ex post* estimated outcome probabilities by employing regression analysis. He found that the subjective probabilities were not significantly different from the outcome probabilities. To further test market efficiency, he developed two strategies, one based on quoted odds and the other incorporating both odds and publicly available information. He concluded that, although the market passes the weak form test, there is substantial evidence that it violates the requirements for semi – strong efficiency, as the strategy that incorporated publicly available information and selected odds yielded positive expected return.

In contrast to testing market efficiency, the examination of the existence of arbitrage opportunities in the betting market seems to have been disregarded in the literature. Nevertheless, Hausch and Ziemba (1990) explore the potential for risk-free arbitrage profits in cross-track betting on U.S. racetracks. Cross-track betting is a form of betting that allows punters to place bets with a bookmaker on one racetrack on races that take place on another racetrack. Hausch and Ziemba found that there are significant differences in prices on the same race from one track to the other, due to the fact that different tracks operate different betting pools. They developed an arbitrage model to exploit such differences and tested it on a number of Triple Crown races. Their model yielded substantial return, which proves that arbitrage opportunities in cross track betting exist and are exploitable. Moreover, Pope and Peel (1989) mention that they discovered a limited number of risk-free arbitrage opportunities in their sample of football odds and results, without further analysis or a presentation of results.

II. METHODOLOGY

Arbitrage

A bookmaker accepts bets on the outcome of sports events, at prices he announces. These prices are called odds, and they reflect the expectations of the bookmaker with respect to the outcome of the events. Bookmakers are not punters. They do not speculate on the outcome of events. They act as market makers in the betting market providing liquidity, namely holding the book. For this service they demand a fee, which is a percentage of the total value of the book. This fee is embedded in the odds.

The odds represent the return of a punter who has placed a bet on a specific outcome of an event, in the case that the actual outcome is the same with the one he placed the bet on. For example, a football match has three possible outcomes, home win, away win and draw. A bookmaker always reports odds on all outcomes of a match. When the match has taken place and the actual outcome is known, the bookmaker pays the backers of the final outcome the amount they have betted plus a profit. Most European bookmakers report odds in the euro-decimal format, in which odds are decimal numbers. Odds in this format include the returned stake of the punter, so that the total amount a winner receives is his stake multiplied by the odd on the outcome he betted on. A 2.51 odd on an outcome means that a punter who has placed a €100 bet will receive €251 back in the case that his prediction is correct. The €251 contains the €100 staked, so that the net profit of the bettor is €151.

Bookmakers employ individuals that have special knowledge of specific sports and are known as odds compilers. Their job is to calculate the probability of each possible outcome of an event. The odd on an outcome *i* is the reciprocal of the probability *P* of the occurrence of that outcome, so that:

$$Odd_i = \frac{1}{P(i)} \tag{1}$$

The bookmaker takes the probabilities calculated by odds compilers and translates them to odds, incorporating his fee, or margin. To understand how this is done, consider the tossing of a fair coin. Since the probability of the two possible outcomes is the same, 0.5, a bookmaker should report odds equal to 2 for both outcomes. This way, if a punter backed heads and another tails, both with a stake of €1, the winner would receive a total of €2, which includes his initial stake and the stake of the other punter. This leaves nothing for the bookmaker, who is without motive to hold the book in the first place. This is why a real bookmaker would report smaller odds, for example 1.90. In this case, the winner receives €1.9 and the bookmaker is left with €0.1, a margin of 5% of the total value of the book. Bookmakers call such a book an "overround" book. In this case, the probabilities that correspond to the odds and satisfy equation (1) are not actual probabilities, but *implied* probabilities.

The same rule applies to all sports, including football. The bookmaker's margin M on an event with n outcomes is easily calculated from the odds on the outcomes i of the event as:

$$M = (\sum_{i=1}^{n} \frac{1}{Odd_i}) - 1 \text{, where } i = 1,2,..,n \tag{2}$$

If we look at equation (1), we can see that $1/Oddi$ equals the implied probability of occurrence of outcome i. So, equation (2) suggests that a bookmaker prices outcomes at lower prices, as if they had larger probabilities of occurrence, which makes the sum of the probabilities of all outcomes implied by the odds larger than 1. The bookmaker's margin then is equal to the difference between 1 and the sum of implied probabilities.

From equation (2) it is obvious that if a punter was to place bets on all outcomes of an event with the same bookmaker he would realize a loss equal to the bookmaker's margin. Thus, in order to construct an arbitrage deal, more than one bookmaker is needed. The purpose of betting arbitrage is to take advantage of differences in the odds set by different bookmakers on the same event, so that the margin is reversed to the bettor's benefit, thus creating an "underround" book. To achieve that, a bettor must select the maximum odds per outcome from all available bookmakers and place a bet on each outcome with the bookmaker who offers the highest odd for that outcome. We call this a combined bet. For an arbitrage opportunity to exist, the condition that must be satisfied is:

$$M = (\sum_{i=1}^{n} \frac{1}{MOdd_i}) - 1 < 0 \text{, where:} \tag{3}$$

$MOddi = \max(Oddi)$

i=1,2,..,n

$MOddi$ is the maximum odd on outcome i reported by all available bookmakers. If equation (3) is true, then the margin of the synthesized book is negative. This means that the margin is in favor of the punter and the profit of the arbitrage deal equals the negative of this margin.

In order to exploit such an opportunity, a combined bet must be structured. The total value of the combined bet must be divided between individual stakes on every outcome proportionally to the odds of each outcome. It follows from equation (3) that the amount Bi placed on the individual stake on outcome i, as a function of the odd on that outcome and the amount W of the guaranteed income, is:

$$B_i = W * \frac{1}{MOdd_i} \tag{4}$$

To illustrate the procedure, we will use an example of an actual arbitrage deal from our database. In 29/6/2003 the national team of Romania faced the national team of Denmark. In a football match, there are three possible outcomes: home win (1), away win (2) and draw (x). The maximum odds for each outcome from all available bookmakers were the following: William Hill priced 1 at 2.37, and Sportingbet priced x at 3.2 and 2 at 5. The margin of the combined bet from equation (3) equals -0.065. If a punter wished a guaranteed payoff of €100, he would have to stake €42.2 on 1, € 31.3 on x and €20 on 2. The total amount staked in the combined bet equals €93.5. The guaranteed payoff is €100, which allows for a risk-free profit of €6.5, as implied by the margin.

