New Entry and Strategic Group Emergence in the soccer betting market: pricing behaviours, group interaction and efficiency implications

1. Introduction

The rapid expansion of online betting in the last twenty years has resulted in a transformation in the financial scale of global soccer betting markets, with billions of dollars traded on a weekly basis (Forrest, 2006, 2012) between a heterogeneous population of bettors and bookmakers. At the same time, the market has been characterised by transformative structural change, fuelled by the significant incursion of new entrants with distinctive behaviours in terms of operating principles and practices. A recent paper (Grant, et al. 2019) charts this change and suggests that traditional bookmakers, referred to as 'position taking bookmakers (PTBs)' are now being challenged by a new strategic group of 'book balancing bookmakers (BBBs)'. In addition, market expansion has coincided with prices offered by bookmakers becoming increasingly competitive (Forrest, 2012), which has attracted sophisticated, price sensitive bettors who seek to take advantage of new 'investment' opportunities. Due to the ease, and relatively low-cost of shifting capital between online bookmakers, these bettors are likely to seek out arbitrage opportunities. In fact, Grant et al (2019) found that the different product portfolios and operating behaviours of the PTBs and BBBs have led to the creation of market inefficiencies, with theoretically highly profitable opportunities for arbitrage. At the same time, however, Grant et al. (2019) present evidence that the PTBs operate a practice of deterring or refusing bets from informed bettors, preventing arbitrage opportunities from being exploited; leading to the concern that these important expanding markets may remain inefficient.

Given the above context, a natural area of enquiry within the industrial organisation tradition is how structural upheaval, in the form of disruptive new entry, affects conduct, in terms of the emergence of behaviourally-distinct strategic groups, pricing and group interaction. This, in turn, raises the issue of how pricing behaviours and outcomes impact on overall market performance in terms of efficiency, where a key question is whether, despite barriers to trade, markets have the capacity to accommodate information embodied in the decisions of sophisticated investors, thereby driving prices towards efficiency.

BBBs act like market makers in other forms of financial market, effectively matching buyers with sellers by adjusting their odds according to the amounts traded on different game outcomes. As a result, their profit is a function of generated turnover and their prices reflect a volume-weighted average of the public's opinion, potentially dominated by 'smart-money'. PTBs on the other hand, attempt to maximize their profit margin from a large customer base, deliberately filtered to avoid 'informed bettors' (Grant et al. 2019).

We contend that informed bettors are more likely to bet with BBBs, ensuring that their final odds represent more efficient predictors of match outcomes than the closing odds of PTBs. However, whilst PTBs operate policies that deter informed bettors, they may realise that these bettors hold information that can inform the setting of appropriate prices. They, therefore, may operate a low risk, low cost policy of capturing this information; moving their own prices in the direction of BBB's prices. This strategy would enable them to capitalise on the informed bettors' information, without risk of their profits suffering from large, informed bets. Consequently, the overall efficiency of the European soccer betting market may be less compromised than the restrictive practices of the BBBs might imply.

It is evident that the, largely online, BBBs are willing (and explicitly so) to accept large stakes from informed traders and typically employ a relatively low overhead, low-margin, high-turnover strategy. To make this a financially viable strategy, they will actively manage their book to ensure that liabilities across outcomes for an event are relatively equalized ('book-balancing'). This will ensure that their profits on a given event will roughly equal the transaction costs incorporated into their odds, so that profits are assured regardless of the event outcome.

Traditional, 'position taking', bookmakers, generally have long-standing reputations and generally cater for recreational (less informed) bettors. These bookmakers include those that provide physical-world betting services (i.e., betting shops). They actively encourage new accounts with small account opening bonuses and advertise low-probability (and low-liquidity) bets. The greater physical-world presence (and associated overheads) means that these bookmakers cannot replicate the low transaction costs (over-rounds) of their exclusively online competitors, and operate relatively low turnover, high-margin strategies.

With a target market of unsophisticated clientele, and relatively high margins, the PTBs have less incentive than BBBs to change their odds frequently and they have less incentive to maintain a balanced book. Based on the belief that they generally have superior information to their customers, they set odds to deliver a high margin, and to maximize profits over time. We argue that even PTBs will change prices in response to public information (when they believe this represents new information) to avoid excessive imbalance in their book. However, we contend that their high over-rounds permit them to change their odds less frequently than those of the BBBs.

To explore these issues of market behaviour and performance in the light of structural change, we compare the evolution and efficiency of prices in European football betting markets of a major online BBB with those of a major UK-based PTB for all matches played in six major European leagues for a whole season. We find that the odds of the BBB change far more often during the active market than those of the PTB. In addition, we show that lagged odds changes at the BBB are significant predictors of the odds changes at the PTB, but not vice-versa. Equally, we show that the closing odds of the PTB are a function of the day-ahead odds of the BBB, but the day-ahead odds of the PTB do not significantly affect the BBB's closing odds. These findings are consistent with the trades of informed bettors moving prices at the BBB and this information then diffusing to the PTB. In addition, we find that the odds of the BBB (cf. the PTB) better forecast match outcome and that closing prices are more efficient predictors of match outcomes than those observed 24 hours prior to kick-off, for both types of bookmaker.

Taken together, the results confirm that informed bettors trade with the BBBs, and that their bets reveal new information about game outcomes over time. Equally, the results suggest that the information held by sophisticated traders also informs prices of PTBs, even though they are deterred from betting directly in these markets, thereby driving the prices of both strategic groups of bookmakers towards efficiency.

The remainder of this paper is structured as follows: Section 2 reviews the literature concerning the behaviour of bookmakers and bettors and this is employed in Section 3 to develop appropriate hypotheses. Details of the data set employed, the nature of the specific bookmakers whose odds are examined and the methodology we employ to test the hypotheses are presented in Section 4. Results are presented in Section 5. A discussion of the implications of the results and suggestions for future research are provided in Section 6.

2. Literature Review

2.1 Bookmakers' Behaviour – Theory and Evidence

Position Taking Bookmakers

Several authors suggest that bookmakers take positions and set prices in order to maximize their expected profit. For example, Levitt (2004) argues that bookmakers are better at predicting game outcomes than the typical bettor and that they set prices in order to exploit this advantage. This can yield greater profit than could be obtained if the bookmakers acted like traditional market makers and attempted to set prices to balance supply and demand. Evidence that some bookmakers `take a position' is provided by Levitt (2004), Paul and Weinbach (2007, 2008) and Humphreys (2010).

Kuypers (2000) argues that such bookmakers seek to maximize profits and can even set odds that deviate from those indicated by unbiased probability estimates if they come to different conclusions regarding how bettors place their bets. Similarly, Marshall (2009) suggests that bookmakers could remove odds discrepancies with competitors, but often do not if they believe that their own odds better reflect the outcome probabilities; suggesting that the objective of bookmakers' price-setting is to maximize profit rather than remove risk. Equally, Franck, et al. (2013) suggest that PTBs purposefully quote some odds above those of their competitors as a loss-leader strategy to attract customers, a policy consistent with the theory of spatial price dispersion (Marshall, 2009)[1]. Whilst this strategy does not necessarily maximize their profit per game, and may involve accepting bets with a negative expected value, it may be consistent with maximization of long-term profit. As Grant et al. (2019) point out, these bookmakers often refuse bets from those they discern to be informed

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

customers. Consequently, their policy of maximizing the customer base should maximize their long-term profits.

