Improving Prediction Market Forecasts by Detecting and Correcting Possible Over-reaction to Price Movements

Ming-Chien Sung , David C.J. McDonald , Johnnie E.V. Johnson , Chung-Ching Tai , Eng-Tuck Cheah

PII: S0377-2217(18)30557-5
DOI: 10.1016/j.ejor.2018.06.024
Reference: EOR 15209

To appear in: *European Journal of Operational Research*

Received date: 15 February 2017
Revised date: 23 February 2018
Accepted date: 11 June 2018

Please cite this article as: Ming-Chien Sung , David C.J. McDonald , Johnnie E.V. Johnson , Chung-Ching Tai , Eng-Tuck Cheah , Improving Prediction Market Forecasts by Detecting and Correcting Possible Over-reaction to Price Movements, *European Journal of Operational Research* (2018), doi: 10.1016/j.ejor.2018.06.024

Highlights

- Behaviour consistent with herding occurs frequently in a prediction market
- Prediction market forecasts are inaccurate after particular price movements
- Errors in prediction market forecasts are of of considerable economic importance
- Methodology developed to correct erroneous prediction market forecasts

Improving Prediction Market Forecasts by Detecting and Correcting Possible Over-reaction to Price Movements

Ming-Chien Sung*[1], David C.J. McDonald[1], Johnnie E.V. Johnson[1], Chung-Ching Tai[2], Eng-Tuck Cheah[1]

[1]*Centre for Risk Research, Southampton Business School, University of Southampton, Southampton, SO17 1BJ, UK.*

[2]*Deartment of Economics, Tunghai Uniiversity, Taichung, Taiwan.*

*Corresponding author. Tel.: +44 23 8059 8974. Fax: +44 23 8059 3844.

Email addresses: m.sung@soton.ac.uk (Ming-Chien Sung), d.mcdonald@soton.ac.uk (David C.J. McDonald), jej@soton.ac.uk (Johnnie E.V. Johnson), chungching.tai@gmail.com (Chung-Ching Tai), jeremy.cheah@soton.ac.uk (Eng-Tuck Cheah)

Abstract

We examine the impact of price trends on the accuracy of forecasts from prediction markets. In particular, we study an electronic betting exchange market and construct independent variables from market price (odds) time series from 6,058 individual markets (a dataset consisting of over 8.4 million price points). Using a conditional logit model, we find that a systematic relationship exists between trends in odds and the accuracy of odds-implied event probabilities; the relationship is consistent with participants over-reacting to price movements. In particular, in different time segments of the market, increasing and decreasing odds lead, respectively, to under- and over-estimation of odds-implied probabilities. We develop a methodology to detect and correct the erroneous forecasts associated with these trends in odds in order to considerably improve the quality of forecasts generated in prediction markets.

Key words: Forecasting, prediction markets, price signals, over-reaction

Improving Prediction Market Forecasts by Detecting and Correcting Possible Over-reaction to Price Movements

Abstract

We examine the impact of price trends on the accuracy of forecasts from prediction markets. In particular, we study an electronic betting exchange market and construct independent variables from market price (odds) time series from 6,058 individual markets (a dataset consisting of over 8.4 million price points). Using a conditional logit model, we find that a systematic relationship exists between trends in odds and the accuracy of odds-implied event probabilities; the relationship is consistent with participants over-reacting to price movements. In particular, in different time segments of the market, increasing and decreasing odds lead, respectively, to under- and over-estimation of odds-implied probabilities. We develop a methodology to detect and correct the erroneous forecasts associated with these trends in odds in order to considerably improve the quality of forecasts generated in prediction markets.

Key words: Forecasting, prediction markets, price signals, over-reaction

Improving Prediction Market Forecasts by Detecting and Correcting Possible Over-reaction to Price Movements

1. Introduction

Prediction markets are vehicles for aggregating information held by a variety of individuals and have been heralded as an effective means of harnessing the 'wisdom of the crowd.' However, the forecasts derived from prediction market prices have not always lived up to expectations. One reason could be that prices which emerge following price movements may not appropriately reflect the information which initiated the price movement. This would arise, for example, if participants in these markets over-react to price signals. We examine prediction markets in which the prices can be interpreted as event probabilities and we aim to develop a methodology to improve the quality of forecasts generated in these markets. To achieve this, we identify systematic relationships between price trends and the accuracy of event probabilities inherent in prices and use these relationships to help correct for any observed inaccuracies.

The earliest prediction markets can be traced back to the wagers on papal selection in Italy in the 16th century, gambling markets on the Parliamentary elections in Britain in the 18th and 19th centuries, and the Presidential betting markets in New York in the 19th and early 20th centuries (Rhode & Strumpf, 2004a, 2008; Snowberg et al., 2008).[1] In an attempt to capitalize on the forecasting potential shown by these early wagering markets, the University of Iowa introduced the first stock exchange style prediction market in 1988 (to forecast results related to the U.S. presidential election). Since then, prediction markets have gained traction as alternative forecasting devices for a wide range of uncertain events, including election outcomes (Berg et al., 2008; Berg et al., 2010), economic aggregates (Gürkaynak & Wolfers, 2007), sports matches (Luckner et al., 2008; Spann & Skiera, 2009), for the success of movie box offices and for the entertainment industry more broadly (Pennock et al., 2001; Gruca et al., 2008).[2] Consequently, a range of companies such as Intel, Hewlett-Packard, Eli Lilly, General Electric, and Google have introduced internal prediction markets to forecast a variety of business activities from the sales of printers to the likelihood of success of new products or the probability of meeting project deadlines (e.g., Chen & Plott, 2002; Hopman, 2007; Cowgill et al., 2009).

[1] Betting markets are a form of prediction market. Snowberg et al. (2008) argue that the distinction is not in terms of their structures, but their functions. Prediction markets provide information externalities that can inform business and policy decision makers, while holders of betting market securities care more about the entertainment and financial effects. Prior to scientific polls, betting odds have long been regarded by society as important sources of forecasts (Rhode and Strumpf, 2004a, 2008).

[2] See Snowberg et al. (2008) for a historical review.

Research suggests that prediction markets can produce accurate forecasts, often beating polls and other statistical forecasting techniques (e.g., Berg et al., 2008; Arnesen & Bergfjord, 2014). Vaughan Williams (2011, p.2) concludes, "the balance of opinion provided by previous research suggests that well designed prediction markets can offer substantial promise as a tool of information aggregation and forecasting".

Prices in prediction markets relate to the likelihood of occurrence of the event in question (e.g., a particular individual becoming president in a forthcoming election). Prices are expressed differently in different types of prediction markets. In the prediction markets which form the focus of this study (betting exchanges), prices are expressed as a number, *z*. Participants can choose the amount or stake they are prepared to risk on the event. Consequently, those who choose to buy (back) at *z*, receive £*z* for a stake of £1(a profit of £ (*z*-1)) if the event occurs and lose their stake (£1) should the event not occur. Those who decide to sell (lay) at *z*, lose £(*z*-1) should the event occur, for each £1 they stake, and they gain a profit of £1 if the event does not occur. Market equilibrium dictates that the expected return to both buyers and sellers at any time is zero (in the absence of transaction costs) and a price of *z*, therefore, suggests that their combined view is that the event has a probability of 1/*z*. The market mechanism offers incentives for individuals to continue to trade until the price represents what they consider to be the probability of the event. The Efficient Market Hypothesis (Fama, 1970) predicts that final prices in these markets should aggregate all the participants' information, leading to efficient prices and, therefore, accurate forecasts of the probability of the event.

However, the most preferred choice in prediction markets in 2016 failed to predict Brexit, Donald Trump's nomination as Republican candidate and his election as president. This led to a flurry of internet articles identifying important issues that may hinder the ability of prediction markets to produce meaningful forecasts (e.g., Kominers, 2016). Clearly, prediction markets, like other forms of market are subject to forces that can lead to inaccuracy.[3] Some factors, if directly observable, can be analyzed and corrected by modifying market mechanisms to achieve higher accuracy.[4] However, there are issues that are difficult to detect but may affect the accuracy of prediction markets. Such issues are often related to traders' behavior, with market manipulations (Wolfers & Leigh, 2002), overconfidence (Wu et al. 2008) and over-reaction to price movements or herding (Schnytzer & Snir, 2008) being the most well-known examples.

We address three fundamental research questions associated with the accuracy of probability forecasts of an event that are derived from prediction market prices. Specifically, under what conditions does a systematic relationship exist between trends in market prices and the accuracy of a

[3] For example, Graefe and Armstrong (2011) found, in a laboratory study, that the forecasting accuracy of prediction markets was inferior to face-to-face meetings in some instances.

[4] For example, the number of traders and their trading volume are found to be positively related to the accuracy of prediction markets (Berg et al., 1997; Gruca et al., 2005; Snowberg et al., 2008). Consequently, a special design known as the market-scoring rule was proposed to tackle the problem of thin markets (Hanson, 2003; 2007, 2009; Abramowicz, 2007).

prediction market's event probability forecasts? To what extent does this reduce the accuracy of prediction market forecasts? What methodology can be employed to account for these systematic relationships, thereby enabling the development of probability forecasts which are more accurate than those derived directly from prediction market prices?

One reason systematic relationships may exist between trends in market prices and event probabilities is over-reaction to price signals. This could arise from the well-known adages amongst market traders such as, 'money talks' and 'follow the money'. These beliefs may lead market participants to trade in line with price movements, believing that they are caused by more informed participants. However, over-reliance on price movements as a guide to trading decisions can lead to prices being driven to a level that no longer reflect new information or knowledge about target events. We refer to this behavior as over-reaction to price signals or herding. Price movements resulting from herding may only reflect the participants' psychological state, and any forecasts relying on them may be inaccurate.

There is a considerable literature addressing over-reaction to price signals in financial markets (Gebka & Wohar, 2013). While no unequivocal definition of herding has emerged (Hsieh, 2013; Goodfellow et al., 2009; Devenow & Welch, 1996), it is generally acknowledged to occur when market participants neglect their own private information and adjust their actions to be more representative of those of other traders. Empirical evidence for the phenomenon from financial markets is inconclusive (Lakonishok et al., 1992; Wermers, 1999; Sias, 2004). Similarly, mixed results have been found in laboratory-based studies, in which both the decisions and the information on which they are based are observable (e.g., Cipriani & Guarino, 2005, 2009; Spiwoks et al., 2008). The common finding from these studies is that participants do herd, but to a lesser extent than theoretical models predict.

The extent to which herding influences market prices has been difficult to confirm using traditional financial market data because of its inherent uncertainty. As Shleifer and Summers (1990, p. 22) note, "price changes may reflect new market information which changes the equilibrium price". Consequently, prices are never entirely derived from the current fundamental information; rather, prices represent the current expectation of future prices. Hence, even if all current fundamental information is fully known, there remains some uncertainty in prices.

Betting market data enable us to overcome this concern because markets for events (e.g., a race) close at a pre-defined end-point, with all traders (referred to as bettors) receiving unambiguous payoffs. Consequently, in these markets, there is a time when all uncertainty in the relation between prices (referred to as 'odds') and fundamental information is resolved (i.e., a winner is declared), thus, allowing us to determine the degree to which the market forecast (the odds-implied probability of the event) accurately predicts the outcome. This resolution of uncertainty is repeated often, with several thousand separate markets per annum. Consequently, we achieve our objective by investigating prices in a liquid electronic exchange market for state contingent claims (a horserace betting market) operated by Betfair, which commenced trading in 2000 and currently has a valuation of over £6,000m ($7,902m).

Betting markets, unlike other forecasting methods that rely on historical data, have the ability to incorporate real-time information and can continuously generate forecasts. This continuous process of market activities can make it difficult to determine when over-reaction to price signals takes place. This is reflected in the disagreement shown in the limited debate concerning the presence of herding in betting markets. A few studies have suggested the presence of herding in betting markets (Law & Peel, 2002; Schnytzer & Snir, 2008), but Shing (2006) concluded that odds changes were mainly caused by genuine information rather than over-reaction. Even if researchers can identify the existence of herding in betting markets, being able to correct its negative impact on the accuracy of forecasts derived from odds, is challenging. For example, Schnytzer and Snir (2008) identified herding by examining whether sudden movements in odds can explain the return on money staked in a betting market. However, the method they employed could not be used to generate a "corrected" forecast for a target event beforehand. Consequently, to help improve the accuracy of prediction market forecasts, we propose a means of detecting systematic relationships between price trends and the accuracy of market-price-implied event probabilities that are consistent with over-reaction to price movements. We also propose a statistical method for fixing the resulting inaccuracy of prediction market forecasts.

The remainder of the paper is organized as follows. In Section 2, we review the literature related to the use of odds to forecast event probabilities and that examining the incidence and impact of herding. In the light of this literature, we discuss the contributions of our approach for both detecting and correcting for inaccuracy in probabilities derived from prices which emerge following price movements. In Section 3, we develop the hypotheses, and outline the data and methods employed to test the hypotheses. We present the results in Section 4. We discuss our findings and the implications learned from our forecasting model in Section 5, and conclude in Section 6.

2. Forecasting using Betting Market Odds and the Impact of Herding

The accuracy of forecasts derived from odds

Betting, as a means of trading on one's ability to make accurate forecasts, has a long history, from betting on foot races and ball games of the North American Indians (Culin, 2012). It is not surprising, therefore, that betting market data have long been used for forecasting. For example, before the rise of scientific polling, newspapers in New York employed betting odds as a means of making election predictions (Rhode & Strumpf, 2004a, 2008) and Wall Street betting odds on U.S. presidential elections from late 19th centuries have been shown to be highly predictive (Rhode & Strumpf, 2004b). Early studies using horserace betting data also found that odds were good approximations to the winning probability (e.g., Griffith 1949, Rosett, 1965; Hoerl & Fallin, 1974).