There are several reasons for the existence of price discrepancies that lead to arbitrage opportunities in the betting market. The first is a difference of opinions between odds compilers. This may happen when a match is difficult to predict, for example, in a match between two teams of equal merit, or on matches in lower divisions, where the expertise of odds compilers is often hazy. Moreover, in our analysis we have made the silent assumption that bets are equally divided between the possible outcomes of an event. This is not the case in the real world and bookmakers have to predict the distribution of bets on the outcomes of events and set the odds accordingly. When a bookmaker fails to predict the distribution of bets successfully, he may find himself vulnerable to the outcome of the event. The bookmaker will then adjust the odds so as to attract bets on the outcome needed in order to balance his book, and this can lead to a profitable arbitrage opportunity.

Betting Strategy

The betting strategy we develop in this paper relies solely on the information contained in past prices (odds). The basis of our betting strategy is a model that forecasts the outcomes of matches.

We define three binary variables:

MO_1 = 1 if the result of a match is home team victory and 0 in all other cases
MO_x = 1 if the result of a match is a draw and 0 in all other cases
MO_2 = 1 if the result of a match is away team victory and 0 in all other cases

These variables are regressed in separate regressions against the odds on all outcomes. Since the dependent variable is binary, we employ logit regression, which is based upon the cumulative distribution function for the logistic distribution:

$$\begin{aligned}\Pr(y_i = 1 \mid x_i, \beta) &= 1 - (e^{-x_i'\beta} / (1 + e^{-x_i'\beta})) \\ &= e^{x_i'\beta} / (1 + e^{x_i'\beta})\end{aligned} \tag{5}$$

The estimated model is used to forecast the outcome of football matches based on the forecasted probability of occurrence of each outcome. The forecasted probabilities, denoted as $\overline{P}_{ij}$, which is the forecasted probability of occurrence of outcome *i* on match *j* are the basis of the betting rule of the strategy. The rule is quite simple: If the forecasted probability $\overline{P}_{ij}$ exceeds a threshold, T_i, then bet on outcome *i* of match *j*. The thresholds are estimated numerically so that to optimize the expected return of the strategy on in-sample data. The strategy is then tested on out-of-sample data, using the previously estimated thresholds.

The basic idea behind this strategy is the exploitation of the favorite-longshot bias, the existence of which in the football betting market has been studied by Cain, Law and Peel (2000). In a market where there is such a bias, the betting public tends to bet on favorites less than it should, according to the probability of the favorite to win, and on longshots more than it should, again with respect to the probabilities. This forces bookmakers to overprice favorites and underprice longshots, so that a bet on a favorite pays more than it should and a bet on a longshot less than it should.

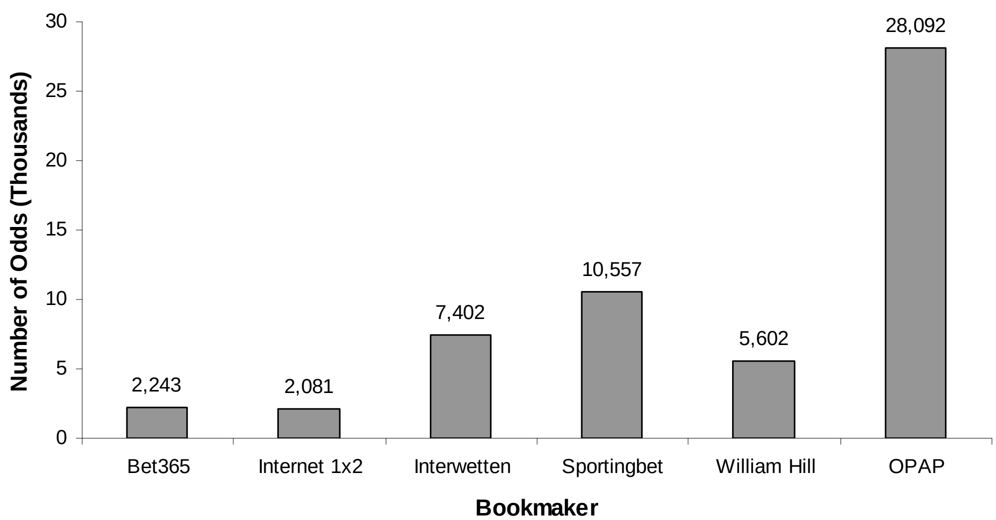
The strategy we describe above places bets on outcomes with extremely low odds, which imply very large probability of occurrence. Since the market exhibits a favorite-longshot bias, a strategy that involves placing bets on extreme favorites can lead to positive return.

III. DATA DESCRIPTION

Our data set contains odds on football matches reported in the Internet sites of 5 major online bookmakers (Bet365, Internet1x2, Interwetten, Sportingbet and William Hill), as well as data extracted from the coupons of 1 fixed-odds bookmaker (OPAP). The major difference between online bookmakers and fixed-odds bookmakers is that online bookmakers are allowed to alter their odds at any time before a match takes place, whereas the odds in fixed-odds betting are fixed for a period before the match takes place. The odds from the online bookmakers are the "closing" odds, which are the odds that were offered just before the matches started, when bookmakers stopped accepting new bets.

The data span over a 3-year period (2002-2004) and cover 26 different countries and events. The database contains a total of 12,841 football matches and 27,885 odds from the online bookmakers and odds on 28,092 matches from the fixed-odds bookmaker. Figure 1 depicts the number of odds quoted by each bookmaker. It is a rather unique database, with respect to its size and the information it contains, and offers the opportunity to compare fixed and non-fixed odds, which can lead to interesting results.