Book Balancing Bookmakers

An alternative operating model is that bookmakers set prices to eliminate risk by balancing the potential liabilities across all possible outcomes; guaranteeing their payoff is close to their over-round, irrespective of event outcome (advocated by Sidney (2003), in a text designed to educate bookmakers). Magee (1990) also states that bookmakers adjust their odds regularly, in order to achieve a balanced book, their odds reflecting the money staked on each possible outcome (Woodland and Woodland, 1991). Theoretical models of bookmaker behaviour also often take this perspective (e.g., Fingleton and Waldron, 1999; Hodges and Lin, 2009) According to this business model, bookmakers effectively act as market makers whose profits are merely a function of the volume traded.

From the bettor's perspective, a BBB offers a similar form of market to a betting exchange, where trades are conducted directly between the different bettors, the betting exchange (which facilitates the transaction) obtaining a commission from the winner. A similar mechanism operates within a BBB, with the BBB effectively acting as an intermediary, receiving a small share of turnover in exchange for providing liquidity to match competing bettors' orders. The bookmaker-driven football betting market offers significantly greater liquidity than the comparable betting exchange market (Franck et al., 2013; Duffie, 2012)

2.2. The Population of Bettors

The betting public is heterogeneous (Gainsbury, et al. 2012) but for the purpose of our study, we distinguish between two types of bettors; namely, the casual betting public, whose bets exhibit negative average expected returns, and a much smaller group of informed bettors whose expected returns may be positive. The latter group may include bettors who (i) possess inside information (Shin, 1991, 1992, 1993) and whose existence in the football market has been observed by Forrest (2012); (ii) arbitrageurs; i.e. those who attempt to benefit from pricing differences across different betting operators (e.g., Marshall, 2009; Franck, et al., 2013), and (iii) bettors who successfully apply mathematical models to profit from betting (e.g., Benter, 1994; Thorp, 2000).

Grant et al. (2019) suggest that BBBs restrict trade with those bettors whom they believe to hold superior information, by implementing restrictions on the size of stake or type of bet they are willing to accept (e.g. restrictions on arbitrage betting), or simply cancelling traders' accounts (Veitch, 2009). Franck et al. (2013) identify potential arbitrage opportunities between odds of BBBs and PTBs and between betting exchanges and bookmakers' odds. However, Grant et al. (2019) argue that these arise because PTBs intentionally misprice events in order to attract customers and these are effectively non-exploitable in the long-run because PTBs restrict the activity of those skilled in identifying these opportunities. Marshall (2009) found the median period for arbitrage opportunities to exist in football markets to be 15.4 minutes. If price dispersion were not intentional, it is likely that these opportunities would be quickly removed. The fact that they are not, provides further evidence that PTBs apply discriminatory policies against skilled bettors. Furthermore, Levitt (2004), analysing bettor-specific data from a PTB, found little evidence of individual bettors who systematically beat the bookmaker, supporting the view that PTBs deter informed traders.

Conversely, for a BBB the objective is to maximize trading volume, as this is the only determinant of such an operator's profitability (Woodland and Woodland, 1991). Hence, a BBB has no incentive to eliminate potentially successful bettors, as BBBs take no position against them[2]. It might not even be possible for a BBB to discern whether a large bet arises from a professional gambler, from a regional PTB bookmaker, or from an insider (Forrest, 2012). As successful bettors are likely to apply reinvestment strategies, their stakes on individual games are likely to grow over time until they are bound by the market's staking limits (applied by BBBs to manage the ratio of volume from informed vs. less informed clients). Consequently, informed clients are likely to be significant suppliers of liquidity for a business model in which trading volume is the decision variable.

Since BBBs allow bets from informed traders, they use regular movements of odds to minimize potential exposure, as they set limits on the size of the stake that they are willing to accept at a given level of odds. In addition, they reduce the odds on whichever outcome receives a sizeable bet and increase the odds correspondingly on other outcomes. The result of this process is that the BBB secures a profit somewhat lower than the over-round, since the stakes on all outcomes might be higher when the corresponding odds were above the average level for a given offer during its life cycle. It is also likely that a greater decrease in the odds will occur after a bettor identified as skilled places a large bet (Forrest, 2012).

3. Theoretical Context and Development of Hypotheses

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

Levitt (2004) argues that it should be expected that “the most talented individuals would be employed as the odds makers” (p. 245) and, as a result, bookmakers (cf. bettors) will be able to forecast event outcomes more accurately. This suggests that PTBs set odds that efficiently reflect the corresponding event probabilities. Consequently, he argues, they do not have to adjust these odds often in order to balance their books, as taking positions will deliver higher profits. Accordingly, bookmakers that very frequently adjust their odds on the basis of trading volume, do not fit Levitt’s model; rather they might better be described as a BBB. So, for Levitt (2004), PTBs are superior forecasters of outcomes relative to BBBs. However, Smith, et al. (2006, 2009) and Franck et al. (2010, 2013) provide evidence that odds derived from betting-exchange markets constitute superior forecasts of match outcomes (cf. bookmaker markets), suggesting that demand-driven markets evaluate the outcomes of sport events better than the expert bookmakers. Kuypers (2000) and Forrest and Simmons (2008) suggest this difference in efficiency could be attributed to inefficient price-setting by bookmakers, as they try to take advantage of punters’ irrational betting or by bookmakers’ loss-leading activity, which creates such inefficiencies intentionally in order to maximise long-term profit. They have no incentive to adjust their odds in response to demand, even if this results in the generation of arbitrage opportunities, since, they can always refuse bets from arbitrageurs or other informed bettors.

By contrast, BBBs, whose profits are maximised by increasing turnover, have little incentive to restrict informed customers. Consequently, informed bettors are likely to have access only to BBBs in order to place large stakes (or, to a much lesser degree, betting exchanges, where sufficient volume may only be available on some events and where there are restrictions regarding the distribution of profit[3]). The implication of this is that information from skilled bettors is likely to be channelled through BBBs. The theories advocated by Levitt (2004), Kuypers (2000), Forrest and Simmons (2008) and Franck, et al. (2010) would suggest that this should have no impact on the prices of PTBs. In particular, Levitt (2004) argues they are superior forecasters to bettors and adjusting their odds in line with betting turnover would therefore be likely to reduce their profits. Equally, Kuypers (2000), Forrest and Simmons (2008) and Franck et al. (2010) argue that their price inefficiencies are intentional.

It should be clear, however, that by eliminating skilled bettors, PTBs lose direct access to the only market players likely to yield superior forecasts to their own, something that would help PTBs move their odds closer to the objective probabilities; thereby enabling them to gain a higher margin from their clients (i.e. casual bettors). As such, we would anticipate that if PTBs believe that their odds can be improved, based on information arising from the informed bettors, they are likely to adjust their odds according to BBBs price changes. Forrest (2012), based on anecdotal information, suggests that PTB's odds do indeed follow trends in Asian bookmakers' odds, and our aim is formally to test this proposition.