More recent studies confirm the accuracy of forecasts derived from betting odds (e.g., for reviews see Sauer, 1998; Vaughan Williams, 2005; Hausch & Ziemba, 2008; Stekler et al., 2010; and Vaughan Williams & Siegel, 2013). This is not surprising, since recent years have seen growth in the number and type of bookmaker organizations across the world and the emergence of online fixed odds betting and betting exchanges (see Appendix 1 for a description of different betting markets). Multiple betting operators offering the opportunity to bet on the outcome of the same event, provides a wide and diverse pool of opinion to inform the odds which emerge in these markets.

Accounting for performance-related information in odds

Many studies have examined the degree to which specific pieces of performance-related information are accounted for in odds. The general conclusion is that bettors effectively incorporate most individual pieces of relevant information into odds, either immediately (e.g., Vaughan Williams, 2000; Smith, 2003; Deschamps & Gergaud, 2008) or over time (e.g., Johnson et al, 2010). The general finding is that, compared to other forecasting methods such as statistical models using sports-related input variables, expert tipsters, and lay predictions, betting odds provide the most accurate forecasts (Forrest et al., 2005; Luckner & Weinhardt, 2008; Spann & Skiera, 2009; Štrumbelj, 2014).

However, it has been shown that bettors do not fully account for the interactions between multiple performance-related factors (e.g., various aspects of a player's previous performances: see Sung & Johnson, 2008a, for survey). Consequently, it is possible to develop superior forecasts which combine and capture complex non-linear relationships between odds from parallel markets or odds and a *range* of variables associated with multiple performance-related factors (e.g., Ma et al., 2016; Lessmann et al., 2009, 2010, 2012; Wunderlich & Memmert, 2016; Goddard, 2013; Dixon & Pope, 2004; Spann & Skiera, 2009; Goddard, 2005; Donniger, 2014).

In addition, some persistent systematic behavioral biases in odds have been shown to exist. For example, odds on favorites/longshots under-/over-represent their chances (e.g., Vaughan Williams & Paton, 1998; Schnytzer & Weinberg, 2008; Ottaviani & Sørensen, 2008, 2010; Snowberg & Wolfers, 2010; Berkowitz et al., 2017), the chances of popular soccer teams are underestimated (e.g., Forrest & Simmons, 2008; Franck et al., 2010; Oikonomidis & Johnson, 2011) and bettors anchor their judgements on particular pieces of information (Johnson et al., 2007), are overly influenced by national sentiment (Braun & Kvasnicka, 2013) and over-weight the chance that runs of success of particular contestants will end (e.g., gamblers fallacy: Terrell & Farmer,1996).

The conclusion to emerge from betting market studies, most of which use data from fixed-odds bookmaker and pari-mutuel markets, is that odds account for a significant amount of publicly available information but bettors are subject to some behavioral biases and they do not fully account for "complex, subtle (or non-linear), and possibly changing relationships between variables" (Sung & Johnson, 2008b, p.302).

The improving accuracy of odds

The spread of the internet, has prompted innovation in the betting industry, with internet betting, via online (fixed-odds) bookmakers and betting exchanges, developing rapidly (Jones et al., 2006). This has resulted in a very competitive, dynamic market with over 600 different sports betting web sites operating worldwide (Malaric et al., 2008) and very large increases in global sports betting volumes (Forrest, 2017). These changes have resulted in odds providing more accurate forecasts (Forrest et al., 2005), even accounting for the predictions of sophisticated mathematical models which draw together a variety of performance related variables (e.g., Fitt et al., 2006; Dobson & Goddard, 2017). For example, there is considerable evidence that odds in liquid betting exchanges provide more accurate probability forecasts that are less prone to biases, than those derived from traditional fixed-odds betting markets (Smith et al., 2006, 2009; Franck et al., 2010). Smith and Vaughan Williams (2008) speculate that this may arise because the exchanges, unlike traditional bookmakers, offer the opportunity for skilled traders and insiders to back contestants to lose, thereby providing an additional mechanism for inaccurate odds to be corrected.

Odds react quickly to new information

Online betting markets have facilitated the development of in-play betting, whereby bets can be placed on the result of an event (e.g., a tennis match) whilst it is taking place (Sauer, 2005; Huang et al., 2011; Dumitrescu et al., 2013; Viney, 2015; Dias, 2016). The general consensus of studies comparing the forecasting ability of in-play odds with the predictions of mathematical models based on information which arrives during the course of an event, is that in-play odds react quickly and appropriately to the arrival of new information, such as a goal being scored in soccer or a point won in tennis (e.g., Newton & Aslam, 2009; Easton & Uylangco, 2010; Sauer et al., 2010; McHale & Morton, 2011; Spanias & Knottenbelt, 2012; Dumitrescu et. al., 2013; Croxson & Reade 2014; Viney, 2015). However, there is some evidence that bettors in these markets may over-/under-react to surprise/expected news (Choi & Hui, 2012).

In summary, odds have been demonstrated to provide accurate forecasts, particularly, in dynamic on-line betting exchanges, where odds react quickly to new information. However, previous studies suggest that bettors are subject to behavioral biases and they are not particularly skilled at accounting for complex data. We suspect that rapidly changing odds in betting exchanges may increase the complexity of a bettor's decision, and increased cognitive load has been shown to be associated with the increasing use of heuristics, leading to biased judgements (Kahneman et al., 1982). We believe that these circumstances could lead bettors to become subject to a behavior which has been observed in wider financial markets, namely over-reaction to price signals. We now briefly review the literature on over-reaction to price signals.

Herding in financial markets

Three main drivers for herding behavior have been proposed: information ('rational herding'), psychology ('irrational herding') and events ('event-driven herding') (Demirer et al., 2010; Demirer & Kutan, 2006; Devenow & Welch, 1996). We focus here on the rational view that traders adjust their views in the belief (perhaps mistaken) that other traders are more informed than themselves. The combined activity of many herding traders can result in extraordinary changes in asset values over a short period, with prices overshooting their fundamental values (Yan, 2010), possibly leading to bubbles, crashes and bank runs (Devenow & Welch, 1996). While the consequences of herding are irrational at the aggregate level, herding may be rational at the individual level (Simonsohn & Ariely, 2008). Some theoretical models, for example, have rationalized herding as 'information cascades', where decisions are made sequentially by different agents who each hold their own private information (e.g., Banerjee, 1992; Bikhchandani et al., 1992; Avery & Zemsky, 1998). There is uncertainty over the validity of price signals, so it may be rational for agents to disregard some of their private information when that held by others appears (as revealed by their actions) to conflict with their own. In fact, in Hong and Stein's (1999) model, momentum traders can earn positive profits, provided they trade early enough in the 'momentum cycle'.

While herd behavior has a theoretically sound basis, empirical evidence for the phenomenon from financial markets is inconclusive (Lakonishok et al., 1992; Wermers, 1999; Sias, 2004). Similarly, mixed results have been found in laboratory-based studies, in which both the decisions and the information on which they are based are observable (e.g., Cipriani & Guarino, 2005, 2009; Spiwoks et al., 2008). The common finding from these studies is that participants do herd, but to a lesser extent than theoretical models predict.

As indicated above, the extent to which herding influences market prices has been difficult to confirm using traditional financial market data because of its inherent uncertainty. Prediction markets, unlike other forecasting methods that rely on historical data, have the ability to incorporate real-time information and can continuously generate forecasts. We now examine the limited literature concerning the presence of herding in betting markets.

Herding in betting markets

Experimental studies provide the opportunity to control information flows, thereby offering the opportunity to clearly identify herding behavior. An early laboratory study of pari-mutuel market betting was conducted by Plott et al. (2003). They created markets for abstract states and subjects could buy "tickets" in different markets to win the prizes if a certain state was realized. To control the information possessed by market participants, Plott et al. (2003) endowed their subjects with different types of private information. Herding was observed in these experiments, but it was not shown to exist in all the markets studied. A more recent laboratory experiment employed sequential pari-mutuel betting games (Koessler et al. 2012). Here, bettors had only one chance to choose between two states (A and B). The information was controlled by revealing private signals to every subject before their choices. Koessler et al. (2012) elicited some subjects' beliefs, by asking them to specify their subjective probabilities pertinent to both states. The elicitation process forced the subjects to think seriously about what could be learned from market information and results from experimental economics predict that this design is more likely to induce herding. The results of the experiment confirmed that more herding took place where subjects were asked to declare their subjective beliefs.

Camerer (1998) conducted field experiments in pari-mutuel horse racing markets to examine herding. He tested whether bettors respond to bets from those who apparently hold privileged information. To achieve this, he placed temporary bets of $500, and in later experiments $1,000, on randomly chosen horses and observed whether the sudden changes in odds induced over-reaction to these price signals. He found that such bets did lead to a temporary distortion of odds. However, later in the market he cancelled these bets and found that the odds returned to their expected levels. Even when he placed his bets at smaller racetracks (where his bets had a bigger distorting effect on odds) and on races where those with privileged information were most likely to bet (e.g., in races where the horses had never won a race), he still did not detect any significant herding. There are many reasons why Camerer (1998) may not have observed herding: first, the bets he placed could have been too small relative to the market size to attract bettors' attention; second, most bettors are aware that holders of privileged information are more likely to bet at the end of the betting period to reduce the chance that the odds are further eroded by herding; third, bettors realized that these early bets could be cancelled, so they 'rationally' failed to respond to these early price movements.

The widespread belief that some bettors have access to privileged information makes betting markets a likely place to observe herding, as less informed bettors see significant reductions in odds as a signal that the horse is being backed by more informed individuals. However, very few empirical studies have been conducted to test the view that herding takes place in betting markets. Those studies that have been undertaken found that odds movements and non-monotonic trends in odds are common (Schnytzer & Snir 2008). It has generally been shown that early betting on outcomes that subsequently attract a high degree of betting interest is profitable (Crafts 1985; Schnytzer & Shilony 1995). However, systematic profits can only be obtained by betting at the odds available *before* the odds fall dramatically and are only likely to be secured by those with privileged information.

Law and Peel (2002) examined occasions when genuine privileged information resulted in odds movements in UK fixed-odds bookmaker markets. Odds in these markets are set by the bookmaker (market operator) and bettors know their potential payoffs at the time the bet is placed. Large falls in odds (plunges) might result from betting by informed bettors, from over-reaction to price movements, or both. However, Law and Peel (2002) proposed that herding is the most likely explanation if the odds plunge while the proportion of informed bettors declines through the duration of the market, and they used a measure developed by Shin to estimate the proportion of informed bettors (Shin, 1991, 1992, 1993). They found that dramatic moves in odds resulted in higher/lower returns at starting odds *if* the Shin measure increased/declined at the same time; suggesting that herding did exist.

Schnytzer and Snir (2008) developed a model of cash-constrained, informed bettors in an on-course bookmaker market. If mispricing becomes apparent early in the market, informed bettors bet to take advantage, and herd betting by less informed bettors may ensue, causing a large odds movement. However, there may be occasions when the odds then return to a level which does not reflect all available information; at this point, it was assumed by Schnytzer and Snir (2008) that the informed bettors have no cash remaining to exploit the fact that odds do not fully discount the available information. Consistent with the model, using UK and Australian on-course bookmaker data, they found that positive returns could be made by betting on horses for which there has been a significant early plunge, but a later reversal in odds. However, the set of such horses was very small, so it is unclear whether the results represent a genuine concern for the forecasting accuracy of probabilities derived from prediction markets.

Contributions of the current study

The studies discussed above have focused on the accuracy of odds and the type of information discounted in odds. The conclusion is that odds generally provide reasonably accurate predictions of event probabilities, particularly those emerging from betting exchanges, but odds can be distorted by behavioral biases of bettors and their inability to fully handle information complexity.

However, very few studies have identified over-reaction to odds movements. No studies have examined this phenomenon in a betting exchange and none propose a method for restoring the accuracy of the forecasts where trends in odds result in them over-shooting appropriate levels.

Recent evidence from the on-line betting markets, and particularly betting exchanges, is that odds in these markets react quickly to new information. However, this speed of reaction presents bettors with significantly greater information, potentially increasing the complexity of their betting decision. Increased cognitive load of this sort has been shown to be associated with the increasing use of heuristics, leading to biased judgements (Kahneman et al., 1982). Consequently, we believe that exchange bettors may, as a short cut procedure for handling the information complexity, overly rely on price signals.   This motivates our desire to identify to what extent this takes place and to develop a means of correcting any inaccuracy which may then arise in the final odds which emerge in these markets.

In achieving this, the paper makes the following contributions: To the best of our knowledge, this is the first to exploration of over-reaction to odds movements in a betting exchange. Odds in exchanges are derived entirely from the relative levels of supply and demand, and assets can be bought and sold (e.g., horses can be backed to win or lose), as in most prediction markets. This is not the case for the betting markets examined in previous explorations of herding. In addition, it avoids the difficulty of interpreting falling odds in bookmaker markets (employed in Schnytzer & Snir, 2008; Law & Peel, 2002; Crafts, 1985; Schnytzer & Shilony, 1995) as evidence of over-reaction, since these odds movements may result from bookmakers artificially lowering odds.

Second, we show, contrary to much of the implications of the betting market literature, that the final market prices in betting exchanges *cannot* be relied upon to provide accurate probability estimates. In particular, we identify significant inaccuracies in probability estimates derived from market prices resulting from price movements in over one third of the 1,514 separate markets considered, indicating that this is a common phenomenon. This finding suggests that caution should be exercised by those who use prediction markets for forecasting.