Figure 1 - Number of Odds per Bookmaker



As can be seen in Figure 1, the database does not include odds from all bookmakers for all matches. There are matches for which odds are available from just one bookmaker, as well as matches for which all bookmakers have quoted odds. For a clearer understanding of the database, Figure 2 shows the number of matches that corresponds to any number of bookmakers offering odds.

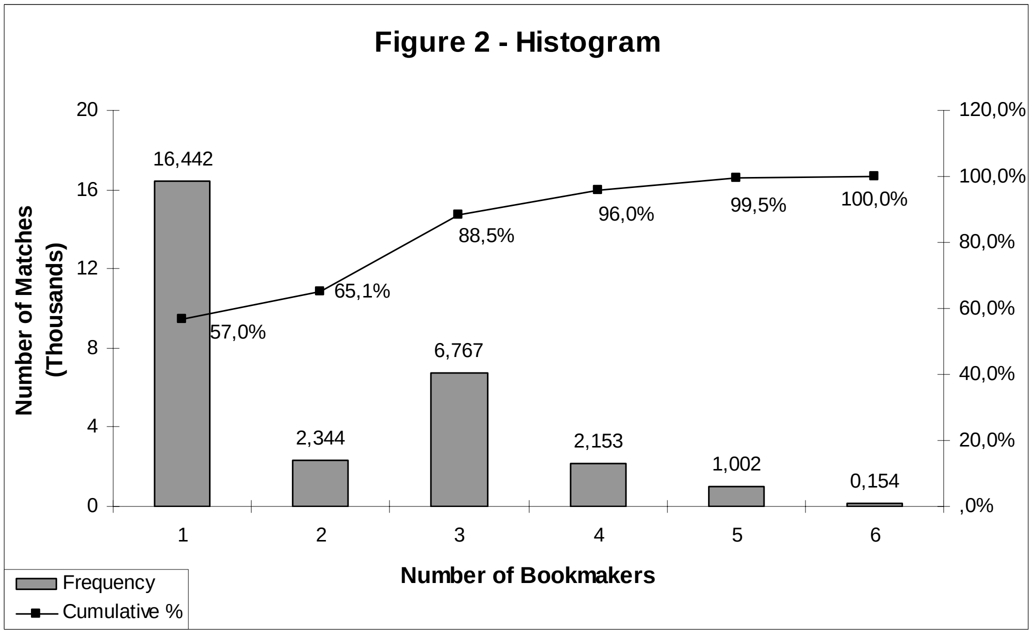


Figure 2 reveals a weakness of the database, the fact that the matches for which there are more than one bookmakers offering odds account for just the 43% of all matches, whereas the matches for which all bookmakers offer odds are only 154. This fact has significant consequences, particularly in the search for arbitrage opportunities. Nevertheless, the size of the database is such that there are more than enough data for our analysis.

In a previous section we mentioned that in an efficient market, no bookmaker should be able to operate at a greater margin than the others. It is, therefore, interesting to examine whether this condition is true in our sample. We calculated the margin of every bookmaker on all matches. Table 1 presents the descriptive statistics of the margin of all bookmakers. The results are somewhat controversial. Although 4 Bookmakers (Bet365, Internet1x2, Sportingbet and William Hill) appear to earn, more or less, the same margin, the other two (Interwetten and OPAP) seem to operate at larger margins.

Table 1 - Descriptive Statistics of Margins

	Bet365	Internet1x2	Interwetten	Sportingbet	William Hill	OPAP
Mean	0.1235	0.1236	0.1533	0.1197	0.1297	0.1685
Standard Deviation	0.0063	0.0096	0.0206	0.0081	0.0070	0.0149
Kurtosis	3.4904	2.9091	-1.1948	75.0802	26.4410	-0.1409
Skewness	0.9744	1.1553	0.3677	2.5873	2.1779	0.6729
Range	0.0518	0.0857	0.1855	0.2877	0.1890	0.1055
Min	0.1103	0.0848	0.1024	0.0734	0.0671	0.1445
Max	0.1621	0.1706	0.2879	0.3611	0.2561	0.2500

On the first hand, the fact that OPAP operates at a greater margin can be expected, since OPAP is a fixed-odds bookmaker. Offering odds that are fixed for a period of about a week before a match takes place has a direct implication on the risk the bookmaker faces, because there is much more uncertainty, as opposed to online bookmakers, who have the ability to change the odds they are offering as new information arrives in the market. The greater risk that OPAP faces is reflected in the standard deviation of the margin of OPAP, which is much greater than those of the other bookmakers. It is therefore understandable that OPAP requires a compensation for the additional risk he assumes, and this compensation is embedded in the margin.

On the other hand, there seems to be no logical explanation for the excessive return that Interwetten obtains, because Interwetten is not a fixed-odds bookmaker. Nonetheless, Interwetten seems to exhibit the same "symptom" with OPAP, namely much greater standard deviation of its margin than the other 4 bookmakers. Consequently, one can argue that the excessive return Interwetten obtains is the compensation for the excessive risk it assumes, exactly like OPAP. In this case, we must identify the source of the excessive risk. It is our opinion that the excessive risk can be attributed to one of the following reasons: Either Interwetten consciously assumes risk by allowing itself to be vulnerable to the outcome of specific events – literally placing bets on those events – or it is not as efficient in predicting the distribution of bets as the other bookmakers and is, therefore, obliged to operate at a greater margin to compensate for the additional risk.

Table 2 presents the expected return to bets on outcomes with different implied probabilities (Pi). As one might recall from equation (1), the implied probability of an outcome equals the reciprocal of the odd on that outcome. The data on Table 2 confirm the existence of the favorite-longshot bias in our sample for all bookmakers, since the expected return is significantly higher for bets on favorites (high Pi) than it is for longshots (low Pi). Moreover, the size of the return on favorites is reversely analogous to the size of the return on longshots. This fact is expected, because in a market that exhibits a favorite-longshot bias, the superior return to bets on favorites is financed by the inferior return to longshots. This behavior is best observed in the case of Interwetten, where betting on favorites has positive expected return, which seems to be financed by the bets on longshots that have the lowest expected return in the sample.