To explore the issues raised in the preceding discussion, we develop three hypotheses. First, we examine to what extent there is evidence that the actions of BBBs and PTBs differ systematically in terms of how frequently they adjust their odds, based on the differences in operating approach outlined above. This gives rise to H1:

[3] According to its regulations, Betfair (the most popular betting-exchange) can withhold up to 60% of a winning player's profit (http://www.betfair.com/www/GBR/en/aboutUs/Betfair.Charges/). This is an obvious deterrent for skilled bettors.
http://www.theguardian.com/sport/2011/jun/29/betfair-premium-charge-increase

The BBBs change odds more frequently and charge lower transaction costs (over-round) per unit stake bet than PTBs.

As explained above, we expect that informed bettors are only likely to trade with the BBBs, whose odds will therefore reflect this information. We suggest that PTBs will use these odds to adjust their own, to ensure that their odds better reflect the true event probabilities. Consequently, we expect PTB's odds to lag those of BBBs and we therefore test H2a: *Odds changes at PTBs converge to lagged odds changes at BBBs (SBOBet)*, and H2b: *Odds changes at BBBs are not influenced by lagged odds of PTBs*.

Levitt (2004) argues that PTBs are superior forecasters (cf. their customers) of match results, suggesting that their final odds will also be superior predictors of event probabilities. However, if we find evidence to support H2a and H2b, we expect closing odds of PTBs to be influenced by BBBs early odds, whereas the closing odds of BBBs to be unrelated to those of PTBs. Consequently, we test H2c: *The closing odds of PTBs will be related to both their own early odds and the early odds of BBBs, whereas the closing odds of BBBs will be related to their own early odds, but not the early odds of PTB.*

The Efficient Market Hypothesis (EMH) would suggest that valuable information held by all bettors should influence odds over time. In the football market, as kick-off approaches, more of the relevant information related to match outcome is revealed (e.g. the teams' final line-ups, weather conditions) and the maximum stakes accepted by BBBs increases significantly. Consequently, bettors who wish to avail themselves of the maximum information and volume (and this is likely to apply to the informed bettors) are likely to bet close to market close (kick-off). As a result, we expect later odds to be more informative of match outcomes than earlier odds for both BBBs (driven by the flow of *smart money*) and PTBs (following odds adjustments based on BBBs odds). However, because the informed bettors trade mainly with BBB, their closing odds are likely to be better predictors of match outcomes. Consequently, we test H3a: *Closing odds will be better predictors of match outcomes than early odds, for both BBBs and PTBs*, and H3b: *Closing odds of BBBs (cf. PTBs) are better predictors of match outcomes.*

4. Data and Methodology

4.1. Identifying typical BBB and PTB bookmakers

To test our hypotheses, we collected time-stamped odds data from typical representatives of a PTB and a BBB, namely Ladbrokes (LAD) and SBOBet (SBO). In order to align our findings with the results presented in Grant et al., (2019), who identified the features which distinguish BBBs from PTB's, we collected data from the 2012/13 season. Importantly, also, the choice of this season enables us to capture the effects of new entry and a transformed market setting early in the period where co-existence of PTBs and BBBs on a non-trivial scale had become clearly established.

We now provide a brief overview of the operations of LAD and SBO in 2012, which clearly identify them as a PTB and a BBB, respectively.

LAD is a traditional UK bookmaker (incorporated in Gibraltar), founded in 1886, with over 16,000 employees, and it claims to be the most recognised betting brand in the United Kingdom.[4] It operates more than 2,800 retail-betting shops in the UK, Ireland, Belgium, and Spain and attracts over 1 million active clients.[5] Its financial statements in 2012, indicate that they laid £17 billion in bets across all events, resulting

[4] Ladbrokes 2012 Annual Report, p.15

[5] http://www.ladbrokesplc.com/about-ladbrokes.aspx

in net revenue of over £1 billion.. LAD offers a diversified range of gambling services, including racing, sports and political betting, online casino games, poker, bingo, and in-shop slot machines. It is likely that LAD uses sports betting to attract customers to these other (less-risky) operations. Its 2012 annual report indicates that customer acquisition costs were £107 per customer, inclusive of promotions and bonuses.

Franck et al. (2013, p. 311) point out that LAD actively discourages informed bettors, 'reserv[ing] the right to refuse part or all of a bet' and use historical trades (e.g., via cookies, log files, clear gifs) to create customer profiles to identify and restrict the activities of potential arbitrageurs or informed bettors who generally bet when odds are favourable. Policies such as these allow LAD to operate under the high-margin, low-turnover model of the traditional PTB.

LAD notes under 'key risks' in its 2012 annual report (p. 23): "the online gambling market is characterized by intense and substantial competition and by relatively low barriers to entry for new participants. In addition, LAD faces competition from market participants who benefit from greater liquidity as a result of accepting bets from jurisdictions in which LAD chooses not to operate." These restricted territories include China. As Forrest (2012) notes, much of the betting on football now comes from South-East Asia (directly or indirectly), and LAD appears to have made the conscious decision not to compete with the Asian bookmaking market.[6] The 2012 Annual Report also notes (p. 24) that the company faces "a relatively high fixed-cost base as a proportion of total costs, consisting primarily of employee, rental and content costs associated with the betting shop estate." Although LAD's high physical presence in the market is not necessarily a driver of its odds-setting process, it does illustrate

[6] The website analytics service, Alexa, reported on Sept. 11, 2013, that 56% of the traffic to Ladbrokes website came from the UK, with 5% from the USA, followed by minor percentages of traffic from elsewhere.

that it is unlikely to be able to operate at the ultra-efficient levels of an online-only sports betting agency.

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

The Institute of International and Strategic Relations (IRIS) Report (2012) presents a detailed analysis of betting markets in Asia. It identifies SBO as one of the four major Asian players, and notes (p. 44) that it ‘represents heritage of an activity begun in 1994 in Singapore that spread to Malaysia, Indonesia and then the Philippines, where the sports betting business acquired an online betting licence in the economic area of Cagyan, a lax jurisdiction.’ The web analytics service, Alexa, reveals that the majority of visitors to SBOBet.com hail from South-East Asia. However, the few Asian online bookmaker sites represent the over-ground section of a vast pyramid (IRIS 2012). SBO represents the highest level of a large pool of regional bookmakers, which collect bets from the wider population through a set of localized bookmakers which hedge their own risks online. Hence, SBO regularly accept very large bets, as an amalgamation of a portfolio of small bets. The IRIS report (2012) notes that ‘these sites offer a particularly high rate of return to the bettor (around 97%), the low margin being offset by the very high volume of bets.’ The figures quoted suggest that a client can place a stake with one of the large Asian bookmakers of around 20 times the amount that a European bookmaker would accept for a major European football championship.

SBO differs from most European bookmakers in that it publishes the maximum amount it is willing to accept on an outcome. If bettors were to seek a larger stake, it would create an imbalance in SBO's liabilities. In reporting market depth, SBO shows that it is relatively indifferent to the identity of the counterparty. The lack of discrimination in counterparties, and the more transparent structure, underpins identification of SBO as a BBB.