Third, the method we propose for detecting behavior consistent with over-reaction to price movements does not require any behavioral assumptions of the bettors or bookmakers. It, therefore, has the potential to be used in a wide range of market mechanisms. Fourth, we propose a general method to identify behavior consistent with over-reaction to price movements, rather than confining our observations to specific types of price movements, such as dramatic movements in odds (Law & Peel, 2002), or a change in the direction of odds (Schnytzer & Snir, 2008).

Fifth, because participants in betting exchanges can 'back' or 'lay' a contestant to win or lose, respectively, this enables us to assess the different reactions to, and importance of, 'buy' and 'sell' signals. We are able to measure at different stages of the market, the likely impact of odds trends on the accuracy of odds-implied probabilities. Consequently, we are able to show that there is a tendency for these trends to result in inaccurate odds-implied probabilities following both 'sell' and 'buy' signals.

Finally, and perhaps most importantly, the method used in previous betting market studies to detect herding could not be used to generate a "corrected" forecast. By contrast, we propose a method which offers the prospect of restoring the accuracy of the forecasts generated in markets where trends in odds result in them over-shooting appropriate levels. In addition, we demonstrate the economic significance of these corrections, by reporting the results of trading strategies designed to capitalize on the winning probabilities predicted by our model.

3. Data and Methodology

We first describe the data used in this study. Second, we explain our strategy for examining when, and to what extent, trends in odds in certain time segments of the betting market lead to odds which under- or over-estimate appropriate levels of a horse's winning probability. Third, we introduce the CL model we use to forecast winning probabilities. Finally, we explain how we improve the accuracy of betting market probability predictions, by using forecasts generated by our CL model that adjust for the systematic relationships we observe between trends in odds and the accuracy of probabilities derived from the odds which emerge in the market.

Data

We employ data from Betfair, the largest electronic exchange betting market, with horserace betting revenues of £105 million ($160 million) in 2010 (Betfair 2011), which overlaps the period in which our real-time data were collected. Betfair is a very liquid online platform (accounting for 90% of the global exchange market: Croxson & Reade, 2008), allowing allows bettors to place bets against each other. We consider 'win' markets, in which bettors must predict which horse will win (or lose). In this case, the Betfair exchange serves to match the bets of those who believe that a particular horse will win (backers) and those who believe it will lose (layers). Backers may bet any amount at the prevailing 'back odds' displayed on the platform (i.e., provided there are sufficient bets of layers to match their bet) and/or they may place orders to bet at higher 'back odds', should these become available later in the market. The liquidity associated with the 'back odds' for a particular horse is provided by those who believe that it will lose. These latter individuals offer to 'lay the horse' at the prevailing 'odds' (or they may place orders to lay the horse at lower lay odds, should those who wish to back the horse accept these odds later in the market). The exchange serves to match the bets of backers and layers.

We denote the best odds available in the exchange at a particular moment in time to back horse i to win race j to be $\overline{R}_{ij}$. This represents the return to a £1 winning 'back bet' (e.g., a winning £1 bet with odds of 3.00 returns £3 for a profit of £2).[5] As is typical of exchanges, Betfair generally take a commission of 5% on net winnings[6]. Consequently, the effective odds, which we use in our analysis, are given by $R_{ij} = 1 + 0.95(\overline{R}_{ij} - 1)$. At any moment in time throughout the market, R_{ij} represent the combined view of both backers and layers of the chance of horse i winning race j. In particular, their combined view can be translated into an odds-implied probability q_{ij} of horse i winning race j, with n_j runners, as follows:

(1)
$$q_{ij} = \frac{1/R_{ij}}{\sum_{i=1}^{n_j} 1/R_{ij}}.$$

While the horses' true winning probabilities are not knowable explicitly, each race j results in a vector of outcomes $(y_{1j}, y_{2j}, ..., y_{n_j j})^T$, where y_{hj} = 1 for the winning horse h and y_{ij} = 0 otherwise. If markets fully incorporate available information, then, over many races, odds-implied probabilities should approximate true winning probabilities as determined by race results.

[5] Exchange odds are expressed inclusive of unit stake and are often referred to as 'decimal' odds; this is in contrast to bookmaker markets where odds are expressed as, say, 2/1 for the equivalent of exchange decimal odds of 3.00.

[6] At the time the data used in this study was collected, the commission structure on Betfair was more complicated than this, with a lower base commission rate applied to high volume bettors, and an additional charge applied to consistent winners. Thus, our assumption of 5% commission on average is an estimate (and the true average commission rate is of little consequence to the results).

The data employed are sequences of odds available in the Betfair electronic betting exchange for 62,124 horses running in 6,058 races in the UK and Ireland from August 2009 through August 2010. We collected data in real-time from the Betfair electronic exchange, using the Betfair API. In particular, we collected the odds available on each runner at 1 minute intervals from 9:00 a.m. each morning, through to the start time of the race for most UK and Irish horseraces run during this period. This resulted in over 8.4 million data points. The odds employed are the best odds at which it is possible to back the horse to win throughout the duration of each market[7]. The race results were collected from the website of the Racing Post newspaper prior to conducting the analyses below.

Identifying over-betting on the basis of price signals

In order to answer our first two fundamental research questions (i.e., when and to what extent odds movements lead to odds which do not accurately reflect a horse's winning probability), we must first devise a means of identifying when these situations occur. In betting exchanges the demand for bets on a particular outcome is directly represented in the odds. Thus, odds movements which over-shoot appropriate levels will arise if a unidirectional odds movement occurs which is greater than is merited by the likelihood of the event occurring. However, two important questions remain unanswered: When will this occur? Does this phenomenon have a symmetric impact on markets for both directions of odds movements?

It is well established that bets placed in the later stages of betting markets are more informative than early bets (Asch et al., 1982; Gandar et al., 2001)[8]. One explanation is that the timing of bets is variably incentivized depending on the quality of bettors' information (Ottaviani & Sørensen, 2005). More informed bettors have an incentive to bet late because liquidity is higher in the later stages of an exchange market and bid-ask spreads are narrower. Consequently, they bet late so they can place sufficiently large bets to compensate for revealing information.

These incentives for informed bettors are well known. Consequently, bettors may place too much faith in odds movements believed to be signals of informed trading, leading them to push odds to levels that do not reflect the horse's winning probability. This is likely to be more commonplace in the later stages of the market when it is assumed by the betting public that informed traders are most likely to bet. In addition, there may be insufficient time for more informed bettors to bet in a manner to capitalize on these inappropriate odds. However, if (unexpectedly), bettors place too much faith in odds movements in the early stages of the market and push odds to inappropriate levels, there would be sufficient time for more informed bettors to bet in a manner to push these odds to a level that appropriately represent horses' winning chances. Consequently, if over-estimation of the information

[7] Instead of 'back' odds we could have used 'lay' odds or the mid-point of 'back' and 'lay' odds. Similarly, we make a minor assumption that odds are equally valid whatever the stake limit. Neither of these considerations has more than a minor effect on our analysis.

[8] While not reported here, we have verified that this is also the case in our exchange market data.

value of odds movements leads to market odds that do not adequately reflect event probabilities, this phenomenon is more likely to occur in the later (cf. earlier) stages of the market These considerations motivate our first hypothesis, H1: *Bettors are more likely to over-estimate the value of 'sell' and buy' signals in the later (cf. earlier) stages of the market.*

In betting exchanges, as in other financial markets, 'buy' or 'sell' signals may be interpreted differently. In particular, betting exchanges facilitate the laying ('selling') of 'known losers': (i.e. offering other bettors the 'opportunity' of backing horses that the 'layer' knows will be deliberately prevented from running to their potential (Marginson, 2010). This practice could benefit horse owners who know that their horse will lose. Despite rules that forbid such behavior, its prevalence is the subject of much debate, suggesting that bettors might be more likely to interpret 'sell' signals as genuine informed trading. Consequently, if bettors over-estimate the information value of odds movements, we would suspect that this is more likely following sell (cf. buy) signals. This motivates our second hypothesis H2: *Inaccurate odds-implied probabilities consistent with bettors over-estimating the value of odds movements, are more likely to occur on 'sell' (lay) than 'buy' (back) signals*.

In order to investigate the timing and impact of trends in odds in the betting market leading to odds which do not accurately represent a horse's chance of winning, we first segment the market on each race into three time periods, which depend on the time left before the race start. This allows us to determine the prevalence and direction of odds movements over different periods. While markets are often active from the evening before the race, or earlier for the most popular events, the vast majority of betting takes place on the day of the race, so segment 1 begins at 9:00 am on the day of the race and ends at the race start time. The most active stage of the market begins 30 minutes before the race start, since this is the typical time between races at each racetrack in the UK. This is the period when the parallel track bookmaker and pari-mutuel markets operate and when most participants direct their attention to the race. We divide this period into two equal segments: segment 2 begins 30 minutes before the race start and ends 15 minutes before the race start; segment 3 begins 15 minutes before the race start and ends at the race start time. Hence, segment 1 lasts at least 4 hours, depending on the race start time and, segments 2 and 3 last 15 minutes.

Next, we generate an *odds curve* for each horse in each segment of the market for a given race, using the method of Johnson et al. (2006). That is, for each horse i in race j, and for each market segment k, we have a sequence S_{ijk} of L_{ijk} pairs of times $t_{ijk}(l)$ and odds $R_{ijk}(l)$, i.e. $S_{ijk} = \{[t_{ijk}(1), R_{ijk}(1)], ..., [t_{ijk}(L_{ijk}), R_{ijk}(L_{ijk})]\}$. We record odds changes so that, for each time in the sequence, the odds are different from the preceding time. Consequently, for any time T, where $t(l) \le T \le t(l+1)$, $R = R(l)$ (here, and in the following, we drop the subscripts i, j and k when their use is not required). The first/last pair is the first/last time in the segment along with the first/final odds recorded. The final odds recorded in segment 1 (the full duration of the market) are the odds at which the horse started the race (or 'starting odds'); this special case is used to calculate the final odds-implied probability, which is given by $q_{ij}(L_{ij}) = [1/R_{ij}(L_{ij})]/\sum_{s=1}^{n_j}[1/R_{sj}(L_{sj})]$. Finally, we rescale all the sequences so that $t(1)=0$, $t(L)=1$, and $R(L)=1$. The result of this procedure is that each odds curve is a piecewise continuous step function $\Phi(t)$ on the interval [0, 1], such that $\Phi(1)=1$. From the odds curve, we measure underlying trends in the odds as illustrated in Figure 1. Specifically, the trend μ is estimated as the slope of the ordinary least squares regression line fitted to the pairs in *S*, constrained to pass through (1, 1), i.e.,

(2) $$Y(t) = 1 + (t-1)\mu,$$

and is, therefore, given by

(3) $$\mu = \frac{\sum_{l=1}^{L}[R(l)-1][t(l)-1]}{\sum_{l=1}^{L}[t(l)-1]^2}.$$

A trend variable is estimated for each horse in each race for each of the three segments. Further, because bettors might infer differing information from 'lay' and 'back' bets, increasing or decreasing odds may be interpreted differently. Consequently, we derive two trend variables, $\mu^+ = \max(\mu,0)$ and $\mu^- = |\min(\mu,0)|$, for each horse in each segment (i.e., six trend variables for each horse in each race). Hence, for horse *i* in race *j*, and for market segment *k*, ${\mu_{ijk}}^+ = \max(\mu_{ijk},0)$ and ${\mu_{ijk}}^- = |\min(\mu_{ijk},0)|$. Descriptive statistics of the data and trend variables are provided in Table 1.

The conditional logit model

The CL model (McFadden 1974) has been employed in many betting market studies (Asch et al., 1984; Bolton & Chapman 1986; Benter 1994; Figlewski 1979; Johnson et al., 2010). It allows us to estimate the winning probability of each horse, taking into account competition between horses in the race. Formulation of the CL model begins with an estimate of horse *i*'s ability to win race *j*,

(4) $$W_{ij} = \sum_{l=1}^{m}\beta(l)x_{ij}(l) + \varepsilon_{ij},$$

where $\beta(l)$, for l = 1, … , m, are the coefficients that determine the importance of the variables $x_{ij}(l)$. If the independent errors ε_{ij} are identically distributed according to the double exponential distribution, the estimated winning probability for horse h, p_{hj}, is given by

(5) $$p_{hj} = \Pr(W_{hj} > W_{ij}, i = 1,2,...,n_j, i \neq h) = \frac{\exp[\sum_{l=1}^{m} \beta(l) x_{hj}(l)]}{\sum_{i=1}^{n_j} \exp[\sum_{l=1}^{m} \beta(l) x_{ij}(l)]}.$$

The coefficients $\beta(l)$ are estimated by maximizing the joint probability of observing all the race results in the dataset; this is achieved by maximizing the log-likelihood (LL) of the full model (i.e., one including all independent variables in which we are interested):

(6) $$\ln L(full) = \sum_{j=1}^{N} \sum_{i=1}^{n_j} y_{ij} \ln p_{ij},$$

where y_{ij} = 1 if horse i won race j and y_{ij} = 0 otherwise, and N is the total number of races in the dataset. For this study, an appropriate measure of the predictive accuracy of the model is Maddala's (1983)[9] pseudo-R^2, given by

(7) $$R^2 = 1 - \exp\{(2/N)[\ln L(naive) - \ln L(full)]\},$$

where $\ln L(naive)$ is the LL of the naive model (where each horse in a race is assigned the same probability of winning):

(8) $$\ln L(naive) = \sum_{j=1}^{N} \ln(1/n_j).$$

The standard normal statistic $z(l) = \beta(l)/S.E.[\beta(l)]$ is used to test if variable coefficients are significantly different from 0, i.e., variables add predictive power to the model. An additional test to justify augmenting simpler models with additional variables utilizes the likelihood ratio (LR) test statistic $2[\ln L(full) - \ln L(naive)]$, which is χ^2 distributed with degrees of freedom equal to the number of additional variables.