Table 2 - Average Return for Different Classes of Implied Probability

Range of Pi	**Bet365**			**Internet1x2**		
	Number of Observations	Average Return	Standard Deviation	Number of Observations	Average Return	Standard Deviation
0 < Pi ≤ 0.2	532	-28.85%	2.01	551	-28.06%	2.09
0.2 < Pi ≤ 0.4	4,020	-15.28%	1.43	3,696	-14.50%	1.42
0.4 < Pi ≤ 0.6	1,580	-4.30%	1.04	1,428	-7.50%	1.03
0.6 < Pi ≤ 0.8	541	-5.74%	0.72	513	-5.11%	0.70
0.8 < Pi ≤ 1	56	-7.55%	0.46	55	-5.18%	0.42
	Interwetten			**Sportingbet**		
	Number of Observations	Average Return	Standard Deviation	Number of Observations	Average Return	Standard Deviation
0 < Pi ≤ 0.2	1,446	-36.27%	2.03	3,026	-33.00%	2.11
0.2 < Pi ≤ 0.4	12,935	-19.66%	1.38	18,570	-13.50%	1.43
0.4 < Pi ≤ 0.6	5,763	-8.57%	1.03	7,195	-7.36%	1.02
0.6 < Pi ≤ 0.8	1,781	-5.69%	0.71	2,518	-4.40%	0.71
0.8 < Pi ≤ 1	281	1.03%	0.37	362	-2.00%	0.41
	William Hill			**OPAP**		
	Number of Observations	Average Return	Standard Deviation	Number of Observations	Average Return	Standard Deviation
0< Pi ≤ 0.2	1,416	-30.74%	2.00	6,656	-25.74%	2.13
0.2 < Pi ≤ 0.4	9,939	-15.27%	1.42	48,160	-18.37%	1.38
0.4 < Pi ≤ 0.6	3,853	-6.71%	1.03	20,425	-11.42%	1.02
0.6 < Pi ≤ 0.8	1,435	-6.17%	0.72	7,855	-9.54%	0.72
0.8 < Pi ≤ 1	163	-4.52%	0.43	1,180	-5.63%	0.44

IV. EMPIRICAL ANALYSIS

Arbitrage

As mentioned before, our sample contains odds on football matches quoted by five online bookmakers and one fixed-odds bookmaker. Because online bookmakers operate differently from fixed-odds bookmakers, the search for arbitrage opportunities was performed in two different samples: The first contains only the odds from the online bookmakers, whereas the second incorporates the odds from the fixed-odds bookmaker as well. This is arranged so that the existence of arbitrage opportunities in the online betting market can be verified and the extent of price discrepancies between fixed-odds and online bookmakers observed.

Table 3 summarizes the results of arbitrage trading for both samples. The second sample, which incorporates the odds from all bookmakers, yielded a significantly larger number of arbitrage opportunities than the first sample. Although the first sample includes less than half the matches of the second sample, the difference in the number of arbitrage opportunities between the two samples cannot be attributed to this reason. For an arbitrage opportunity to exist there must be at least two bookmakers offering odds for the same match. Therefore, it is not the difference in the total number of matches but rather the difference in the number of matches with odds from more than one bookmaker that has a significant impact to the number of arbitrage opportunities observed. As Table 3 reveals, the difference in the number of matches with more than one bookmaker does not account for the difference in the number of arbitrage opportunities discovered. The two samples perform differently and this is depicted in the percentage of arbitrage deals over the number of matches with more than one bookmaker.

Table 3 - Arbitrage Trading Results

	1st Sample	2nd Sample
Number of Arbitrage deals	10	63
Total number of Matches	12,841	28,862
Number of Matches with more than one bookmaker	10,374	12,420
Percentage of Arbitrage deals over all matches	0.078%	0.218%
Percentage of Arbitrage deals over matches with more than one bookmaker	0.096%	0.507%
Descriptive Statistics of Arbitrage Return		
Mean	0.4072	0.2178
Standard Error	0.2008	0.0379
Median	0.1601	0.1429
Standard Deviation	0.6349	0.3006
Sample Variance	0.4030	0.0904
Kurtosis	6.5102	24.1464
Skewness	2.5064	4.2247
Range	2.0777	2.0858
Minimum	0.0101	0.0020
Maximum	2.0878	2.0878

The poor performance of the first sample can be attributed to the characteristics of the sample. The sample includes odds from five online bookmakers taken at the closing of the betting period – just a few minutes before the matches started. Since online bookmakers have the privilege to alter their offered odds at any time before the closing of the betting period, they can eliminate any price discrepancies between them. Such differences in prices are not rare between bookmakers early in the betting period and reflect differences in opinions regarding the outcomes of matches and, more importantly, the distribution of bets. As the time passes, the distribution of bets becomes more and more certain and new information about the matches themselves – for example, the availability of key players, weather e.t.c. – reaches the market. The flow of information decreases uncertainty and alters the expectations of bookmakers, which leads to a convergence of the odds from all bookmakers at the end of the betting period. Nevertheless, it is possible to observe price discrepancies between bookmakers even at closing prices. When a bookmaker finds the distribution of bets on his book overweighed towards a specific outcome near the end of the betting period, he faces increased risk. This is not desirable for the bookmaker who will, in an effort to attract bets towards the desirable outcomes so that he can balance his book, offer very attractive prices for those outcomes. This is the main explanation for the arbitrage opportunities that were found in the first sample.

The performance of the second sample is expected for similar reasons. Fixed-odds bookmakers cannot alter their odds for a specific time period before the match takes place. This means that they cannot incorporate in their odds any new information that becomes available. Since the second sample includes both online and fixed-odds bookmakers it is expected that a significantly larger number of price discrepancies – and, consequently, arbitrage opportunities – would be detected in this sample. In fact, the number of arbitrage opportunities in the second sample would be much greater, if the overlapping between the matches for which the online bookmakers offer odds and the matches for which the fixed-odds bookmaker offers odds was greater – it is no more than roughly 2,000 matches in our sample.