4.2. Betting data

We analyse time-stamped home team, draw, and away team (1x2) odds collected from both LAD and SBO, at points 1 day, 16 hours, 8 hours, 4 hours, 2 hours, 1 hour, 30 minutes and 1 second prior to kick-off for all 2132 football matches played in season 2012/13 in the 6 most prominent European football leagues: The English Premier League, the Spanish La Liga, the Italian Serie A, the German Bundesliga, the French Ligue 1 and the Dutch Eredivisie. This resulted in a total of 51, 168 [2,132 × 8 (time points) x 3 (1x2 outcomes)] individual odds offers per bookmaker on potential game outcomes.[7]

We posted requests to the bookmakers' servers to obtain the data and decided on eight time points at which to collect odds, to trade-off between data availability and reliability. Increasing the number of time points per game would have increased the number of requests we made to the bookmakers' servers, posing the risk that the bookmakers might restrict our access to their websites. The volume of betting tends to increase as match kick-off approaches. Consequently, we collected data at smaller

[7] Although the traditional match outcome (Home team win, draw, away team win) betting is most popular in Europe, SBO specialises in Asian Handicaps. We chose to use match outcome data as these were most likely to be liquid at the position-taker, and hence most likely to lead to reliable odds movements.

intervals in time as kick-off approached, in order to capture a similar amount of betting activity within each period.

4.3. Methodology

4.3.1. Differences between bookmakers

Calculation of transaction costs for each bookmaker

The level of transaction costs for each bookmaker ('over-round') is measured by the extent to which the sum of their odds-implied probabilities across all match outcomes exceeds unity. We measure the over-round, $\rho_{i,k,t}$, at each time point, t in the lead-up to game i, for each bookmaker $k \in \{SBO, LAD\}$, by adding the inverse of the gross payoffs per dollar bet $X_{i,k,t,o}$ for each of the three outcomes $o \in E = \{H, D, A\}$ in the football match (H denoting home win, D denoting draw, and A denoting away win):

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

The average overround, $\bar{\rho}_{k,t}$ for each bookmaker, across all matches, for all points in time is computed and the bookmakers' average trading costs are compared using the pooled t-test, with test statistic T:

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

where

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

is the estimate of the common standard deviation of the two samples, and n_1 and n_2 are the number of observed matches for SBO and LAD, respectively. Our first hypothesis predicts that the transaction costs for SBO should be significantly lower than their PTB counterpart, LAD; i.e. the test statistic *T* should be significantly negative.

Estimation of Frequency of Odds

We calculated the average frequency of odds changes per game across the sample in order to test whether the BBB (cf. PTB) moves its odds significantly more often. First, we calculate the difference in prices for each bookmaker for each match-outcome-time point:

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

As we are only interested at this point in whether the odds move, rather than the actual size of the movement, we tally each match-outcome-time point based on a categorical score, $M_{i,kt,o}$:

$$\begin{array}{c} M_{i,k,t,o} = 1 \quad if \quad \Delta X_{i,k,t,o} \neq 0 \\ M_{i,k,t,o} = 0 \quad if \quad \Delta X_{i,k,t,o} = 0 \end{array} \qquad (3)$$

A binomial test is carried out to determine if the frequency of odds movements at SBO (cf LAD) is greater, i.e. $\frac{M_{i,SBO,t,o}}{n_1} > \frac{M_{i,LAD,t,o}}{n_2}$.

4.3.2. Transmission of Information within the Market

Serial Correlation in Odds Movement

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

Odds movements across outcomes within games are complementary, inasmuch as an increase in odds on the favourite winning will generally coincide with a decrease in odds on the longshot winning. Draw outcomes are notoriously difficult to predict for both experts and models (e.g. Pope and Peel, 1989; Goddard, 2005). Moreover, movements in draw odds are similarly likely to be driven by bets on one of the other outcomes, rather than based on specific information concerning the likelihood of this outcome. Thus, we restrict our analysis to only the favoured team (defined as the team with the higher gross payoff-reciprocal, or odds-implied probability at SBO, one-day before kick-off in each game[8]).

We conduct a Fisher-type unit root test on panel data (Choi, 2001) for each bookmaker separately to test odds movements for stationarity. Each panel consists of match-time point observations of bookmaker odds. We use the BIC to find the optimal

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

number of lags to include in our regression model to test H2; allowing a trade-off between model fit and parsimony.

Modelling Odds Movements

In order to test H2a, that PTBs' odds converge to those of BBBs and H2b, that BBBs' odds are not influenced by PTBs' odds, we employ a random-effects model. A random effects model is chosen to account for any unobserved variation in odds due to match-specific factors. As with the examination of lag length, the panel in the random-effects model consists of match-time observation points for each bookmaker. The general form of the random-effects models is as shown in Equations (4a) and (4b):

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

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

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

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

where n is the number of observations[9], $\Delta X_{i,t,k}$ is the realised first difference in odds between time $t-1$ and time t at bookmaker k, and g is the degrees of freedom (the number of parameters in the model minus one.) The first term in the BIC penalises poor model fit, while the second term penalises the number of parameters required to achieve the model fit. Thus, a smaller value of the BIC means the model is preferred.

We expect that the models described by Eqs. (4a) and (4b) can be simplified. In particular, since H2 is concerned with whether a particular bookmaker's prices will converge to those of the other bookmaker, we expect the difference in the two bookmakers' prices in the previous lag to be the main determinant of odds. Therefore, Eq (3) should reduce to:

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

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

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

As we suspect that informed bettors can only access BBBs, we expect that prices in that market will reflect quality information, which is unavailable directly to PTBs. Therefore, the latter are expected to react to odds changes in the BBBs' market and adjust their odds when these deviate from those offered by BBB. Therefore, we expect a significantly positive value for B when we estimate Eq.(6a). Similarly, Eq.(6b)

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

tests the influence of LAD's odds on the odds changes of SBO. If the informed traders are driving the market, support for H2b would be found if B in Eq.(6b) is close to zero. Alternatively, if bookmakers are the most accurate forecasters of sports events (Levitt, 2004), there should be little convergence from PTBs to BBBs' odds (i.e. the coefficient B would be close to zero in Eq.(6a)).

Early-to-Late Odds Movements between Bookmakers

To further test our view that odds of PTBs are driven by movements in BBBs odds (but not *vice versa*), we examine whether the PTB's closing prices are related to early prices of themselves *and* the BBB, whereas the BBB's closing prices are only related to their own early prices. To achieve this, we utilize a fixed-effects approach, incorporating normalized odds-implied probabilities for all three outcomes along one dimension of the panel, and all games along the other.

First, both bookmakers' payoffs ($X_{i,t,k,o}$) one day before kickoff (time t_0) and at the time one second before kick-off (time T) are converted to normalised probabilities ($Y_{i,t,k,o}$). We achieve this by dividing bookmaker k's payoff reciprocal for outcome *o* by the gross over-round for the game i at time t from Eq. (1):

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

The fixed-effects models used to test the information content of odds takes the following form:

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

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

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

The models in Eq.(7a) and Eq.(7b) help test the influence of the early SBO odds on the terminal odds of LAD and the influence of early LAD odds on the terminal odds of SBO, respectively. If, as we suspect, the influence of informed bets at the BBB drives the odds in the PTB's market, whereas the PTB's odds do not influence the BBB, then β_2 in Eq. 7a should be significant and positive, and the coefficient β_1 in Eq. 7b should be close to 1, while the coefficient β_2 should be insignificant and close to zero.