The first variable in our CL models is the logged final odds-implied probability, i.e. $x_{ij}(1) = \ln[q_{ij}(L_{ij})]$. If the estimated value of the coefficient of this variable, $\beta(1)$, is equal to one when there are no other variables in the model, this implies that there is no favorite-longshot bias (FLB: the widely reported phenomenon whereby favorites/longshots are under-/over-bet (e.g., Gramm and Owens 2006)). The greater the value of $\beta(1)$, the greater is the degree of the FLB (Bacon-Shone et al., 1992). Previous studies have indicated that betting exchanges display little or no FLB (Smith et al., 2006), suggesting that $\beta(1) = 1$. Whatever its value, having developed a model incorporating an

[9] We use Maddala's pseudo-R^2 (rather than McFadden's (1974) more popular definition) because McFadden's R^2 has the unfortunate property of varying with the average number of horses in each race, which is not the case for Maddala's pseudo-R^2. Our results would be the same using the McFadden pseudo-R^2, but if we used this measure we could not account for variations in forecasting accuracy due to the differing numbers of runners.

appropriate value of $\beta(1)$ (i.e., having adjusted for any FLB), the pseudo-R^2 of a single-variable CL model is an appropriate measure of the predictive accuracy of market odds.

The second variable in our CL model captures potential trends in odds. Over-reaction to odds movements occurs when bettors alter their actions to be more representative of the actions of others. Many bettors behaving in this manner can lead to significant odds movements that cannot be accounted for by the underlying objective information of a handful of informed bettors. Hence, when a horse's odds change significantly (from an imbalance of bets on that horse to win or lose), and that odds movement is not fully attributable to an equivalent change in that horse's probability of winning, then the horse's final odds-implied probability will differ from its true winning probability, i.e., the final odds will not reflect accurately its chance of winning. Therefore, in order to identify this is likely to have taken place, we look for occasions when significant odds movements resulted in odds-implied probabilities deviating from realized winning probabilities.

To assess when this is the case, we estimate six CL models, each having two variables: log of final odds-implied probability and one of the two trend variables (+/-) for each market segment. These trends were defined to be either increasing or decreasing. Consequently, the trend variables for a given horse in a given race within a time segment were either zero or had a positive value (e.g., for segment 1, when $\mu_{ij1}{}^{+}$ was positive, $\mu_{ij1}{}^{-}$ took the value zero, and vice versa). Higher values for the trend variables imply steeper odds changes. If odds movements are systematically associated with the accuracy of the resulting odds-implied probabilities, such that these odds appear to have been pushed too far in the same direction as the odds movement, we believe that this is likely to have arisen from over-reaction to odds movements. In this case, the coefficients of the trend variables should be significantly different to zero (corresponding LR tests should also be significant) and their sign should reflect the fact that a trend of increasing odds leads to odds which *under*-estimate a horse's winning probability and a trend of decreasing odds leads to odds which *over*-estimate a horse's winning probability: i.e., horses whose odds-implied probabilities decrease/increase over time win more/less often than implied by the odds. Determination of the extent to which bettors over-bet in the different time periods when odds increase or decrease allows us to test our first and second hypotheses. We preferred to estimate six separate models including the various trend variables (rather than one model incorporating all the variables) since this enabled us to show clearly the additional information content of each of these back and lay trend variables (i.e. the improved log-likelihood of the models including these individual variables over that simply including odds); thus enabling us to demonstrate the relative degree to which lay and back odds signals may have led to over-betting in the different segments. In addition, we did not include the trend variables for segments 2 or 3 with the trend variables for segment 1 in the same models as the time periods for segments 2 and 3 overlap with that for segment 1. Consequently, the trend variables in these periods may have been correlated and this could have led to misleading results.

We split the dataset into a training set of the first 75% of races (4,544: 47,196 horses), and a holdout set of 25% of races (1,514: 14,928 horses). We estimate CL models on the training set, in order to determine whether bettors over-bet certain horses. We use these models to predict horses' winning probabilities in the holdout set and construct betting strategies for the holdout set of races (see below) based on these probabilities to test the accuracy of these forecasts compared to those directly derived from market odds. Hence, our conclusions about the accuracy of final odds probabilities, which are based only on the holdout set, can be relied upon, because they are out-of-sample and, thus, are not influenced by fitting our models on the training set.

Assessing forecast accuracy with betting strategies

In assessing the degree to which the CL model we develop produces accurate winning probability forecasts, a direct and intuitive approach might be to compare the forecasts delivered by our model with the actual results. By undertaking such a comparison over a large enough sample, we can conclude with a measure of deviation, how accurate in general are our model forecasts. However, what we observe are realizations of uncertain events in binary form (win/lose), while our model produces probabilities. A direct comparison between these two might not always be the most reasonable evaluation. However, as indicated above, we can determine the LL of the model's predictions. This enables us to assess the predictive accuracy of the model using Maddala's (1983) pseudo-R^2. A LR test comparing two alternative forecasting models provides us with a means of determining which model most accurately predicts winning probabilities (i.e. a CL model simply incorporating odds-implied probabilities and a CL model incorporating odds implied probabilities together with the trend in odds variables we develop from the training data).

We are also interested to know if the predictions of the model incorporating trend in odds variables offer an *economic advantage* over predictions simply based on odds-implied probabilities. In particular, it is important to test the effectiveness of the model's outputs against the wins of horses in races in the holdout period. This will help to confirm if a real economic advantage accrues when betting in real-time, appropriately using knowledge of the degree to which bettors may over-bet horses whose odds follow a trend. To achieve this, we can determine whether it is possible to secure profits when betting using the winning probability forecasts from a CL model incorporating odds implied probabilities together with the trend in odds variables. The logic behind this is that if the betting market is free from over-betting on horses whose odds follow a trend, the final odds-implied probability should be the unbiased estimator of the true winning probability. On the contrary, the final odds-implied probability will deviate from the true value if over-betting occurs. If such deviation exists, there will be the opportunity to gain profits by betting horses whose odds under-estimate, and laying those horses that over-estimate, their true probability. That is, the odds which emerge in these markets following movements in odds, produce significantly inaccurate forecasts if the probabilities which they represent deviate sufficiently from the true winning probabilities, such that profitable opportunities arise. We test, therefore, whether bets based on the probability predictions of a CL model incorporating odds implied probabilities, together with the trend in odds variables that we found were significant in segments 2 and 3 (μ_{ij2}^{-} and μ_{ij3}^{+}, respectively) (estimated using the training data), can yield profits in the holdout races, where profits are directly related to which horses win the races. In particular, we estimate Model 7 (see Table 2), using the training sample data. We then fix the coefficients of the odds implied probability variable together with the segment 2 and segment 3 trend in odds variables at the values determined by this estimation (i.e. 1.026, -0.442 and 0.165, respectively; as shown in Table 2). We then use this model to estimate the winning probabilities for the holdout sample of races. We employ these probabilities as the basis of a range of betting strategies to test hypothesis H3: *Odds changes consistent with bettors over-estimating the value of 'buy' and 'sell' signals, lead to considerable mispricing in this prediction market.*

We aim to show that Model 7 enables us to generate probability estimates which can be used in betting strategies applied to the holdout set of races to make significantly higher profits than could be obtained from probability estimates generated by a benchmark CL model simply incorporating an odds implied probability variable. This will suggest that incorporating variables associated with trends in odds in segments 2 and 3 (μ_{ij2}^{-} and μ_{ij3}^{+}, respectively) alongside an odds implied probability variable in a CL model can produce probability estimates which are significantly more accurate than those which can be derived simply from the final odds which emerge in these markets. Consequently, this will indicate that the incorporation of odds trend variables alongside an odds implied probability variable in a CL model will provide a means of adjusting the probabilities derived from prediction markets to correct for errors caused by odds movements consistent with bettors over-estimating the value of 'buy' and 'sell' signals.

In order to proceed with this analysis, we develop a range of betting strategies on the holdout set of races to test whether such profitable opportunities exist. If bettors' actions lead to odds-implied probabilities that are sufficiently out of line with true winning probabilities, it should be possible to find profitable betting opportunities. If over-betting results in odds that do not reflect all available information, betting strategies based on the model developed using the training dataset should be profitable when applied to the holdout races and involve relatively low risk. Consequently, considering each holdout race j in turn, with initial wealth £1000 and current wealth W_j, we use the estimated probabilities from the model as the basis for the following betting strategies:

1. *Level staking*: For each horse i, if $p_i > 1/R_i$, bet 1% of current wealth on horse i. Therefore, if a bet is to be placed, the size of the bet is £W_j / 100.
2. *Proportional staking*: For each horse i, if $p_i > 1/R_i$, bet an amount such that the profit from a win, after commission, is 10% of W_j, i.e., bet size is £$W_j / 10(R_i - 1)$. The advantage of this

betting strategy is that returns are not unduly influenced by 'lucky' wins on horses with very high odds (Schnytzer and Snir, 2008).

3. *Kelly staking*: The Kelly (1956) strategy assigns bet sizes x_i over all n horses in the race to maximize the log of expected wealth after the race, $G(x_1, x_2, ..., x_n) = \sum_{i=1}^{n} p_i \ln F_i$, where $F_i = 1 + 0.95(x_i \overline{R}_i - \sum_{s=1}^{n} x_s)$ if $x_i \overline{R}_i > \sum_{s=1}^{n} x_s$, and $F_i = 1 + x_i \overline{R}_i - \sum_{s=1}^{n} x_s$ otherwise (since 5% commission is only paid if bets result in an overall profit). The x_i are estimated using numerical optimization. The Kelly strategy is optimal in that it maximizes the asymptotic rate of growth of wealth and minimizes the expected time to reach a pre-defined wealth target (Breiman 1961). However, since recommended bets may be very large, the volatility of returns from a full Kelly strategy over the 1,514 holdout races may not lead to a positive overall return.
4. *Half Kelly staking*: Some authors (e.g., Benter 1994) recommend a fractional Kelly strategy, whereby bet sizes are a fixed proportion (in this case, half) of those recommended by the full Kelly strategy. This is sub-optimal, as it no longer maximizes the asymptotic growth rate of wealth. However, fractional Kelly strategies are less risky, and, over medium-length time horizons, may result in a higher expected return on turnover (MacLean et al., 2010).

For a Kelly betting strategy, the order of the races in the holdout set is inconsequential to the result (Johnstone 2011) and the above strategies all entail a zero probability of ruin, assuming that arbitrarily small bets can be placed. A non-intuitive property of the Kelly strategy is that it may recommend bets with negative expected returns (to reduce risk). However, this is only optimal if our estimates for the true winning probabilities, p_i, are accurate. If this is not the case, over-betting will occur (MacLean et al., 1992). Consequently, since our estimates may not be completely accurate, we do not place bets on horses for which the expected return is negative, i.e., $p_i R_i < 1$ (Hausch et al., 1981). Second, the Kelly betting strategies might recommend very large bets on horses with a high probability of winning, so a single unfortunate loss may skew the overall returns. Similarly, skewed returns may result from a fortunate win on a horse with a low winning probability. We, therefore, restrict single bet sizes to a maximum of 10% of current wealth.

We assess the performance of the betting strategies using the following measures:

1. *Rate of return*: the ratio of the profit (or loss) achieved to the total amount bet.
2. *Risk-adjusted return*: the return per unit risk (Edelman 2000), given by $R/[Var(R)]^{1/2}$, where R is the rate of return and its variance (Var (R)) is estimated using a bootstrap procedure, by sampling with replacement from the holdout set 1,000 times and calculating returns on each sample.

3. *Expected final wealth*: $W_0 \prod_{j=1}^{N} X_j$, where N is the number of races bet and, for race j, X is the expected increase in wealth factor $\sum_{i=1}^{n} p_i \ln \tilde{F}_i$. Here, $\tilde{F}_i = 1 + 0.95(\tilde{x}_i \overline{R}_i - \sum_{s=1}^{n} \tilde{x}_s)$ if $\tilde{x}_i \overline{R}_i > \sum_{s=1}^{n} \tilde{x}_s$, and $\tilde{F}_i = 1 + \tilde{x}_i \overline{R}_i - \sum_{s=1}^{n} \tilde{x}_s$ otherwise, where $\tilde{x}_i$ is the fraction of current wealth bet on horse i after any restrictions are imposed (MacLean et al., 1992).
4. *Probability that final wealth is above x% of initial wealth*: this is given by $1 - \Phi\left\{\dfrac{(1/N)\ln(x/100) - E[\ln X_j]}{\sigma[\ln X_j]}\right\}$, where Φ is the standard normal cumulative distribution function (MacLean, et al., 1992).

Should our results suggest that it is possible to secure profits using betting strategies based on estimates derived from our combined CL model (i.e. including variables related to trends in odds and the odds-implied probability) this will have two important implications: First, trends in odds in certain time intervals have led to final odds-implied probabilities that are out of line with true winning probabilities. Second, by incorporating the variables related to trends in odds in our CL model we have a means of correcting for some of the inaccuracies caused by bettors’ apparent tendency to over-estimate the value of both ‘sell’ and ‘buy’ signals.

4. Results

Revising winning probabilities in light of over-reaction to odds signals

The results of estimating models (using the training set of 4,544 races) including trends in odds in the three time segments of the market (separated by whether odds increase or decrease) using the training data are presented in Table 2. The coefficient of log of final odds-implied probability in Model 0 is significantly different from zero (z = 51.45, p = 0.0000). In addition, the model’s LL is -8337.6, confirming that the odds, as expected, add significant predictive power over the naive model (LL = -10,307.7).