The expected return of arbitrage trading is significantly high for both samples and suggests that the arbitrage opportunities discovered can be exploited efficiently. In order to support this claim, Figures 3 and 4 depict the distribution of arbitrage trading return for the first and second samples, respectively. Again, it is obvious that the second sample provides a more suitable terrain for arbitrage trading than the first sample. Nevertheless, both return distributions confirm the exploitability of the arbitrage opportunities discovered, as the return yielded by arbitrage trading is considerable in most occasions. It is notable that, on both samples, 60% of the arbitrage trades returned more than 12.5%. Naturally, there are costs involved in arbitrage trading, such as commissions for the transferring of money to bookmakers and taxes. An examination of the betting terms of online bookmakers revealed that these costs can be set aside, since betting profits are tax-free and there is the ability to maintain an account with the bookmakers in order to avoid the charging of commissions by banks. Although there are no commissions involved in fixed-odds betting – a punter can simply walk to the nearest betting shop to place his bet in cash – the betting profits are not tax-free in this case. Despite this fact, the size of the return of arbitrage trading is such, that even if the associated costs cannot be avoided, most arbitrage opportunities can be profitably exploited.

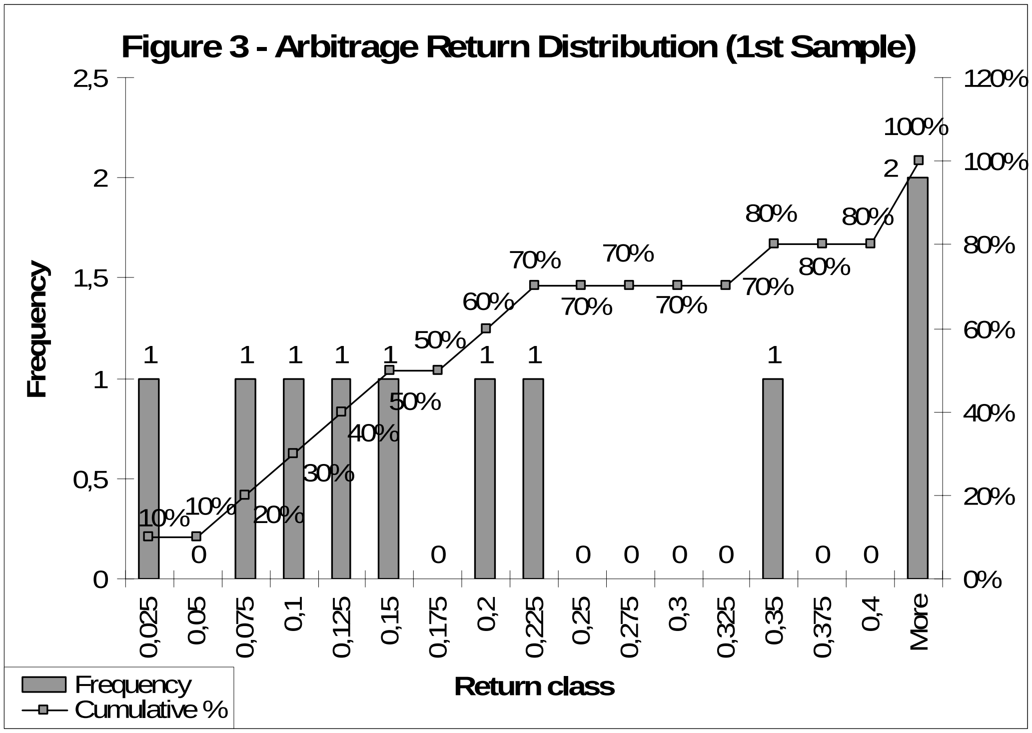


Figure 3 - Arbitrage Return Distribution (1st Sample)
Frequency
Return class
Frequency
Cumulative %
0,025
0,05
0,075
0,1
0,125
0,15
0,175
0,2
0,225
0,25
0,275
0,3
0,325
0,35
0,375
0,4
More

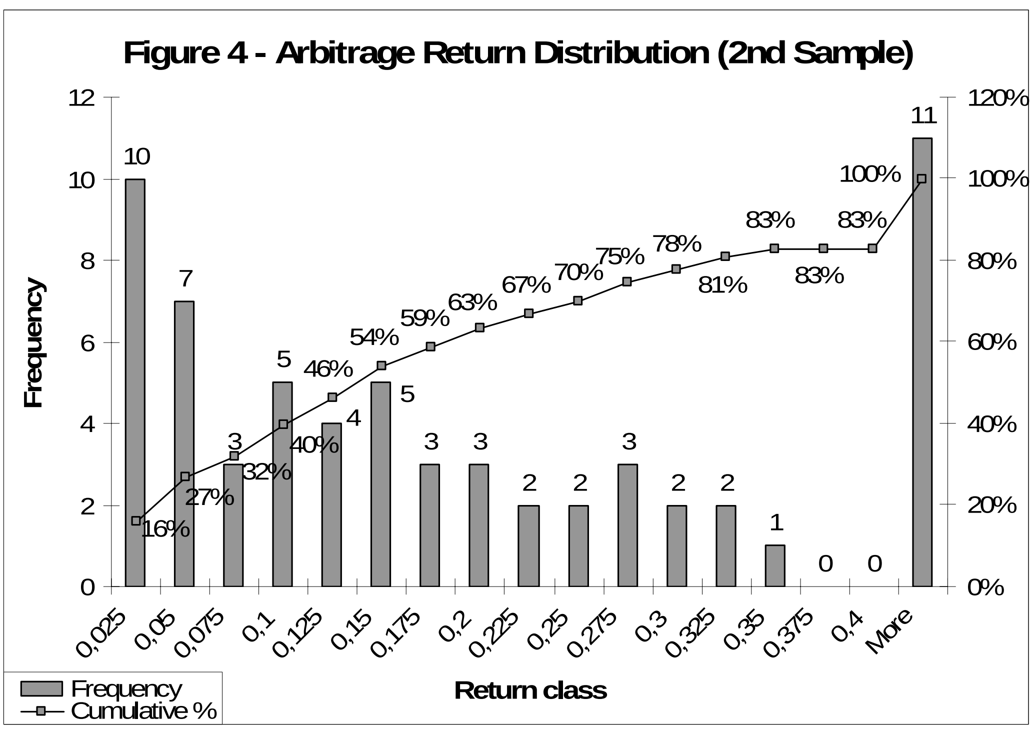


Figure 4 - Arbitrage Return Distribution (2nd Sample)
Frequency
Return class
Frequency
Cumulative %
0,025
0,05
0,075
0,1
0,125
0,15
0,175
0,2
0,225
0,25
0,275
0,3
0,325
0,35
0,375
0,4
More