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

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

4.3.3. The Efficiency of Odds-Based Estimates

According to H3a and 3b, bookmakers' closing odds are expected to be more efficient predictors of actual game outcomes compared to early odds and BBBs' odds are expected to be more accurate than those of the PTBs. To test these hypotheses, we estimate conditional logit (CL) models (McFadden, 1974); these have been used to test the efficiency of odds in many previous betting studies (e.g., Bolton and Chapman, 1986; Benter, 1994; Sung and Johnson, 2010). The CL model is used with only odds as an explanatory variable, and takes the following form, where $E_{i,t,k,o}$ is the probability of the outcome $E \in \{H, D, A\}$ occurring in game i at time t, with odds on outcome o from bookmaker k:

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

where

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

We estimate model (8a) for each bookmaker at times t_0 one day prior to kick-off and $T,$ one second before kickoff. The model fit is evaluated using McFadden's (1974) pseudo-R^2 statistic; a higher pseudo-R^2 implying a superior model fit. Support for H3a would arise if earlier odds have less explanatory power than later odds (i.e. if the model fit at time T is greater than at time t_0 for both bookmakers). Support for H3b would be given if the model fit is higher for SBO cf. LAD at times t_0 and T).

5. Results

5.1. Differences between Bookmakers

Our results clearly demonstrate that SBO odds are subject to more changes than LAD odds. In the first three columns of Table 1 we show the proportion of occasions in various time intervals prior to kick off when LAD and SBO changed their odds. It is clear from this table that SBO (cf. LAD) change their odds more frequently in a given time interval prior to a kick-off. This is confirmed by the difference in the mean number of odds changes across time intervals during the 24 hours prior to the game between SBO (5.36 per match, SD= 1.03) and LAD (0.795 per match, SD= 1.24). A t-test confirmed that the difference is unlikely to be random (p-value= 0.000). These results provide strong support for our view, expressed in H1, that BBBs change their odds more frequently than PTBs and confirms the view of Levitt (2004) that PTBs rarely move prices in the lead-up to games.

Table 1: Average over-round and proportion of occasions when LAD and SBO changed their odds, in various time-periods prior to kick-off

	Proportion of occasions odds changed		**Average over-round**	
Time period prior to kick- off	**LAD**	**SBO**	**LAD**	**SBO**
over 1 day-16 hrs	0.09	0.64	0.077	0.073
16 to 8 hrs	0.11	0.75	0.077	0.070
8hrs to 4 hrs	0.22	0.86	0.077	0.070
4hrs to 2 hrs	0.15	0.77	0.077	0.065
2hrs to 1hr	0.09	0.73	0.077	0.064
1hr to 30 mins.	0.09	0.78	0.077	0.064
(5 mins to 1 sec)	0.14	0.87	0.077	0.064

In columns 4 and 5 of Table 1, we present results regarding the bookmakers' over-rounds (calculated using Eq.(1)) at various stages of the market. The t-tests confirm that SBO (cf. LAD) clearly operate with a lower over-round at all points in time, and the difference in over-round increases as the kick-off approaches. These results provide support for the view, expressed in H1, that BBBs operate with lower margins than PTBs; the lower margin operated by BBBs generating higher volumes, as more informed traders are attracted by more advantageous odds.

5.2. Transmission of Information from Bettors to Bookmakers

The results of the Fisher type unit-root test, presented in Table 2, show that the first differences in the odds are stationary for both SBO and LAD. Therefore, we proceed to modelling such differences using Equations (4a) and (4b).

Table 2: Fisher-Type Unit-Root test statistics for the four stationarity tests described by Choi (2001), with Ho: The panels contain a unit-root.

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

We apply the BIC in order to identify the optimal number of lags, trading-off between fit and complexity. The results are presented in Table 3. For SBO (Panel A), the model producing the lowest BIC value (-9,075) is that incorporating the constant term only, and thus Equation (4b) is best modelled using $\Delta X_{i,t,SBO} = c$. The model incorporating the lagged odds terms from both SBO and LAD, $\Delta X_{i,t,SBO} = c + b_{11}X_{i,t-1,LAD} + b_{12}X_{i,t-1,SBO}$, produces a significantly worse BIC (-9,062, p-value 0.002). This implies that odds movements at SBO are not related to previous odds movements in either SBO or LAD prices, or at least there is no improvement in forecasting power from adding lagged prices from either bookmaker.

Table 3: Results related to model selection for odds changes at SBO and LAD.
The first column shows the right-hand side of the non-nested model under consideration, with one, two, and three lags of each bookmaker's odds used, respectively. The second and third columns report the BIC and R^2 of the non-nested model, respectively. The fourth column shows the nested model under consideration (i.e. the nested model having produced the lowest current BIC value). The fifth column reports the BIC of the nested model. The final column reports the significance of the difference in BIC values, calculated as p-value = $\exp\{(\mathrm{BIC}_{\mathrm{lower}} - \mathrm{BIC}_{\mathrm{higher}})/2\}$.

Panel A: Model Selection for Odds Changes at SBO.
Dependent Variable: Change in Prices at SBO ($\Delta X_{i,t,SBO}$)

Non-Nested Model	BIC	Model R^2	Nested Model	BIC	p-value*
$c + b_{11}X_{i,t-1,LAD} + b_{12}X_{i,t-1,SBO}$	9,062	0.20%	c	9,075	0.002
$c + b_{11}X_{i,t-1,LAD} + b_{12}X_{i,t-1,SBO} + b_{21}X_{i,t-2,LAD} + b_{22}X_{i,t-2,SBO}$	-6,068	0.30%	c	-6,082	0.001
$c + b_{11}X_{i,t-1,LAD} + b_{12}X_{i,t-1,SBO} + b_{21}X_{i,t-2,LAD} + b_{22}X_{i,t-2,SBO} + b_{31}X_{i,t-3,LAD} + b_{32}X_{i,t-3,SBO}$	-3,509	0.55%	c	-3,510	0.479

Panel B: Model Selection for Odds Changes at LAD.
Dependent Variable: Change in Prices at LAD ($\Delta X_{i,t,LAD}$)

Non-Nested Model	BIC	Model R^2	Nested Model	BIC	p-value*
$c + b_{11}X_{i,t-1,LAD} + b_{12}X_{i,t-1,SBO}$	-53,973	8.75%	c	-52,822	0.000
$c + b_{11}X_{i,t-1,LAD} + b_{12}X_{i,t-1,SBO} + b_{21}X_{i,t-2,LAD} + b_{22}X_{i,t-2,SBO}$	-45,102	9.81%	$c + b_{11}X_{i,t-1,LAD} + b_{12}X_{i,t-1,SBO}$	-45,051	0.000
$c + b_{11}X_{i,t-1,LAD} + b_{12}X_{i,t-1,SBO} + b_{21}X_{i,t-2,LAD} + b_{22}X_{i,t-2,SBO} + b_{31}X_{i,t-3,LAD} + b_{32}X_{i,t-3,SBO}$	-37,326	10.08%	$c + b_{11}X_{i,t-1,LAD} + b_{12}X_{i,t-1,SBO}$	-37,242	0.000

*Lower BIC model better than higher BIC model

Examination of the models' BICs estimated for changes in LAD odds, shows that there may be marginal improvement over the single-lag model by adding in second- and third-lags of both LAD and SBO odds. For example, the second row of Panel B of Table 4 shows that the BICs of the models with two and one lags are -45,102 and -45,051, respectively. However, the marginal improvement in model fit, although significant (p-value of 0.000), reduces the sample size significantly (by 12.5% for each additional lag). Consequently, we retain the single lag specification for modelling the changes in LAD odds.