Models 1 to 6 include a second variable that describes a trend in odds over time. Note that our CL model is used to predict a horse’s probability of winning. If a trend variable plays a significant role in this model, it suggests that the horse’s probability should be revised in response to that trend. In Models 1 and 2, we incorporate trends in odds across the full duration of the market. The coefficient of the trend in odds variable in these models is not significantly different from zero (z = 0.92, p = 0.3576 and z = -0.13, p = 0.8966, respectively). These results suggest that odds movements over the full duration of the market do not necessarily result in odds-implied probabilities differing from true winning probabilities, i.e., over-betting is not apparent when considering the full duration of the market.

However, considering the period between 30 and 15 minutes from the race start (Models 3 and 4), the coefficient of the trend in odds variable is significant when odds decrease (Model 4: z = -3.56, p = 0.000), i.e., bettors over-react to decreasing odds, but not increasing odds (Model 3: z = 0.91, p = 0.3682). To be exact, Model 4 suggests a decrease in predicted winning probability when bettors appear to over-estimate the value of 'buy' signals; leading to them over-betting horses whose odds decrease. We observe the opposite effect in the last 15 minutes (Model 5 and 6), the coefficients of the second variable are significant when odds increase (Model 5: z = 2.14, p = 0.0324), but not when odds decrease (Model 6: z = -0.74, p = 0.4592). Model 5 suggests an increase in predicted winning probability when bettors appear to over-estimate the value of 'sell' signals. Therefore, odds movements in the last 15 minutes of the market do result in odds-implied probabilities differing from true-winning probabilities, but only when odds increase, i.e., bettors appear to under-bet horses whose odds increase in this late stage..

These results are all supported by LR tests vs. Model 0. In particular, only Models 4 and 5 add significant predictive power over the model simply incorporating odds (Model 0) (Model 4: χ_1^2 = 14.27, p = 0.0002; Model 5: χ_1^2 = 4.54, p = 0.0331). These findings were supported by the results of a 10-fold cross validation, in which only Models 4 and 5 add predictive power over Model 0 (Model 4: χ_1^2 = 10.85, p = 0.0010; Model 5: χ_1^2 = 4.66, p = 0.0308). Consequently, it is only in the periods between 30 and 15 minutes and particularly in the last 15 minutes before the race start that we observe inaccurate odds-implied probabilities that appear to arise from bettors over-estimating the value of 'buy' and 'sell' signals. These results support H1, that bettors are more likely to over-estimate the value of 'sell' and buy' signals in the later (cf. earlier) stages of the market.

However, we do not find evidence to support our second hypothesis that inaccurate odds-implied probabilities, caused by bettors over-estimating the value of odds movements are more likely to occur on 'sell' (lay) than 'buy' (back) signals. In particular, bettors appear to over-estimate the information value of 'sell' signals in the last 15 minutes (the most active betting period) (Model 5), but also on 'buy' signals between 30 and 15 minutes from to the race start (Model 4). As indicated above, these results were confirmed when performing a 10-fold cross-validation.

Economic significance of correcting for odds which appear to arise from over-estimation of the 'buy' and 'sell' signals

We estimate Model 7 using the training data of 4,544 races and evaluated it on the holdout set of 1,514 races. This model includes two variables to account for the under-estimation of odds we observed following increasing odds in the last 15 minutes, and the over-estimation of odds we observed on decreasing odds in the 30 to 15 minute period prior to the race start (μ_{ij3}^{+} and μ_{ij2}^{-}), together with an odds-implied probability variable. The results of estimating Model 7 using the training data are presented in the last row of Table 2. We also conducted a 10-fold cross validation and this confirmed that Model 7 added significant predictive power over odds alone (Model 7: $\chi_2^2 = 13.16$, $p = 0.0014$).

In betting market studies, to demonstrate the effect size of the phenomena observed, it is normal to show what additional profits could be earned from a model incorporating both an odds variable and the particular variables in question, over a benchmark model, simply incorporating odds (e.g., Sung et al, 2016; Lessmann et al., 2012): We follow this approach to identify the strength of the combined effect of μij3+ and μij2- in predicting winning probabilities over that from simply employing the odds variable (lnqij(Lij)). To achieve this, we estimate Model 7 (see Table 2), using the training sample data. The coefficients of the odds implied probability variable together with the segment 2 and segment 3 trend in odds variables are then fixed at the values determined by this estimation (i.e. 1.026, -0.442 and 0.165, respectively; as shown in Table 2). We then use this model to estimate the winning probabilities for the holdout sample of races and develop betting strategies to exploit any mispricing. The results of this analysis are presented in Table 3 and Figure 2.

These results show that a strategy of betting against the trend in odds is profitable for all the betting strategies we employ, even when accounting for transaction costs. However, the level stakes strategy (rate of return: 5.20%) and the proportional stakes strategy (6.49%) spend a significant portion of the holdout period betting at a loss relative to initial capital (see Figure 2 for cumulative wealth for each strategy). On the other hand, the full Kelly (6.16%) and half Kelly (10.39%) strategies rarely drop below initial capital. The greatest monetary accumulation is achieved with the full Kelly strategy, with initial capital increasing by over 126%. However, it is also the riskier of the two Kelly strategies, with 20.5 times the initial capital bet over the course of the holdout period (cf. just 9.1 times for the half Kelly strategy). For the half Kelly strategy, initial capital increases by over 94%. Consequently, the risk-adjusted return is greatest for the half Kelly strategy, with a value of 0.90. Similarly, the full Kelly strategy has the highest expected final wealth and the highest probability of doubling wealth (0.49) but also the lowest probability of retaining at least half of initial wealth (0.76). In summary, positive returns are identified for the various betting strategies, including a sizeable return of 10.39% from our preferred strategy (half Kelly). These results provide strong support for H3, that odds changes consistent with bettors over-estimating the value of buy and sell signals, lead to considerable mispricing in this prediction market.

In order to demonstrate the value of the approach we adopt for correcting for the possibility that bettors over-estimate the value of ‘buy’ and ‘sell’ signals in the market, we compare the results presented in Table 3 and Figure 2 with those from a betting strategy based on the winning probabilities forecast using a benchmark CL model simply incorporating the odds variable (Model 0). This benchmark model produces no recommended bets in the holdout set and the improvement in returns from the various betting strategies over the benchmark are, therefore, those presented in Table 3 and Figure 2. These results indicate that the incorporation of odds trend variables alongside an odds implied probability variable in a CL model (Model 7) provides a means of adjusting the probabilities derived from prediction markets to correct for errors caused by odds movements consistent with bettors over-estimating the value of ‘buy’ and ‘sell’ signals.

To confirm this conclusion, we also compare the returns shown in Table 3 and in Figure 2, with the returns that would have been achieved had we simply bet randomly on the holdout races. Specifically, we find that had we bet £1 on each horse in the 1514 holdout races, this would have resulted in 14928 bets and a loss of £724.59 (-4.85%) and if we had bet sufficient on each horse in the holdout races to win £10, this would have led to a loss of £335.59 (-1.61%). We also developed a strategy of betting £1 on a random horse in each race in the holdout sample and this strategy was repeated in 1000 simulations. This resulted in a loss of 9.72%. Similarly, we developed a strategy where of betting sufficient on a randomly selected horse in each holdout race to win £10, and repeated this strategy in 1000 simulations. This resulted in a loss 4.09%.

Clearly, all of the random betting strategies result in significant losses (ranging from -1.61% to -9.72%) when applied to the holdout races. By contrast, all of the betting strategies shown in Table 2, employing winning probability forecasts which correct for errors caused by odds movements consistent with bettors over-estimating the value of ‘buy’ and ‘sell’ signals (i.e. using Model 7), lead to positive returns in excess of 5%.

Taken together, these betting simulation results indicate that the CL model incorporating variables to account for odds changes consistent with bettors over-estimating the value of buy and sell signals (Model 7) can generate forecasts that are significantly more accurate than the final odds-implied probabilities observed in the prediction market; thus Model 7 offers a means of adjusting the probabilities derived from the final odds to correct for errors caused by odds movements. In addition, the size of the profits achievable using this model suggest that the corrections are economically significant.

5. Insights into Trends in Odds and Forecast Accuracy

In response to the three key questions proposed in the introduction, we begin our discussions by summarizing the facts that our statistical method reveal about the *timing*, *degree*, and *direction* of behavior consistent with bettors misevaluating ‘buy’ and ‘sell’ signals. Based on these facts, we explain why misevaluation of odds signals may occur and how it affects the prediction market’s forecast accuracy. We then discuss why we need a device that can detect such behavior and correct for the error in market forecasts, and how our model meets this need.

Observation 1: The degree of odds movements in later stages is negatively correlated with forecast accuracy.

Our results suggest that under certain conditions, odds movements which we believe may be caused by over–reaction to odds movements, have a detrimental effect on the forecast accuracy of prediction markets. In particular, our modeling of these odds movements shows that the larger the odds movement, the greater the disparities between final odds-implied and true winning probabilities. However, this behavior only becomes significant in the later stages of the market. Odds changes over the full duration of the market do not generally lead to inaccuracies in final market odds. This is not unexpected, since there is a lengthy period during which market odds that do not appropriately reflect a horses' winning probabilities, can be corrected. Moreover, much of the information pertaining to horses' chances is revealed on the day of the race. For example, information concerning jockey changes, and horses' condition and behavior may not be revealed until the market on a race has opened (Bruce and Johnson 1995). Therefore, odds *are expected* to change before the final stages of the market (resulting from revealed fundamental information). As a result, any over-estimation of the value of information associated with odds movements is unlikely to take place because of early stage market odds changes. Consequently, it appears that odds that do not appropriately reflect horses' winning probabilities, are more likely to occur when (i) there is little time remaining to correct any inaccuracy caused by over-estimation of the information value of odds changes, and (ii) when bettors perceive odds movements as evidence of trading by those with privileged information.[10]

Observation 2: Over-betting on decreasing odds tends to take place 15 to 30 minutes before a race starts.

Our finding that, in the later stages of the market, the *degree* of odds movement depends on its *direction*, serves to support our prediction that less informed bettors' perceptions of the actions of informed bettors are key to the prevalence of behaviour consistent with them over-estimating the information content of odds movements. In particular, while previous studies of herding in betting markets (Law and Peel 2002; Schnytzer and Snir 2008) have focused on bookmaker markets, where bettors may only back their preferred horse to win (leading to a reduction in its odds), our study examines a betting exchange, where bettors may also lay horses to lose (leading to an increase in their odds). There is little qualitative difference between backing/laying a horse one thinks will win/lose. Consequently, the differences we observe in odds movements are likely, we believe, to be due to differences in the bettors' perceptions of 'buy' and 'sell' signals. In particular, we argue that this stems from their belief that those bettors with privileged information will trade at different times, depending upon whether they believe a horse will win or lose a race.

[10] This conclusion chimes well with classic cases of herding in regular financial markets, such as that evidenced in the South Sea Bubble (Dale et al., 2005).

We find that there is no over-betting on decreasing odds in the last 15 minutes of the market, suggesting that bettors do not over-value these odds signals.[11] It implies that the average bettor does not consider a late 'plunge' to be a signal containing valuable information, or bettors realize that by the time the plunge has happened, the information is assimilated in the odds (previous studies have found that bets on 'plungers' are not profitable once the odds change has occurred: Crafts 1985; Bird and McCrae 1987). An alternative explanation may be that bets placed in the last 15 minutes by informed bettors cancel out the bets of those who over-estimate earlier movements in odds.

However, we find over-betting on plunges that occur earlier in the active betting (in the period between 30 minutes and 15 minutes before race starts). In this case, the odds change to such an extent that further odds movements, which happen in the last 15 minutes of the market, are insufficient to ensure that odds reflect all available information. This might be explained by: (i) cash-constrained informed bettors bet early, but they may not have the funds to correct odds for a second time should odds fail to account for available information (Schnytzer and Snir 2008), or (ii) bets placed in the 30 to 15 minute market segment are generally those of less informed bettors (who might be more likely to over-value the information content of odds movements), since more informed bettors benefit from placing their bets later so as not to divulge their own information (Ottaviani and Sørensen 2005). In either case, it appears that uninformed bettors perceive that odds that decline in the period 30 to 15 minutes before the race start result from the actions of informed bettors (presumably believing that any fundamental information would have been discounted in odds in the earlier stages of the market). It has been found in empirical studies of financial markets that less informed bettors are able to detect informed trading via the volume and direction of the informed trades (e.g., Meulbroek 1992). Hence, a tendency to over-value 'buy' and 'sell' signals may or may not occur depending on (i) uninformed bettors' *perception* of the degree of influence over market odds held by bettors with privileged information, and (ii) informed bettors' *actual* degree of influence. Consequently, the over-valuing of 'buy' and 'sell' signals will occur only if less informed bettors believe that market odds movements are currently reflecting the opinions of more informed bettors. On the other hand, the extent of over-valuing the information content of odds movements will be reduced if informed bettors have sufficient market power to restore odds to the point where they reflect all available information.

Observation 3: Under-estimation of event probabilities following increasing odds tends to take place in the last 15 minutes prior to races.

The perceptions of uninformed bettors also appear to play a part when considering whether bettors over-estimate of the information value of increasing odds. Increases in odds in the last 15 minutes prior to market close often lead to situations where the odds are too high (i.e., the horse is relatively under-valued). This suggests that bettors may over-estimate the negative signals concerning

[11] This finding is consistent with the literature on herding in financial markets, which has found little conclusive evidence that investors display herd behavior.

a horse's chance of winning following increases in odds (by laying unfavored horses) even in the late stages of the market. It seems, therefore, that 'sell' signals are treated differently to 'buy' signals. A 'sell' signal is taken seriously even late in the market. This may arise because bettors perceive that it is more likely that individuals with access to privileged information (e.g., horse owners) lay horses to lose (rather than back them to win), since it is easier for them to predict (and/or influence) that their horse will lose (Marginson 2010). It may be perceived that they are more likely to do this later in the market when positive information concerning the prospects of other runners has been fully discounted in odds. Indeed, it is not even necessary that this practice of laying known losers is prevalent, provided bettors perceive that it is.