Betting Strategy

A betting strategy is simply a rule based upon which a punter places bets on the outcomes of specific matches. Most betting strategies are based on some kind of forecast of the match outcomes. In our case, we employed a logit regression model estimated using past odds data from 6 different bookmakers to forecast the results of the matches in the sample. In the previous section we analyzed the differences between online and fixed-odds bookmakers that dictated the use of two different samples. In this section we use again two samples in order to explore potential differences in the predictive power of the odds quoted by online bookmakers at the end of the betting period with those quoted by a fixed-odds bookmaker. This led to the estimation of two different models: the first incorporates the data from the online bookmakers, whereas the second is comprised of the data from the fixed-odds bookmaker only. There is no need to include the data from any of the online bookmakers in the second sample, as the betting strategy has no requirements of more than one bookmaker quoting odds in order to operate successfully, as opposed to arbitrage trading.

In order to establish the robustness of our model, both samples were divided to subsamples. Roughly 77% of the data was used for the estimation of the logit model and the betting thresholds, whereas the remaining 33% of the data was set aside for out-of-sample evaluation purposes.

The regressions were estimated using only past odds data as the explanatory variables and a binary outcome variable as the dependent variable. Because the first sample includes odds from 5 different bookmakers, we used the mean of the odds from all bookmakers for each outcome in the estimation of the model, in order to consolidate the information embedded in the odds of all bookmakers to a single number. Each model contains three different regressions, one for each binary outcome variable. Any variables that were not found significant at the 10% level were omitted.

Table 4 reports the results of the models estimation. There are two tables included, one for the model estimated using online bookmakers data and the other for the model estimated using the fixed-odds bookmaker data. Each column represents a different regression, one for each binary outcome variable. As expected, the coefficient of the odd on the outcome that is used as dependent variable is negative on all regressions, depicting the negative relationship between the odd on an outcome with its probability of occurrence. It is notable that the variables that account for the variability of the dependent variable on each regression are almost the same on both samples. This confirms the fact that bookmakers use similar criteria in the odds-setting procedure. For example, in the case of draw occurrence, the only significant variable on both models is the odd for draw. This fact can be attributed to the odds-setting mechanism of bookmakers. Because draws are difficult to predict, bookmakers usually calculate the odds for home and away wins and then set the odd for draw so that the book is overround. As a result, the odds on draws usually contain only the information embedded in the other odds. The models reject the other variables simply because they offer no additional information. This fact is confirmed when the draw regression is estimated using the other two variables and omitting the odd for draw variable. In this case, the other two variables are significant at 1% level, but the best model is still the one that incorporates only the odd for draw variable.

Table 4 – Logit Estimation Results

	Online Bookmakers Data*		
Variable	Home team victory (MO_1)	Draw (MO_x)	Away team victory (MO_2)
Constant		0.7972	-1.4410
		(0.1783)	(0.2585)
Average odd for home team victory	-0.4353		0.2398
	(0.0168)		(0.0517)
Average odd for draw		-0.5477	0.3922
		(0.0544)	(0.0946)
Average odd for away team victory	0.2029		-0.3940
	(0.0089)		(0.0353)
	Fixed-odds Bookmaker Data**		
	Home team victory (MO_1)	Draw (MO_x)	Away team victory (MO_2)
Constant		0.6762	-1.4699
		(0.1127)	(0.1613)
Odd for home team victory	-0.5877		0.2920
	(0.0350)		(0.0342)
Odd for draw	0.1996	-0.5324	0.3126
	(0.0410)	(0.0357)	(0.0609)
Odd for away team victory	0.1138		-0.3327
	(0.0174)		(0.0231)

Notes: All coefficients are significant at 1% level.

Standard errors of estimated coefficients are shown in parentheses.

* number of observations = 10,000 ; ** number of observations = 21,600

The estimated models were used to forecast the probability of occurrence of each outcome for all matches in the sample. The in-sample proportion of the estimated probabilities was used for the estimation of the betting thresholds. Table 5 presents the estimated thresholds for both models. It is obvious that the thresholds are very high, which means that the strategy encourages the betting on extreme favorites. This result is to be expected in a market that exhibits a favorite-longshot bias. Although the threshold for the betting on draws, T_x , seems to be relatively low, this is not true. Draw is usually the least probable outcome and this is naturally reflected in the odds. For this reason, the forecasted probabilities for draws are lower than those for the other outcomes. In fact, the estimated thresholds are very near to the maximum values of the forecasted probabilities for draws, a fact that indicates that in the cases that we placed bets on draws, the draw was the ‘favorite’.

Table 5 – Estimated Strategy Thresholds

Model	T_1	T_x	T_2
Online Bookmakers	0.82	0.39	0.97
Fixed-odds Bookmaker	0.96	0.41	0.99

Once the thresholds are estimated, the strategy can be evaluated on out-of-sample data using the same threshold values. Tables 6 and 7 summarize the results of the betting strategy for both samples on in-sample and out-of-sample data, respectively. For the calculation of the strategy return for the online bookmakers sample we used the maximum odd available in each match, so as not to use the average of all odds, which is a non-existent number. The Tables indicate that the strategy is successful, as it was able to yield significant expected return both on in-sample and out-of-sample data for both samples. Moreover, the risk-adjusted return (Sharpe Ratio) on both samples is also significantly high, as the standard deviation of return is, in most cases, quite low, which is expected of a strategy that places bets on favorites.