We present the results of estimating the random-effects model in Eq. (4a), using one period lagged odds from both bookmakers as the independent variables, in Table 4. The coefficients for the lagged odds for the two bookmakers are nearly identical in magnitude, but with opposing signs. Rearranging the model:

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

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

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

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

Consequently, the results presented in Table 4, suggest that at any time up to kick-off, around 9.5% of the odds at LAD can be explained by the lagged odds at SBO, while the remaining 90.5% are explained by the lagged odds at LAD. Importantly, the coefficient of SBO's lagged odds is positive and significant at the 1% level. Hence, these results support H2a, that LAD's odds converge towards the odds of SBO.

Table 4: Results of estimating the random effects panel model.

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

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

This model produces a relatively low R^2 value (8.74%), partially because in most cases LAD do not change their odds, i.e. $\Delta X_{i,t,LAD} = 0$. This leads us to undertake further analysis of the cases in which LAD's odds *have* moved, i.e. when $\Delta X_{i,t,LAD} \neq 0$, which occurs in 966 matches (with 1,572 odds movements). We estimate the model represented by Eq. 4a with one lag and we present the results in Table 5. The increase in the magnitude of the coefficients when examining cases for which

LAD's odds move is substantial; nearly 50% of the variation in LAD's odds moves can be explained by the deviation of their lagged odds from those of SBO. In addition, the R^2 of the model has increased significantly to 49.89% (from 8.74%). These results provide strong support for the view that a key driver for PTBs to change their odds are differences between their odds and those of BBBs.

Table 5: Results of estimating the random effects panel model for the sample of 966 European football matches at time points for which LAD's odds moved, $\Delta X_{i,t,LAD} \neq 0$.
This table reports the coefficients from estimating the following regression model (Eq. 4a with one lag): $\Delta X_{i,t,LAD} = c + b_{11}X_{i,t-1,LAD} + b_{12}X_{i,t-1,SBO} + U_i + e_{i,t}$, where$X_{i,t,k}$ indicates the gross odds X offered on game i at time t by bookmaker k, on the favoured outcome. The coefficient and corresponding standard errors are reported in the second column, the third column shows the significance level of each coefficient in the regression: (*), (**), and (***) denoting significance at the 10%, 5%, and 1% level, respectively. The fourth column shows the p-value of the statistic. The standard deviation due to the random-effects design is σ_U and the standard deviation due to the white noise error term is σ_e.

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

To further test Hypothesis 2a, we model the change in LAD's odds, for the restricted sample of 1,572 occasions when their odds did move, against the lagged difference in SBO's and LAD's prices ($X_{i,t-1,SBO} - X_{i,t-1,LAD}$) directly. We present the results of estimating this model in Table 6. The single-factor specification shows a similar magnitude of coefficients; with about 50% of the movement in LAD's odds explained by the lagged difference in SBO's and LAD's odds.

Table 6: Results of estimating the random effects panel model for the sample of 966 European football matches which involve periods when LAD's odds did move, $\Delta X_{i,t,LAD} \neq 0$. This table reports the coefficients from equation 4a with one lag: $\Delta X_{i,t,LAD} = c + B(X_{i,t-1,SBO} - X_{i,t-1,LAD}) + U_i + e_{i,t}$, where $X_{i,t,k}$ indicates the gross odds X offered on game i at time t by bookmaker k, on the favoured outcome. The coefficient and corresponding standard errors are reported in the second column, the third column shows the significance level of each coefficient in the regression: (*), (**), and (***) denote significance at the 10%, 5%, and 1% level, respectively. The fourth column shows the p-value of the Z-statistic. The standard deviation due to the random-effects design is σ_U and the standard deviation due to the white noise error term is σ_E.

	Coefficient (Std. Error)	Sig.	Z-Stat. (P-value)	
$X_{i,t-1,SBO} - X_{i,t-1,LAD}$	0.5038 (0.0129)	(***)	38.91 (0.0000)	
c	-0.0234 (0.0018)	(***)	-12.80 (0.0000)	
σ_U	0.0367			
σ_e	0.0503			
ρ (Fraction of Variance due to U_i)	0.3476		R^2	
n. observations	1,572		Within	0.3210
n. groups	966		Between	0.5828
Wald $\chi^2(2)$	1,513	(***)	Overall	0.5000

The results reported in Table 6, show a significantly negative sign for the constant on the random-effects regression, suggesting that a bettor's payoff for a winning bet with LAD is typically slightly lower than they could have obtained with SBO on the same favoured outcome (largely due to the differences in over-round). Thus, if LAD's odds were below (above) SBO's odds in the previous period, we expect less (more) than half of the difference to be made up by the change in odds at LAD. For example, if at $t-1$, SBO and LAD were offering gross odds of \$1.90 and \$1.70, respectively, the results suggest that LAD's prices at time t would be \$1.70 + 0.5038(\$1.90 - \$1.70) – 0.0234 ≈ \$1.78. The predicted price would also be \$1.78 at time t if LAD were offering \$1.90 and SBO were offering \$1.70 at time $t-1$. These results therefore support H2a, suggesting that when a PTB moves its prices, it does so towards BBB's odds. If we can demonstrate that BBB's (cf. PTB's) odds discount more information concerning match outcome, this will support our view that information arising from informed bettors is indirectly transmitted to PTBs, via BBB's odds.

5.3. The Relative Efficiency of Odds-Based Forecasts

We present the results of the estimation of Eq. (7a) and (7b), using, respectively, last-minute odds of LAD and SBO as the dependent variables in Tables 7 and 8.

Table 7: Results of estimating the fixed effects regression model (Eq. 7a).

The coefficients and corresponding standard errors are reported in the second column, the third column shows the significance level of each coefficient in the regression: (*), (**), and (***) denoting significance at the 10%, 5%, and 1% level, respectively. The fourth column shows the p-value of the statistic. The standard deviation due to the fixed-effects design is σ_A and the standard deviation due to the white noise error term is σ_E.

Bookmaker's odds-implied probabilities one day before kickoff	Coefficient (Std. Error)	Sig.	P-value	
$Y_{i,t_0,SBO,o}$	0.4618 (0.0252)	(***)	(0.0000)	
$Y_{i,t_0,LAD,o}$	0.5547 (0.0248)	(***)	(0.0000)	
c	-0.0040 (0.0010)	(***)	(0.0000)	
σ_A	0.0060			
σ_e	0.0236			
ρ (Fraction of Variance due to α_i)	0.0612		R^2	
n. observations	3,397		Within	0.9921
n. groups	1,700		Between	0.9426
F(2, 1695)	105,770	(***)	Overall	0.9912

Table 8: Results of estimating the fixed effects regression model (Eq. 7b) with the odds-implied probabilities of SBOBet $Y_{i,T,SBO,o}$ as dependent variable and the day-ahead odds-implied probabilities of LAD and the day-ahead odds-implied probabilities of SBO as independent variables, using the full sample of 2132 European football matches: $Y_{i,T,SBO,o} = c + \beta_1 Y_{i,t_0,SBO,o} + \beta_2 Y_{i,t_0,LAD,o} + \alpha_i + u_{io}$, where $Y_{i,t,k,o}$ indicates the odds-implied probability Y offered on game i at time t by bookmaker k, on match outcomes o. The coefficients and corresponding standard errors are reported in the second column, the third column shows the significance level of each coefficient in the regression: (*), (**), and (***) denoting significance at the 10%, 5%, and 1% level, respectively. The fourth column shows the p-value of the statistic. The standard deviation due to the fixed-effects design is σ_A and the standard deviation due to the white noise error term is σ_E.