The advantages of the conditional logit model

Having acknowledged the three observations outlined above, an intuitive forecasting strategy might be to discard final odds-implied probabilities but use, say, opening odds to develop forecasts. While this strategy can largely bypass the detrimental effects of over-estimation of the information content of trend in odds, it also forgoes a large part of the valuable information assimilated in market odds. A more favorable forecasting strategy is the one we propose in this study, a contingent method that can diagnose the *timing*, *degree* and *direction* of likely over-estimation of the information value of a trend in odds and counter with a revised forecast.

We demonstrate that considerable disparity between true winning probabilities and those arrived at by the prediction market can occur and that this represents a large and economically significant that the inaccuracy. We speculate that the inaccuracy in prediction market forecasts we observe may arise from bettors over-estimating the information value of a trend in odds. In particular, we find that odds often over- and under-estimate winning probabilities following increasing and decreasing trend in odds, respectively, and that trading strategies can be constructed that show consistent positive returns from betting in a manner which capitalises on these systematic relationships. Such a betting strategy is based on a model that accounts for likely differences in less informed bettors' perceptions of the actions of informed bettors at different times in the market. In fact, we find that a half Kelly strategy with some restrictions provides a substantial rate of return (10.39%) over the holdout sample; the return is sufficiently large to compensate for potential variation in returns and the model risk involved. We find that if we employ probability forecasts from a CL model simply incorporating odds, no bets are recommended in the holdout races. The difference in returns obtainable from this benchmark model and our model, which helps to account for the systematic relationships we observe between trends in odds and the accuracy of odds-implied probabilities, are substantial. Clearly, it may be possible to develop a model which accounts for other, perhaps more subtle effects of these systematic relationships, and this may earn even higher returns.

However, our results certainly demonstrate that some trends in odds are associated, in a systematic manner, with inaccuracy in probabilities derived from odds in prediction markets.

Previous studies of over-betting on the basis of apparent information in odds movements (Law and Peel 2002; Schnytzer and Snir 2008) have demonstrated that positive returns can be made by betting against the trend, but these approaches offer very few betting opportunities. Importantly, the approach adopted in these studies does not offer the means of correcting probabilities derived from prediction markets for the errors caused by trends in odds. By contrast, our results show that it is possible to develop a profitable strategy betting ‘against the trend’, and that this strategy provides a significant number of betting opportunities (betting in over 33 percent of markets).

Our results run counter to the expectation that further movement in the direction of trends in odds in certain time segments move odds closer to an appropriate level to reflect a horse’s chance of winning. Rather, we find that increasing odds in the last 15 mins (segment 3) is a *positive* indicator, whereas decreasing odds in the period 30 minutes to 15 minutes from the race start (segment 2) is a *negative* indicator, of the horse’s chance of winning. Consequently, in segment 2 when odds decrease, this is a sign that the odds over-estimate the horse’s chance of winning and in segment 3 when odds increase, this is a sign that the odds under-estimate the horse’s chance of winning. We find that the relationships between odds trends in these time segments and the acuracy of odds- implied probabilities are consistent with bettors reacting to ‘buy’ and ‘sell’ signals and driving odds to unrealistically low and high levels, respectively. Consequently, the winning probability is best estimated by correcting for the apparent over-betting on price signals which occurs in these segments. It is difficult to identify a cause for the systematic relationships we observe, other than over-reaction to price movements, but the establishment of the reason for this clearly requires further research.

Our results clearly demonstrate a systematic relationship between odds movements and the accuracy of probability estimates derived directly from the final market odds. We show that this relationship is likely to have damaging effects on the accuracy of probabilities derived from prediction market odds and that these effects have considerable economic importance. In addition, we demonstrate a means of adjusting the probabilities derived from prediction markets to correct for errors caused by these inappropriate odds movements. Despite the significant improvements in prediction market probabilities which we show can arise from the procedure demonstrated here, it is important to note that the model that we develop is not the ultimate model. Identification of different and/or more subtle aspects of inaccuracy of odds-implied probabilities associated with trends in odds could also be incorporated in a predictive model, and this may help improve even further the probability forecasts obtained from prediction markets.

6. Conclusion

This study is the first to examine the impact of on prediction market forecasts of odds movements in an electronic prediction market. We propose a conditional logit model to detect inaccurate odds-implied probabilities arising from what we believe to be bettors over-estimating the information content of trends in odds. We demonstrate that this model can also be used to adjust market odds appropriately to produce forecasts that are more accurate. We find that over-betting occurs on odds signals, but only under certain conditions. In particular, we believe our results suggest that the impact of over-estimation of both 'buy' and 'sell' signals is concentrated in the later, more active stages of the market, and that this over-estimation affects odds-implied probabilities differently at different times in the market. However, whilst our results are consistent with the view that bettors over-react to odds movements, our analysis does not prove that this is the case. Further research is certainly needed to categorically identify the reasons for the systematic relationship we observe between trends in odds and the accuracy of the winning probabilities which can be derived from the final odds which emerge in prediction markets.

Our results suggest that inaccurate odds following odds movements can often occur in prediction markets. We find that the apparent tendency to over-estimate the information content of odds moves adversely affects the accuracy of odds in prediction markets and that this tendency is most notable in periods when participants have little time to correct the inaccuracy. One of the implications for those who rely on the forecasts developed from prediction markets is that these markets should always be allowed sufficient time for the more informed bettors to correct odds that do not discount all available information.

Clearly, our results suggest that caution needs to be exercised in using prediction market forecasts. However, our results also provide encouragement for those who rely on the forecasts emerging from prediction markets. In particular, provided sufficient data concerning the outcomes of previous prediction markets are available, the methodology presented here can be used to correct the final odds for any tendency to over-estimate the information value of odds movements during the later stages of the market. Consequently, the predictability of the relationship between odds movements and the accuracy of odds-implied probabilities in some prediction markets might be used to improve considerably the forecasts that emerge from these markets.

Finally, our work contributes new evidence to the literature on over-reaction to odds movements, where results of empirical studies have been inconsistent. At the aggregate level, over-reaction to odds movements can result in odds failing to discount all available information, particularly when the market has insufficient time to correct the resulting mispricing, or when informed bettors are not participating actively. We demonstrate that the systematic relationship we observe between odds movements and the accuracy of odds-implied probabilities, offers the prospect of abnormal returns for those who seek to capitalize on the apparent over-betting of others. Most importantly, we find that these systematic relationships, which we speculate may arise from less-informed bettors over-estimating the information value associated with odds movements, occur frequently, with a significant deviation between odds-implied probabilities and true winning probabilities in over one third of the markets that we examine.

Appendix 1. A Description of UK Betting Markets

The 'Tote' is the UK's only pari-mutuel market. Bets can be placed online, in betting offices around the UK, as well as on-course at racetracks. In this market structure, bets are collated in a 'pool' covering all the horses in a race. After the race, the pool is divided between all the bettors who placed bets on the winning horse, proportionally to the amount that they wagered, and net of a fixed proportional track take of 13.5%. However, the Tote has only a marginal presence in UK horserace betting. The two major types of betting market in the UK are bookmakers and exchanges, accounting for 94% of horserace betting turnover (over £5.7 billion) in the year to March 2010 (Gambling Commission 2010).

In bookmaker markets, bets are placed at fixed odds set by the bookmaker. The bettor must accept the odds currently offered by the bookmaker or the unknown starting odds (calculated later). Bookmakers have the highest operating costs (e.g., maintaining an estate of betting offices) so their margins are typically higher than exchange and pari-mutuel markets. Bets can be placed at racetracks ('on-course' market) as well at betting offices around the UK or online ('off-course' market). Odds are distinct in the on-course and off-course markets until 10 minutes before the race starts, at which point the two markets converge.

A betting exchange is an online platform that allows bettors to back horses to win or to lay them to lose. Bets are only matched when two or more bets of the appropriate stake and odds are made, with the exchange automatically pairing backers and layers to settle bets. Exchanges typically have lower margins than bookmakers. For more information on exchanges, see Smith and Vaughan Williams (2008).

References

Abramowicz, M. 2007. The hidden beauty of the quadratic market scoring rule: A uniform liquidity market maker, with variations. *Journal of Prediction Markets* **1** 111-125.

Arnesen, S., O. Bergfjord. 2014. Prediction markets vs polls- An examination of the accuracy for the 2008 and 2012 elections. *Journal of Prediction Markets* **8** 24-33.

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

Asch, P., B. G. Malkiel, R. E. Quandt. 1984. Market efficiency in racetrack betting. *The Journal of Business* **57** 165-175.

Avery, C. and Zemsky, P. 1998. Multidimensional uncertainty and herd behavior in financial markets. American Economic Review, 88, 724 – 748.

Bacon-Shone, J., V. S. Y. Lo, K. Busche. 1992. Modelling winning probability. *Research Report 10,* Department of Statistics, University of Hong Kong.

Banerjee, A.V. 1992. A simple model of herd behavior. *The Quarterly Journal of Economics* **107** 797-817.Benter, W. 1994. Computer based horse race handicapping and wagering systems: A report. D. B. Hausch, V. S. Y. Lo, W. T. Ziemba, eds. *Efficiency of Racetrack Betting Market.* Academic Press, London.

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

Berg, J. E., R. Forsythe, T.A. Rietz. 1997. What makes markets predict well? Evidence from the Iowa Electronic Markets. Albers, W., W. Güth, P. Hammerstein, B. Moldovanu, E. Damme, eds. *Understanding Strategic Interaction: Essays in Honor of Reinhard Selten*, Springer, New York, 444-463.

Berg, J., J. Geweke, T. Rietz. 2010. Memoirs of an indifferent trader: Estimating forecast distributions from prediction markets. *Quantitative Economics* **1** 163-186.

Berg, J., F. Nelson, T. Rietz. 2008. Prediction market accuracy in the long run. *International Journal of Forecasting* **24** 285-300.

Berkowitz, J.P., Depken, C.A., II, Gandar, J.M. 2017. A favorite-longshot bias in fixed-odds betting markets: Evidence from college basketball and college football. *Quarterly Review of Economics and Finance*, 63, 233-239.

Betfair. 2011, Annual Report. Press release, June 29. Accessed August 2011. http://corporate.betfair.com/investorrelations/~/media/Files/B/Betfair/pdf/annual-Report-2011.pdf.

Bikhchandani, S., Hirshleifer, D, and Welch, I. (1992). Theory of fads, fashion, custom, and structural change as informational cascades. Journal of Political Economy, 100, 992 – 1026.

Bird, R., M. McCrae. 1987. Tests of the efficiency of racetrack betting using bookmaker odds. *Management Science* **33** 1552-1562.

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

Braun, S. and Kvasnicka, M. 2013. National sentiment and economic behavior: evidence from online betting on European football. *Journal of Sports Economics,* 14(1), 45-64.

Breiman, L. 1961. Optimal gambling systems for favorable games. *Proceedings 4th Berkeley Symposium on Mathematical Statistics and Probability* **1** 65-78.

Bruce, A. C., J. E. V. Johnson. 1995. Costing excitement in leisure betting. *Leisure Studies* **14** 48-63.

Camerer, C. F. 1998. Can asset markets be manipulated? A field experiment with racetrack betting. *Journal of Political Economy* **106** 457-482.

Chen K.Y., C. R. Plott. 2002. Information aggregation mechanisms: Concept, design and implementation for a sales forecasting problem. CalTech, Division of Humantities and Social Sciences, Working paper 1131.

Choi, D. Hui, S.K. 2012. The role of surprise: Understanding over- and under-reactions using in-play soccer betting. *Mimeo*, Hong Kong University of Science and Technology and Stern School of Business, New York University.

Cipriani, M. and Guarino, A. 2005. Herd behavior in a laboratory financial market. *The American Economic Review*, 95, 1427-43.

Cipriani, M. and Guarino, A. 2009. Herd behavior in financial markets: An experiment with financial market professionals. *Journal of the European Economic Association*, 7(1), 206 – 233.

Cowgill, B., J. Wolfers, J., E. Zitzewitz. 2009. Using prediction markets to track information flows: Evidence from Google, Working paper. http://faculty.haas.berkeley.edu/bo_cowgill/GooglePredictionMarketPaper.pdf

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

Croxson, K., & Reade, J. 2008. *Information and efficiency: Goal arrival in soccer betting*. Working Paper, Oxford University.

Croxson, K., Reade, J.J. 2014, Information and efficiency: Goal arrival in soccer betting. *Economic Journal*, 124, 62-91.

Culin, S. 2012. *Games of the North American Indians*, Dover Publications. (Original work published 1907).

Dale, R. S., J. E. V. Johnson, L. Tang. 2005. Financial markets can go mad: Evidence of irrational behaviour during the South Sea Bubble. *The Economic History Review* **58** 233-271.

Demirer, R. and Kutan, A.M. 2006. Does herding behaviour exist in Chinese stock markets? *Journal of International Financial Markets, Institutions and Money*, 16, 123 – 142.

Demirer, R., Kutan, A.M. and Chen, C.-D. 2010. Do investors herd in emerging stock markets? Evidence from the Taiwanese market. *Journal of Economic Behavior and Organization*, 76, 283 – 295.