Table 6 – Strategy Results on In-sample Data

	Online Bookmakers			
	Home Team Victory	Draw	Away Team Victory	Total
Number of Bets	238	1	14	253
Expected Return	0.0039	1.1000	0.0721	0.0120
Standard Deviation of Return	0.3714	N/A	0.0372	0.3671
Sharpe Ratio	0.0105	N/A	1.9369	0.0327

	Fixed-odds Bookmaker			
	Home Team Victory	Draw	Away Team Victory	Total
Number of Bets	16	13	11	40
Expected Return	0.0200	0.3500	0.0418	0.1333
Standard Deviation of Return	0	0.7754	0.0140	0.4564
Sharpe Ratio	N/A	0.4514	2.9842	0.2919

Table 7 – Strategy Results on Out-of-sample Data

	Online Bookmakers			
	Home Team Victory	Draw	Away Team Victory	Total
Number of Bets	53	2	3	58
Expected Return	0.0081	0.1000	0.1267	0.0160
Standard Deviation of Return	0.3988	1.5556	0.0231	0.4379
Sharpe Ratio	0.0203	0.0643	5.4848	0.0365

	Fixed-odds Bookmaker			
	Home Team Victory	Draw	Away Team Victory	Total
Number of Bets	9	3	4	16
Expected Return	0.0233	0.1833	0.0275	0.0544
Standard Deviation of Return	0.0100	1.0324	0.0150	0.3825
Sharpe Ratio	2.3333	0.1776	1.8333	0.1422

There seems to be a difference in the number of bets placed between in-sample and out-of-sample data. This is to be expected, as the predictive power of the model is much greater for the data on which it was estimated. Furthermore, there is a difference in the number of bets placed between the online bookmakers sample and the fixed-odds bookmaker sample both on in-sample and out-of-sample data. The reasons for the better performance of the online bookmakers sample over the fixed-odds bookmaker sample lie in the information contained in each sample. We mentioned in the previous section that the main advantage of online bookmakers is their ability to incorporate new information in their odds as it reaches the market. The most important information is the distribution of bets, because it reflects the expectations of the betting public. These expectations are forged through every bettor's experience and all the available information. For this reason, they are a valuable source of information, despite the fact that they may be biased. Fixed-odds bookmakers do not have this advantage, therefore their odds are a poorer source of information.

Moreover, the first sample contains the consolidated information from 5 different bookmakers, whereas the second only contains the information from one bookmaker. This is another advantage for the first sample, because, even in cases where one bookmaker fails to set the odds successfully, the correct odds of the other four will compensate so that the mean is very close to the right value.

The difference in the information contained in the two samples is depicted in Figures 5 and 6. Each figure presents a 3-dimentional scatter diagram of the odds for all matches in each sample (Figure 5 for online bookmakers and Figure 6 for the fixed-odds bookmaker). Every point in the 3-dimentional space represents a match. The position of each point depends on the value of the odds for the three possible outcomes, whereas the color represents the actual outcome (red for home team victory, blue for draw and yellow for away team victory).

In the edges of the accumulation of points in both figures one color, different for each edge, seems to be dominant. This indicates that extreme favorites are usually winners. For instance, in Figure 5, the matches that have extremely low odds for home team victory (which is accompanied by extremely high odds for away team victory and draw), a fact that suggests that the home team is an extreme favorite, are dominantly in red color, which indicates that the actual outcome was a home team victory.

A comparison of the two figures reveals the difference in the quality of the information contained in the two samples. The data from the fixed-odds bookmaker are much noisier than the data from the online bookmakers. This fact reflects the increased uncertainty that exists at the time the fixed-odds bookmaker sets his odds, compared to the end of the betting period, at which time the online bookmakers odds were collected. Furthermore, the domination of one color on each edge of the accumulation of points on Figure 6 is not as absolute, as it is for Figure 5, a fact that suggests that the fixed-odds bookmaker data are not as good a predictor of match results, as are the online bookmakers data.

Nonetheless, the strategy was able to yield substantial positive expected return for both samples on out-of-sample data. This fact has important implications for the efficiency of the football betting market. Since the strategy incorporated only the information embedded in past prices, the results suggest weak-form inefficiency for the market, although the number of profitable opportunities detected is relatively low.

Figure 5 - 3D Scatter Diagram of Average Odds and Outcomes (Online Bookmakers)