Bookmaker's odds-implied probabilities one day before kickoff	Coefficient (Std. Error)	Sig.	p-value	
$Y_{i,t_0,SBO,o}$	1.0111 (0.0506)	(***)	(0.0000)	
$Y_{i,t_0,LAD,o}$	-0.0074 (0.0498)		(0.8810)	
c	-0.0034 (0.0020)	(*)	(0.0810)	
σ_A	0.0140			
σ_e	0.0475			
ρ (Fraction of Variance due to α_i)	0.0796		R^2	
n. observations	3,397		Within	0.9673
n. groups	1,700		Between	0.7399
F(2, 1695)	25,069	(***)	Overall	0.9624

It is evident that early odds posted by SBO do exhibit significant forecasting power in predicting closing LAD odds, whereas the converse is not true. In particular, the results presented in Table 7 show that the early odds-implied probabilities of match outcome from SBO, $Y_{i,t_0,SBO,o}$, are highly significant in predicting the closing LAD's odds-implied probabilities, $Y_{i,T,LAD,o}$; the coefficient (0.4618), indicating that closing LAD's odds converge, on average, about halfway towards early SBO odds. Furthermore, the results in Table 8 show that the closing odds of SBO are unrelated to the early odds of LAD. Consequently, LAD's price adjustments appear to be influenced by the prices in the BBB market, whereas SBO's odds moves are not predictable; providing further support for H2a and H2b.

As a further robustness check, we present the results of estimating a fixed effects model of the closing LAD's odds-implied probabilities at kick-off as a function of the early LAD's odds implied probabilities only (see Table 9). As we expected, the coefficient of the early LAD's probabilities is very close to 1 when no other variables are added to the model. However, when comparing this model to that presented in Table 7, the improvement in forecasting power by adding the SBO's odds-implied probabilities is clearly shown (a likelihood-ratio test exploring if the model in Table 7 nests the model in Table 9, is emphatically rejected: $\chi^2(1)$ = 615.87, p-value = 0.000). Hence, we conclude that SBO's odds are a significant determinant of the odds posted by LAD at kick-off.

In order to test the forecasting power of different sets of odds for predicting the outcome of football games (H 3a and 3b), we estimate Eqs. (8a) and (8b), using the early and late (log of) odds-implied probabilities from the two bookmakers (see Table 10). The magnitude of the coefficients (around 1.10) and their standard errors (0.50) indicates that each of the odds sets imply a small bias towards pricing favourites (e.g. Bacon-Shone et al, 1992).

Table 9: The Results of estimating the fixed effects regression model (Eq. 7b) with closing odds-implied probabilities of LAD $Y_{i,T,LAD,o}$ as dependent variable and the day-ahead odds-implied probabilities of LAD as independent variable, using the full sample of 2132 European football matches: $Y_{i,T,LAD,o} = c + \beta_1 Y_{i,t_0,LAD,o} + \alpha_i + u_{io}$, where $Y_{i,t,k,o}$ indicates the odds-implied probability Y offered on game i at time t by bookmaker k, on all match outcomes o. The coefficient and corresponding standard errors are reported in the second column, the third column shows the significance level of each coefficient in the regression: (*), (**), and (***) denoting significance at the 10%, 5%, and 1% level, respectively. The fourth column shows the p-value of the statistic. The standard deviation due to the fixed-effects design is σ_A and the standard deviation due to the white noise error term is σ_E.

Bookmaker's Odds-Implied Probabilities One Day Before Kickoff	Coefficient (Std. Error)	Sig.	P-value	
$Y_{i,t_0,LAD,o}$	1.0076 (0.0024)	(***)	(0.0000)	
c	-0.0026 (0.0011)	(**)	(0.0140)	
σ_A	0.0057			
σ_e	0.0258			
ρ (Fraction of Variance due to α_i)	0.0469		R^2	
n. observations	3,397		Within	0.9905
n. groups	1,700		Between	0.9486
F(2, 1695)	176,287	(***)	Overall	0.9897

Although the odds-implied forecasts are highly correlated, the Psuedo-R^2 is higher and the BIC lower, for later odds (bottom half of Table 10) than for earlier odds (top half of the Table 10), at both SBO and LAD. The BICs are formally compared using the formula:

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

where BIC_{Lower} and BIC_{Higher} are, respectively, the models with the lower and higher observed BIC values. The improvement in the BIC value for late *vs*. early odds is significant for both SBO (p-value =0.000) and LAD (p-value = 0.020). This confirms H 3a; later odds are more efficient for both types of bookmaker.

Table 10: Results of comparing the efficiency of bookmaker's odds using CL models. The modelling is conducted using the method outlined in Equations (8a) and (8b): $P(E_{i,t,k,o} = 1) = \frac{e^{Z_{i,t,k,o}}}{\sum_{o=1}^{3} e^{Z_{i,t,k,o}}}$, where $E_{i,t,k,o}$ is the event home win, draw, or away win (one of which is 1 for each game i), with odds from bookmaker k at time t for outcome o, $Z_{i,t,k,o} = b \ln(Y_{i,t,k,o})$ (i.e. the logged odds-implied probabilities for each game i).

Bookmaker (time pre kick-off)	Coef. (SE)	p-value	Pseudo-R^2	Bookmaker (time pre kick-off)	Coef. (SE)	p-value	Pseudo-R^2
$Y_{i,t_0,SBO,o}$ SBO (-1 day)	1.1283 (0.0588)	(0.000)	0.1087	$Y_{i,t_0,LAD,o}$ LAD (- 1 day)	1.0996 (0.0571)	(0.000)	0.1072
BIC	3,623			BIC	3,630		
$Y_{i,T,SBO,o}$ SBO (-1 second)	1.1111 (0.0572)	(0.000)	0.1123	$Y_{i,T,LAD,o}$ LAD (-1 second)	1.0963 (0.0565)	(0.000)	0.1092
BIC	3,609			BIC	3,622		

We also compare the efficiency of the LAD's and SBO's late odds. We find, in line with H3b, that there is a relatively greater increase in efficiency from early to late odds at SBO than at LAD (pseudo R^2 increases from 0.1087 to 0.1123 c.f. 0.1072 to 0.1092). In addition, the difference in the efficiency of the late odds of these two bookmakers is substantial. We use the BIC comparison test from Eq. (10), to compare the efficiency of late odds at SBO and LAD and find that the former have a significantly lower BIC value (p-value = 0.000). Consequently, these results support H3b, that late odds at BBBs (cf. PTBs) are more efficient in predicting game outcomes.