Deschamps, B., Gergaud, O. 2008. Efficiency in horserace betting markets. The role of the professional tipster. Hausch, D.B., Ziemba, W.T. eds. *Handbook of Sports and Lottery Markets.* North Holland (Elsevier).

Devenow, A. and Welch, I. 1996. Rational herding in financial economics. *European Economic Review*, 40, 603 – 615.

Dias, A.C. 2016. *Live Betting Markets Efficiency: The NBA Case*. Master thesis, Instituto Universitário de Lisboa.

Dixon, M.J. Pope, P.F. 2004.The value of statistical forecasts in the UK association football betting market. *International Journal of Forecasting*, 21(2), 331-340.

Dobson, S. Goddard, J. 2017. Evaluating probabilities for a football in-play betting market. Rodriguez, P., Humphreys, B.R. Simmons, R. eds. *The Economics of Sports Betting.* Edward Elgar, Cheltenham.

Donninger, C. 2014. Towards the perfect football betting bot: A first preliminary report of Zoccer (August 14, 2014). Available at SSRN: https://ssrn.com/abstract=2498910 or http://dx.doi.org/10.2139/ssrn.2498910

Dumitrescu, G., Huang, X., Knottenbelt, W.J., Spanias, D., Wozniak, J. 2013. Inferring the score of a tennis match from in-play betting exchange markets. Proc. 4th Mathematics in Sport Conference, Leuven, Belgium, June 2013.

Easton, S., Uylangco, K. 2010. Forecasting outcomes in tennis matches using within-match betting markets. *International Journal of Forecasting*, 26(3), 564-575.

Edelman, D. 2000. On the financial value of information. *Annals of Operations Research* **100** 123-132.

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

Figlewski, S. 1979. Subjective information and market efficiency in a betting market. *Journal of Political Economy* **87** 75-88.

Fitt, A.D., Howls, C.J., Kabelka, M. 2006. Valuation of soccer spread bets. *Journal of the Operational Research Society*, 57, 975-985.

Forrest, D. 2017. Sports corruption and developments in betting markets. Rodriguez, P., Humphreys, B.R. Simmons, R. eds. *The Economics of Sports Betting.* Edward Elgar, Cheltenham.

Forrest, D., Goddard, J., Simmons, R. 2005. Odds-setters as forecasters: The case of English football. *International Journal of Forecasting*, 21(3), 551–564.

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

Franck, E., Verbeek, E., Nüesch, S. 2010. Prediction accuracy of different market structures — bookmakers versus a betting exchange. *International Journal of Forecasting*, 26(3), 448-459.Gambling Commission. 2010. Industry Statistics 2009/10. Accessed February 2011. http:www.gamblingcommission.gov.uk/pdf/Gambling Industry Statistics 2009 2010 WEB – January 2011.pdf.

Gandar, J. M., R. A. Zuber, R. S. Johnson. 2001. Searching for the favourite-longshot bias down under: An examination of the New Zealand pari-mutuel betting market. *Applied Economics* **33** 1621-1629.

Gebka, B. and Wohar, M.E. 2013. International herding: Does it differ across sectors? *Journal of International Financial Markets, Institutions and Money*, 23, 55 – 84.

Goddard, J. 2005. Regression models for forecasting goals and match results in association football. *International Journal of Forecasting*, 21(2), 331-340.

Goddard, J. 2013. The efficiency of soccer betting markets. Vaughan Williams, L., Siegel, D.S. eds.*The Oxford Handbook of the Economics of Gambling*. Oxford University Press.

Goodfellow, C., Bohl, M. T. and Gebka, B. 2009. Together we invest? Individual and institutional investors' trading behaviour in Poland. *International Review of Financial Analysis*, 18, 212 – 221.

Graefe, A., J. Armstrong. 2011. Comparing face-to-face meetings, nominal groups, Delphi and prediction markets on an estimation task. *International Journal of Forecasting* **27** 183–195.

Gramm, M., D. H. Owens. 2006. Efficiency in pari-mutuel betting markets across wagering pools in the simulcast era. *Southern Economic Journal* **72** 926-937.

Griffith, R.M. 1949. Odds adjustment by American horse race bettors. *American Journal of Psychology*, 62(2), 290-294.

Gruca, T. S., J. E. Berg, M. Cipriano. 2005. Consensus and differences of opinion in electronic prediction markets. *Electronic Markets* **15** 13-22.

Gruca, T. S., J. E. Berg, J. E., M. Cipriano. 2008. Incentive and accuracy issues in movie prediction markets. *Journal of Prediction Markets* **2** 29-43.

Gürkaynak, R. S., J. Wolfers. 2007. Macroeconomic derivatives: An initial analysis of market-based macro forecasts, uncertainty, and risk. Frankel, J. A., C. A. Pissarides, eds. *NBER International Seminar on Macroeconomics 2005*. MIT Press, Cambridge, MA 11-50.

Hanson, R. 2003. Combinatorial information market design. *Information Systems Frontiers* **5** 107–119.

Hanson, R. 2007. Logarithmic market scoring rules for modular combinatorial information aggregation. *Journal of Prediction Markets* **1** 1–15.

Hanson, R. 2009. On market maker functions. *Journal of Prediction Markets* **3** 61–63.

Hausch, D.B., Ziemba, W.T. (Eds.). 2008. *Handbook of Sports and Lottery Markets*. North Holland (Elsevier).

Hausch, D. B., W. T. Ziemba, M. Rubinstein. 1981. Efficiency of the market for racetrack betting. *Management Science* **27** 1435-1452.

Hoerl, A.E., Fallin, H.K. 1974. Reliability of subjective evaluations in a high incentive situation. *Journal of the Royal Statistical Society. Series A (General)*, 137(2), 227-230.

Hong, H. and Stein, J.C. (1999). A unified theory of underreaction, momentum trading, and overreaction in asset markets. *The Journal of Finance*, 54, 2143-84.

Hopman, J. W. 2007. Using forecasting markets to manage demand risk. *Intel Technology Journal* **11** 127-135.

Hsieh, S.-F. 2013. Individual and institutional herding and the impact on stock returns: Evidence from Taiwan stock market. *International Review of Financial Analysis*, 29, 175 – 188.

Huang, X., Knottenbelt, W., Bradley, J. 2011. Inferring tennis match progress from in-play betting odds. (Final year project), Imperial College London, South Kensington Campus, London, SW7 2AZ.

Johnson, J. E. V., O. Jones, L. Tang. 2006. Exploring decision makers' use of price information in a speculative market. *Management Science* **52** 897-908.

Johnson, J. E. V., R. O'Brien, M. Sung, M. 2010. Assessing bettors' ability to process dynamic information: policy implications. *Southern Economic Journal* **76** 906-931.

Johnson, J.E.V., Schnytzer, A., Liu, S. 2007. To what extent to investors in a financial market anchor their judgments excessively? Evidence form the Hong Kong horserace betting market. *Journal of Behavioral Decision Making* 22(4) 410-434.

Johnstone, D. J. 2011. Economic interpretation of probabilities estimated by maximum likelihood or score. *Management Science* **57** 308-314.

Jones, P., Turner, D., Hillier, D., Comfort, D.2006. New business models, and the regulatory state: A retail case study of betting exchanges. *Innovative Marketing*, 3, 112-119.

Kahneman, D., Slovic, P. and Tversky, A. 1982. *Judgement under Uncertainty: Heuristics and Biases*, New York: Cambridge University Press.

Kelly, J. L. 1956. A new interpretation of information rate. *Bell System Technical Journal* **35** 917-926.

Koessler, F. Noussair, C., Ziegelmeyer, A. 2012. Information aggregation and belief elicitation in experimental parimutuel betting markets. *Journal of Economic Behavior & Organization*, 83 (2), 195-208

Kominers, J. D. 2016. https://www.bloomberg.com/view/articles/2016-11-15/prediction-markets-didn-t-call-trump-s-win-either

Lakonishok, J., Shleifer, A. and Vishny, R.W. 1992. The impact of institutional trading on stock prices. *Journal of Financial Economics*, 32, 23-43.

Law, D., D. A. Peel. 2002. Insider trading, herding behaviour and market plungers in the British horse-race betting market. *Economica* **69** 327-338.

Lessmann, S., Sung, M., Johnson, J. E. V. 2010. Alternative methods of predicting competitive events: An application in horserace betting markets. *International Journal of Forecasting, 26* (3), 518–536

Lessmann, S., Sung, M.-C., Johnson, J. E. V. 2009. Identifying winners of competitive events: A SVM-based classification model for horserace prediction. *European Journal of Operational Research,* 196 (2), 569–577.

Lessmann, S., Sung, M., Johnson, J. E. V., Ma, T. 2012. A new methodology for generating and combining statistical forecasting models to enhance competitive event prediction *European Journal of Operational Research,* 218 (1), 163-174.

Luckner, S., J. Schröder., C. Slamka. 2008. On the forecast accuracy of sports prediction markets. H. Gimpel, N. R. Jennings, G. E. Kersten, A. Ockenfels, C. Weinhardt, eds. *Negotiation, Auctions, and Market Engineering*, Springer Berlin Heidelberg 227-234.

Luckner, S., Weinhardt, C. 2008. Arbitrage opportunities of market-making traders in prediction markets. *IEEE Joint Conference on E-Commerce Technology (CEC'08) and Enterprise Computing, E-Commerce and E-Services (EEE'08)*, 21-24 July, Washington, DC.

Ma, T., Tang, L. McGroarty, F., Sung, M. Johnson, J.E.V. 2016. Time is money: Costing the impact of duration misperception in market prices. *European Journal of Operational Research*, 255(2), 397-410.

MacLean, L. C., E. O. Thorp, Y. Zhao, W. T. Ziemba. 2010. Medium term simulations of the full Kelly and fractional Kelly strategies investment. L. C. MacLean, E. O. Thorp, W.T. Ziemba, eds. *The Kelly Capital Growth Investment Criterion: Theory and Practice*. World Scientific, Singapore 543-561.

MacLean, L. C., W. T. Ziemba, G. Blazenko. 1992. Growth versus security in dynamic investment analysis. *Management Science* **38** 1562-1585.

Maddala, G. S. 1983. *Limited Dependent and Qualitative Variables in Econometrics*. Cambridge: Cambridge University Press.

Malarić, R., Katić, T., Sabolić, D. 2008. The market efficiency of the soccer fixed odds internet betting market, *Journal Applied Economics Letters, 15 (3), 171-174.*

Marginson, D. 2010. Market *in*efficiency in person-to-person betting: Examining 'known loser' insider trading on the exchanges. *Southern Economic Journal* **76**: 932-952.

McFadden, D. 1974. Conditional logit analysis of qualitative choice behavior. P. Zarembka, eds. *Frontiers in Econometrics*. Academic Press, New York 105-142.

McHale, I., Morton, A. (2011). A Bradley-Terry type model for forecasting tennis match results. *International Journal of Forecasting*, 27(2), 619-630.

Meulbroek, L. K. 1992. An empirical analysis of illegal insider trading. *The Journal of Finance* **47** 1661-1699.

Newton, P. K. & Aslam, K., (2009) Monte Carlo Tennis: A Stochastic Markov Chain Model. *Journal of Quantitative Analysis in Sports*, 4(3), pp. 1-42

Oikonomidis A., Johnson J.E.V. 2011. Who can beat the odds? The case of football betting reviewed. L. Vaughan Williams, ed. *Prediction Markets: Theory and Applications*, Routledge, 2011, 202-220.

Ottaviani, M., P. N. Sørensen. 2005. The timing of bets and the favorite-longshot bias. *Mimeo*, London Business School and University of Copenhagen.

Ottaviani, M., Sørensen, P.N. 2008. The favorite-longshot bias: An overview of the main explanations. D.B. Hausch, W.T. Ziemba, eds. *Handbook of Sports and Lottery Markets*. North-Holland, Amsterdam, 83-101.

Ottaviani, M., Sørensen, P.N. 2010. Noise, information, and the favorite-longshot bias in parimutuel predictions. *American Economic Journal: Microeconomics*, 2 (1), 58-85.

Pennock, D. M., S. Lawrence, C.L. Giles, F. A. Nielsen. 2001. The real power of artificial markets. *Science*, **291**(5506) 987-988.

Plott, C.R., Wit, J. Yang, W.C. 2003. Parimutuel betting markets as information aggregation devices: experimental results. *Economic Theory*, 22(2), 311- 351.

Rhode, P.W., K. S. Strumpf. 2004a. Historical presidential betting markets. *Journal of Economic Perspectives* **18**(2) 127-142.

Rhode, P., Strumpf, K. 2004b. Historical prediction markets: Wagering on presidential elections. Working paper.

Rhode, P.W., K. S. Strumpf. 2008. Historical political futures markets: An international perspective. National Bureau of Economic Research Working paper 14377.

Rosett, R.N. 1965. Gambling and rationality. *Journal of Political Economy*, 73(6), 595-607.

Sauer, R.D. 1998. The economics of wagering markets. *Journal of Economic Literature*, 36(4), 2021-2064.

Sauer, R.D. 2005. The state of research on markets for sports betting and suggested future directions. *Journal of Economics and Finance*, 29(3), 416-426.

Sauer, R.D., Waller, J.K., Hakes, J.K.2010. The progress of the betting in a baseball game. *Public Choice*, 142, 297-313

Schnytzer, A., Weinberg, G. 2008. Testing for home team and favorite biases in the Australian rules football fixed-odds and point spread betting markets. *Journal of Sports Economics*, 9(2), 173-190.

Schnytzer, A., Y. Shilony. 1995. Inside information in a betting market. *The Economic Journal* **105** 963-971.

Schnytzer, A., A. Snir. 2008. Herding in imperfect betting markets with insider traders. *The Journal of Gambling Business and Economics* **2** 1-15.