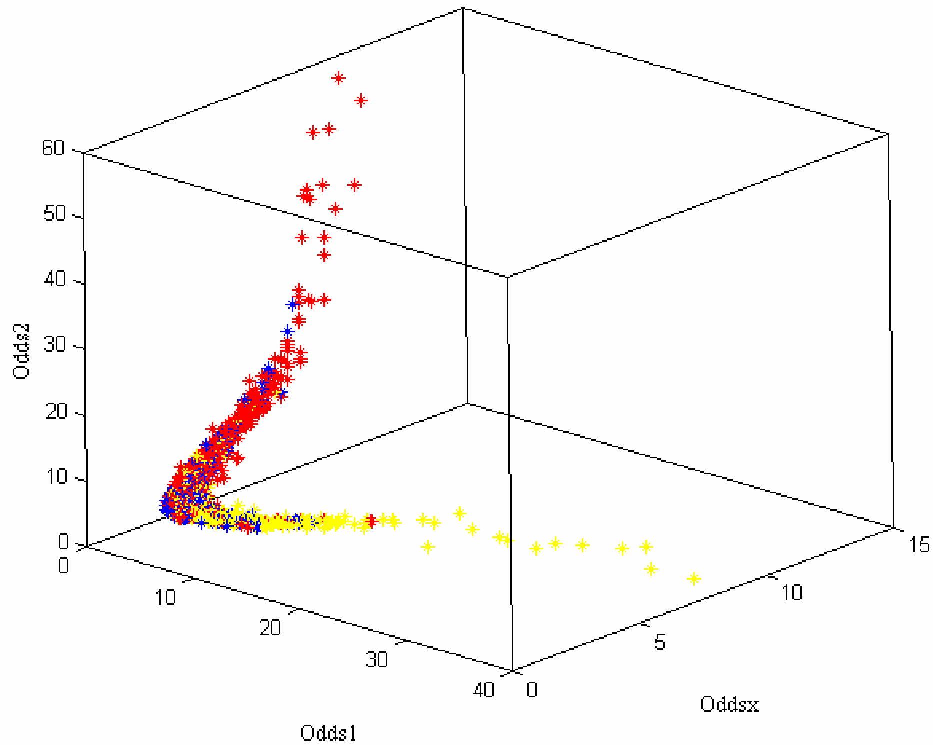
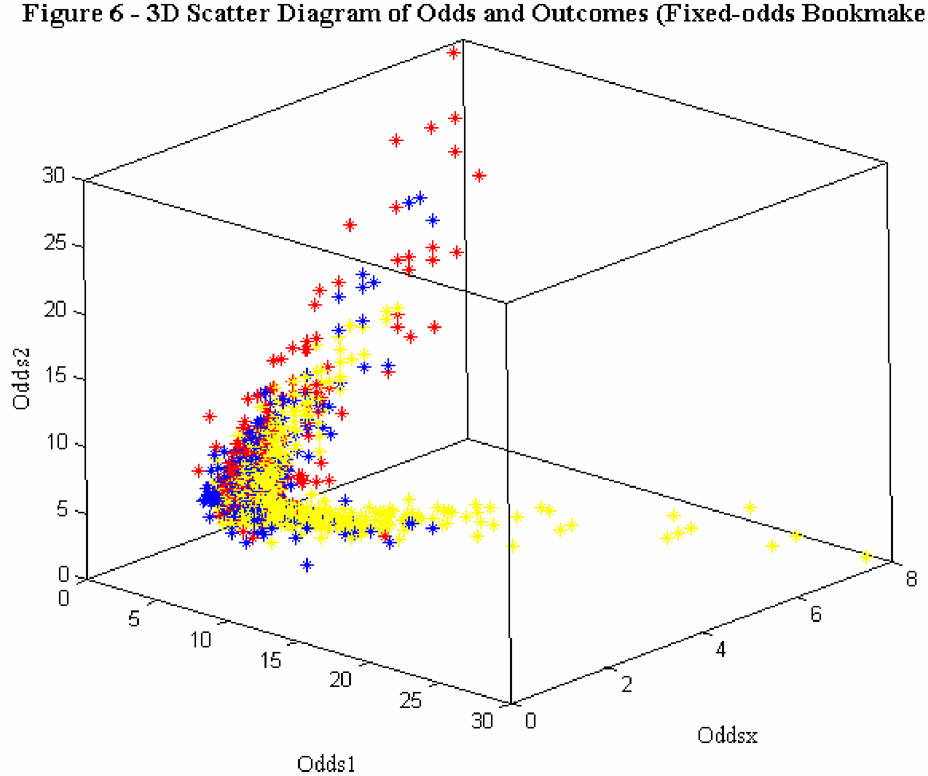


Figure 6 - 3D Scatter Diagram of Odds and Outcomes (Fixed-odds Bookmaker)



V. CONCLUSIONS

The purpose of this paper is to explore the existence of profitable opportunities in the football betting market. We used arbitrage trading and a betting strategy based on a logit regression model of past odds in an effort to generate positive returns. The methodology developed was implemented on a sample of past football matches results and odds on the outcomes of those matches from six different bookmakers. We found a significant number of exploitable arbitrage opportunities originated in the price discrepancies between bookmakers. The betting strategy yielded substantial positive expected return on out-of-sample data. The results presented in this paper pose serious doubts on the validity of the efficient markets hypothesis for the football betting market.

REFERENCES

Cain, M., Law, D. & Peel, D. (2000) "The Favourite – Longshot Bias and Market Efficiency in UK Football Betting", *Scottish Journal of Political Economy,* vol. 47, no. 1, pp. 25-36.

Dixon, M.J. & Coles, S.G. (1997) "Modelling Association Football Scores and Inefficiencies in the Football Betting Market", *Applied Statistics,* vol. 46, no.2, pp. 265-280.

Dowie, J. (1976) "On the Efficiency and Equity of Betting Markets", *Economica,* New Series, vol. 43, no. 170, pp. 139-150.

Fama, E.F. (1970) "Efficient Capital Markets: a Review of Theory and Empirical Work", *Journal of Finance,* vol. 25, no. 2, pp. 383-487

Gabriel, P.E. & Marsden, J.R. (1990) "An Examination of Market Efficiency in British Racetrack Betting", *The Journal of Political Economy,* vol. 48, no. 4, pp. 874-885.

Goddard, J. & Asimakopoulos, I. (2004) "Forecasting Football Results and the Efficiency of Fixed-odds Betting", *Journal of Forecasting,* vol.23, pp. 51-66.

Hausch, D.B. & Ziemba, W.T. (1990) "Arbitrage Strategies for Cross-Track Betting on Major Horse Races", *The Journal of Business*, vol. 63, no. 1, part 1, pp. 61-78.

Jones, P., Clarke-Hill, C.M.& Hillier, D. (2000) "Viewpoint: Back Street to Side Street to High Street to e-street: Sporting Betting on the Internet", *International Journal of Retail and Distribution Management*, vol. 28, no. 6, pp. 222-227.

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

Pope, P.F. & Peel, D.A. (1989) "Information, Prices and Efficiency in a Fixed Odds Betting Market" *Economica,* New Series, Vol. 56, No. 223, pp. 323-341.

Quandt, R.E. (1986) "Betting and Equilibrium", *The Quarterly Journal of Economics,* Vol. 101, No.1, pp. 201-208.

Smith, V.L. (1971) "Economic Theory of Wager Markets", *Western Economic Journal,*vol. 9, p. 242.

Thaler, R.H. & Ziemba, W.T. (1988) "Anomalies: Pari-mutuel Markets: Racetracks and Lotteries", *The Journal of Economic Perspectives,* Vol. 2, No. 2, pp. 161-174.