Taken together, these results suggest that SBO improves the accuracy of its odds by responding to information arising from informed traders, and LAD reacts to the trend in SBO's odds, improving the predictive ability of its odds in turn.

6. Discussion

The results confirm that bookmakers can be classified as PTBs and BBBs, with PTBs attracting an unsophisticated clientele and being willing to lose money in the short-run by either providing incentives to bet or promotional odds, to earn profits in the long run against its customer base as a whole. There is evidence that they actively maintain a book of unsophisticated clients by restricting or excluding those bettors who are believed to be superior traders. Hence, PTBs operate against a relatively uninformed clientele, which places small bets at high margins. Our results also suggest that BBBs attract the bets of informed clients by operating under a regime of high-turnover and low margins. Prices at BBBs change regularly in response to the volume of bets made by clients as they move their prices in order to achieve a low, risk-free margin from a high volume of stakes.

Our results confirm that LAD (a typical PTB) charges higher transaction costs (over-round), and moves prices less frequently than a typical BBB (SBO). When LAD's prices do move, they tend to converge to SBO's odds; lagged SBO prices explaining around 50% of LAD price movements. These results are consistent with informed traders moving BBBs' prices, and this information slowly diffusing to the odds of PTBs. The actions of BBBs (e.g., lower transaction costs) seem to encourage informed traders to bet, resulting in significant information being transmitted from the informed betting public to their odds. As a result, the efficiency of their odds substantially improves as the event draws nearer. This practice contrasts with that of PTBs, which discourage the activity of informed traders. However, PTBs tend to be influenced by trends in the BBBs' odds, attempting to benefit indirectly from the flow of *smart money* in order to increase their odds' efficiency (potentially gaining a higher margin from their casual clients).

Overall, our results suggest that Levitt's (2004) view that bookmakers set the market and are superior forecasters compared to bettors is certainly no longer true across the whole market. Levitt's proposition, which is confined to the operation of PTBs, suggests that they have no incentive to react to price moves arising from a demand-driven market, as the betting public should not be able to improve on the bookmaker's expert estimates. Consequently, even when PTBs do move their odds because of betting volume, Levitt's argument would suggest that this can only be the result of risk-aversion rather than maximization of expected profit. As a result, this reaction should lead to less accurate estimations, as experts' (seen as bookmakers by Levitt (2004)) forecasts are adjusted to the opinions of noise-traders. However, we show that the supply-side of the market is *following* a demand driven market (BBBs), shaped by the stakes of informed traders. Consequently, prices in these markets are effectively processing information from trading volumes, leading market prices towards efficiency. Thus, we conclude that the population of bettors, whilst involving many noise-traders, does in fact include informed traders who are capable of significantly improving the accuracy of the prices.

Our results, therefore, suggest that a highly liquid globalized market such as the football betting market is capable of efficiently processing information arising from diverse global sources. This finding has broad implications. For example, it strongly supports the value of prediction markets, since we show that provided liquidity is present, volume weighted average prices constitute more accurate estimations of unknown true probabilities (cf., those of expert forecasters).

Finally, and importantly, whilst our study suggests that setting barriers to trade, such as the discrimination against skilled players, is likely to lead markets away from efficiency, the emergence of a strategic group which behaves more like book-balancers, suggests that information from informed traders will, to some extent, be transmitted to the prices offered by those practising discrimination.

References:

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

Baye, M. R., Morgan, J., and Scholten, P. A. (2004). Price dispersion in the small and in the large: Evidence from an internet price comparison site. *Journal of Industrial Economics*, 52, 463-496.

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

Benter, W. (1994). Computer based horse race handicapping and wagering systems: A report. In D. B. Hausch, V. S. Y. Lo, and W. T. Ziemba (Eds.), *Efficiency of Racetrack Betting Markets* (pp. 193-198). London: Academic Press.

Bolton, R. N., and Chapman, R. G. (1986). Searching for positive returns at the track: A multinomial logit model for handicapping horseraces. *Management Science*, 32, 1040-1060.

Choi, I. (2001). Unit root tests for panel data. *Journal of International Money and Finance*, 20, 249-272.

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

Fingleton, J., and Waldron, P. (1999). Optimal determination of bookmakers' betting odds: Theory and tests. *Working Paper, Trinity College Dublin*.

Forrest, D. (2012). The threat to football from betting-related corruption. *International Journal of Sports Finance*, 7, 99-116.

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

Franck, E., Verbeek, E., and Nüesch, S. (2010). Prediction accuracy of different market structures – Bookmakers versus a betting exchange. *International Journal of Forecasting*, 26, 448-459.

Franck, E., Verbeek, E., and S. Nüesch, E. (2013). Inter-market arbitrage in betting. *Economica*, 80, 300-325.

Gainsbury, S., Sadeque, S., Mizerski, D., and Blaszczynski, A. (2012). Wagering in Australia: A retrospective behavioural analysis of betting patterns based on player account data. *Journal of Gambling Business and Economics*, 6, 50-68.

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

Grant, A., Oikonomidis, A., Bruce, A.C. and Johnson, J.E.V. (2019) 'New entry, strategic diversity and efficiency in soccer betting markets: the creation and suppression of arbitrage opportunities', *The European Journal of Finance* (forthcoming)

Hodges, S., and Lin, H. (2009). Fixed odds bookmaking with stochastic betting demand. *Working Paper, Cass Business School*, *City College London*.

Humphreys, B. R. (2010). Prices, point spreads and profits: Evidence from the National Football League. *Working Paper*, *University of Alberta*.

Institute of International and Strategic Relations (IRIS) Report (2012). Sports betting and corruption: How to preserve the integrity of sport. Paris: L'Institutde Rélations Internationales et Stratégiques.

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

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

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

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

McFadden, D. (1974). Conditional logit analysis of qualitative choice behaviour. In P. Zaremba (Ed.), *Frontiers in Econometrics* (pp. 105-142). Academic Press.

Paul, R. J., and Weinbach, A. P. (2007). Does Sportsbook.com set pointspreads to maximize profit? Tests of the Levitt model of sportsbook behavior. *Journal of Prediction Markets*, 1, 209-218.

Paul, R. J., and Weinbach, A. P. (2008). Price Setting in the NBA gambling market: Tests of the Levitt model of sportsbook behavior. *International Journal of Sports Finance*, 3, 2-18.

Pope, P. F., and Peel, D. A. (1989). Prices and efficiency in a fixed-odds betting market. *Economica*, 56, 223-241.

Shin, H. S. (1991). Optimal betting odds against insider traders. *The Economic Journal,* 101, 1179-1185.

Shin, H.S. (1992). Prices of state contingent claims with insider traders, and the favourite-longshot bias. *The Economic Journal*, 102, 426-435.

Shin, H.S. (1993). Measuring the Incidence of insider trading in a market for state-contingent claims. *The Economic Journal*, 103, 1141-1153.

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

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

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

Sung, M.-C., and Johnson, J. E. V. (2010). Revealing weak-form inefficiency in a market for state-contingent claims: The importance of market ecology, modelling procedures, and investment strategies. *Economica*, 77, 128-147.

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

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

Woodland, B. M., and Woodland, L. M. (1991). The effects of risk-aversion on wagering: Point spread versus odds. *Journal of Political Economy*, 99, 638-653.