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 traders in a market for state-contingent claims. *The Economic Journal* **103** 1141-1153.

Shing, H. F. 2006. *The Economics of Betting Markets*. Unpublished Ph.D. thesis, University of London.

Shleifer, A., L.H. Summers. 1990. The noise trader approach to finance. *The Journal of Economic Perspectives* **4** 19-33.

Sias, R. W. 2004. Institutional herding. *The Review of Financial Studies* **17** 165-206.

Simonsohn, U., and Ariely, D. 2008. When rational sellers face nonrational buyers: Evidence from herding on eBay. *Management Science*, 54, 1624-37.

Smith, M.A. 2003. The impact of tipster information on bookmakers' prices in UK horserace markets. Vaughan Williams, L. ed. *The Economics of Gambling*, Routledge, London.

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

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

Smith, M.A., Vaughan Williams, L. 2008. Betting exchanges: A technological revolution in sports betting. Hausch, D.B., Ziemba, W.T. eds. *Handbook of Sports and Lottery Markets*. North Holland (Elsevier).

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

Snowberg, E., J. Wolfers, E. Zitzewitz. 2008. Prediction markets: From politics to business (and back). W. T. Ziemba, D. Hausch, eds. *Handbooks in Finance: Handbook of Sports and Lottery Markets*, Elsevier 385-402.

Spanias, D. & Knottenbelt, W., (2012) Predicting the outcomes of tennis matches using a low-level point model. *IMA Journal of Management Mathematics*.

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

Spiwoks, M., Bizer, K., Hein, O. 2008. Informational cascades,: A mirage? *Journal of Economic Behavior and Organisation*, 67(1), 193-199.

Stekler, H.O., Sendor, D., Verlander, R. 2010. Issues in sports forecasting. *International Journal of Forecasting*, 26(3), 606-621.

Štrumbelj, E. 2014. On determining probability forecasts from betting odds. *International Journal of Forecasting*, 30(4), 934-943.

Sung, M.-C., Johnson, J.E.V. 2008a. Revealing weak-form inefficiency in a market for state contingent claims: The importance of market ecology, *Economica* 77, 128-147.

Sung, M.-C., Johnson, J.E.V. 2008b. Semi-strong form information efficiency in horse race betting markets. D.B. Hausch, W.T. Ziemba, G. Constantinides, H.M Markowitz, R.C. Merton, S.C. Myers, P.A. Samuelson, W.F. Sharpe, K.J. Arrow, eds. *Handbook of Sports and Lottery Markets*. North Holland, Amsterdam, 275-306.

Sung, M., McDonald, D.C.J., Johnson, J.E.V. 2016. Probabilistic forecasting with discrete choice models: Evaluating predictions with pseudo-coefficients of determination. *European Journal of Operational Research*, 248, 1021-1030.

Terrell, D., Farmer, A. 1996. Optimal betting and efficiency in parimutuel betting markets with information costs. *Economic Journal*, 106(437), 846-868.

Vaughan Williams, L. 2000. Can forecasters forecast successfully? Evidence from UK betting markets. *Journal of Forecasting*, 19(6), 505-514.

Vaughan Williams, L. 2011. Introduction. L. Vaughan Williams, eds. *Prediction Markets: Theory and Applications*. Routledge Oxford 1-2.

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

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

Vaughan Williams, L., Siegel, D.S. (Eds.).2013.*The Oxford Handbook of the Economics of Gambling*. Oxford University Press.

Viney, M. A., 2015. Prediction of in-play tennis. PhD thesis, School of Mathematical and Geospatial Science, RMIT University: https://researchbank.rmit.edu.au/eserv/rmit:161474/Viney.pdf

Wermers, R. (1999). Mutual fund herding and the impact on stock prices. *The Journal of Finance*, 54, 581-622.

Wolfers, J., A. Leigh. 2002. Three tools for forecasting Federal elections: Lessons from 2001. *Australian Journal of Political Science* **37** 223-240.

Wu, S., M. Sung, J.E.V. Johnson. 2008. Overconfidence in judgements: The evidence, the implications and the limitations. *Journal of Prediction Markets*, **2**, 2008, 73-90.

Wunderlich, F., and Memmert, D. 2016. Analysis of the predictive qualities of betting odds and FIFA World Ranking: Evidence from the 2006, 2010 and 2014 football World Cups. *Journal of Sports Sciences*, 34 (24), 2176-2184.

Yan, H. (2010). Is noise trading cancelled out by aggregation? *Management Science*, 56, 1047-59.

Figure 1. Example of the least squares regression method for determining the trend of the odds curve for a horse.

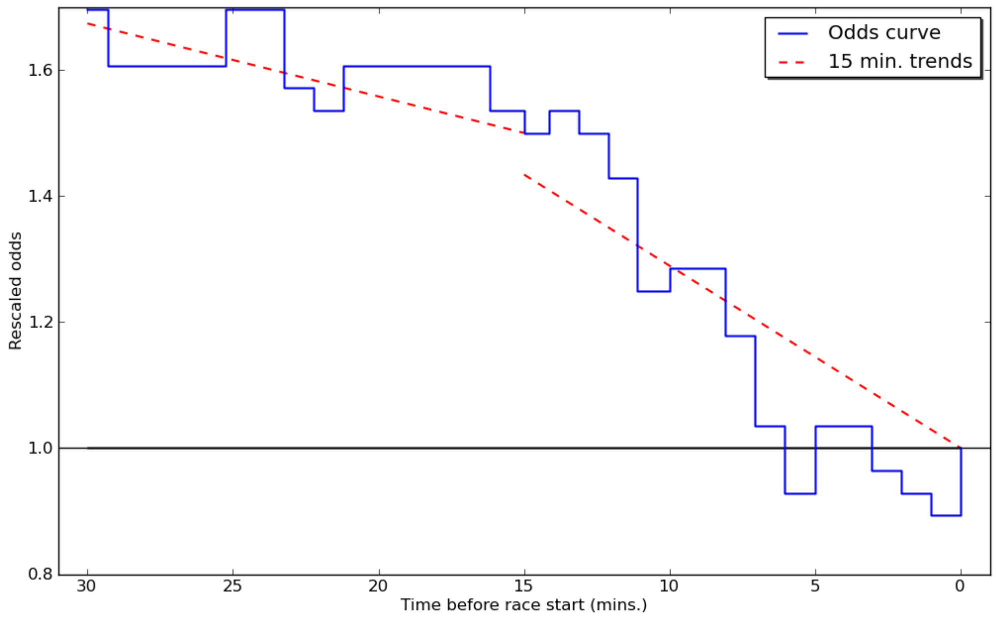

Figure 2. Log of cumulative wealth relative to initial wealth from betting strategies designed to capitalize on mis-pricing resulting from herd behavior. A benchmark model simply incorporating the odds variable (Model 0) recommends no bets, and the log of final accumulated wealth using the winning probabilities form Model 0 would be £0.

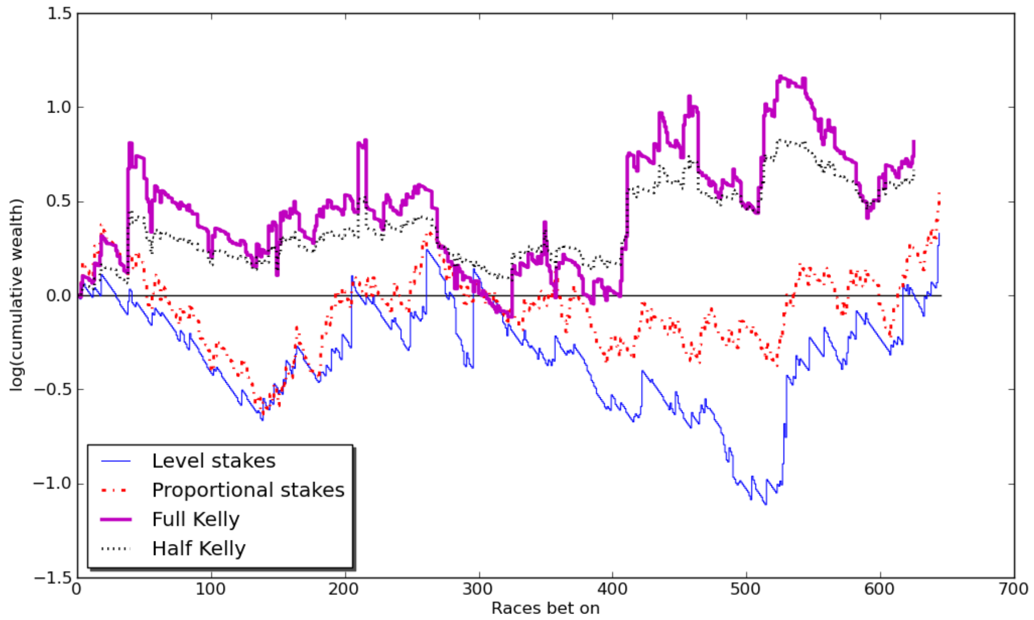

Table 1. Descriptive statistics for the 6,058 horseraces (62,124 runners) run in the UK and Ireland from August 2009 through August 2010 and employed in the study.

	Min.	Max.	Mean*	Median*	Std. dev.	No. of zeroes
No. of runners per race	2	30	10.25	10	3.83	
Odds	1.12	1000	59.52	16.00	145.52	
μ_{ij1}^{+}	0.000	36.937	0.578	0.188	0.496	21,600
μ_{ij1}^{-}	0.000	15.406	0.594	2.810	0.698	40,600
μ_{ij2}^{+}	0.000	10.023	0.159	0.108	0.226	30,730
μ_{ij2}^{-}	0.000	9.019	0.166	0.147	0.270	32,734
μ_{ij3}^{+}	0.000	30.001	0.302	0.126	0.228	27,004
μ_{ij3}^{-}	0.000	15.030	0.362	0.144	0.410	35,526

*Values of zero for the trend in odds variables are excluded when calculating the mean, median and std. dev., since zero implies that the trend in odds was in the opposite direction to that captured by a particular trend in odds variable (e.g., when μ_{ij1}^{+} has a positive value then μ_{ij1}^{-} takes the value zero).

Table 2. Results of fitting conditional logit models on the training set of 4,544 races. The first variable in each model is the log of final odds-implied probability. Models 1 to 7 also include trend variables as indicators of herd behavior.

Model	Variable	Coefficient $\beta(l)$	$z(l) = \beta(l) / S.E.[\beta(l)]$ ($S.E.$)	ln$L$	LR test vs. Model 0	Pseudo-R^2
Naive	-	-	-	-10,307.7	-	-
0	$\ln q_{ij}(L_{ij})$	1.015	51.45** (0.020)	-8,337.6	-	0.5798
1	$\ln q_{ij}(L_{ij})$	1.025	45.68** (0.022)	-8,337.2	0.82	0.5799
	μ_{ij1}^{+}	0.058	0.92 (0.063)			
2	$\ln q_{ij}(L_{ij})$	1.016	50.28** (0.020)	-8,337.6	0.02	0.5798
	μ_{ij1}^{-}	-0.004	-0.13 (0.028)			
3	$\ln q_{ij}(L_{ij})$	1.019	50.75** (0.020)	-8,337.2	0.78	0.5799
	μ_{ij2}^{+}	0.117	0.91 (0.129)			
4	$\ln q_{ij}(L_{ij})$	1.014	51.35** (0.020)	-8,330.5	14.27**	0.5812
	μ_{ij2}^{-}	-0.471	-3.56** (0.132)			
5	$\ln q_{ij}(L_{ij})$	1.031	48.82** (0.021)	-8,335.3	4.54*	0.5803
	μ_{ij3}^{+}	0.215	2.14* (0.100)			
6	$\ln q_{ij}(L_{ij})$	1.017	51.16** (0.020)	-8,337.3	0.56	0.5799
	μ_{ij3}^{-}	-0.038	-0.74 (0.051)			
7	$\ln q_{ij}(L_{ij})$	1.026	48.52** (0.021)	-8,329.1	16.89**	0.5814
	μ_{ij3}^{+}	0.165	1.63 (0.102)			
	μ_{ij2}^{-}	-0.442	-3.32** (0.133)			

* and ** denote significance at the 5% and 1% level in a 2-tailed test, respectively.

Table 3. Returns from betting strategies applied to the 1,514 holdout races, developed using probabilities estimated from Model 7.

Strategy*		*Level stakes*	*Proportional stakes*	*Full Kelly*	*Half Kelly*
Number of races bet on		546	546	532	532
Total number of bets		644	644	625	625
Number of winning bets		111	111	111	111
Total amount bet (£)		7,479.9	11,759.0	20,495.0	9,114.3
Final capital (£)		1,389.2	1,762.9	2,262.2	1,946.9
Profit or loss (£)		389.2	762.9	1,262.2	946.9
Rate of return *R* (%)		5.20	6.49	6.16	10.39
Risk-adjusted return		0.24	0.48	0.53	0.90
Expected final wealth (£)		1,448.4	1,637.0	1,922.5	1,387.6
Probability that final wealth is above *x*% of initial wealth	50%	0.93	0.69	0.48	0.33
	100%	0.94	0.74	0.54	0.40
	150%	0.76	0.64	0.55	0.49
	200%	0.86	0.64	0.47	0.35

*The characteristics of the four betting strategies and the five means of assessing the performance achievable from these strategies (Number of winning bets, total amount bet (£), final capital (£), profit or loss (£), rate of return *R* (%), risk-adjusted return, expected final wealth (£), probability that final wealth is above x% of initial wealth) are described in Section 3 (Data and Methodology), in the sub-section *'Assessing forecast accuracy with betting strategies.*