University of Reading

Information and predictability: bookmakers, prediction markets and tipsters as forecasters

Article

Accepted Version

It is advisable to refer to the publisher's version if you intend to cite from the work.

www.reading.ac.uk/centaur

Information and Predictability: Bookmakers, Prediction Markets and Tipsters as Forecasters

by
J. James Reade

Information and Predictability: Bookmakers, Prediction Markets and Tipsters as Forecasters

J. James Reade

Department of Economics, University of Reading
SAIS — The Johns Hopkins University

April 29, 2014

Abstract

The more information is available, and the more predictable are events, the better forecasts ought to be. In this paper forecasts by bookmakers, prediction markets and tipsters are evaluated for a range of events with varying degrees of predictability and information availability. All three types of forecast represent different structures of information processing and as such would be expected to perform differently. By and large, events that are more predictable, and for which more information is available, do tend to be forecast better.

1 Introduction

The more information is available, and the more intrinsically predictable are events, the better forecasts ought to be. Further, the manner in which forecast models process that information must matter for forecast quality. In this paper we investigate these basic tenets of forecasting theory using three market-based mechanisms for generating forecasts of sporting outcomes; we consider traditional bookmakers, prediction markets, and the forecasts of a self-selected group of experts, or tipsters, for a range of sporting events. While all three are market based and hence subject to commercial pressures which may influence forecasts produced, the context of sports betting is a highly competitive market in which poor forecasts embodied in offered odds (prices) would be either ignored by agents who can bet elsewhere, or exploited to make guaranteed returns. As such we argue there are strong incentives for accurate forecasts to be produced. We consider a large number of football matches in England across a range of divisions; these different divisions, or leagues, provide us with events that have different associated costs to procure information, and different levels of predictability.[1] The mix of bookmakers, traders on prediction markets and tipsters providing forecasts yields a range of different types of forecasts that will rely, to differing extents, on statistical models and judgement, and are also subject to differing degrees of market pressures. As we are evaluating different aspects of the betting industry, we also investigate the favourite-longshot bias, commonly found in previous empirical investigations.

---

[1]When using the term 'football' in this paper, we are referring to *Association Football*, or *soccer* for short.

1

We assess forecasts produced by our different sources for a common set of events using a number of frequently used metrics before evaluating a small number of basic trading strategies based upon different forecasts. We find that, consistent with commercial pressures, tipster forecasts are inferior to those from bookmakers and prediction markets, and we argue our results lend tentative support to one particular, information-centred explanation for the favourite-longshot bias. Finally we find that different strategies based on bookmakers and prediction markets yield different, but never significantly positive, expected returns.

In Section 2 we introduce our different forecast models and outline how we will compare them, in Section 3 we introduce our dataset, in Section 4 we present the results of our comparisons, and in Section 5 we conclude.[2]

2 Methodology

Forecasting outcomes has always been an important pursuit of economists, and as such, various methods have been developed to carry out such forecasts. All methods differ, fundamentally, in how they process the available information, and it is that information availability that in part determines to what extent an event is forecastable (see, for example Hyndman and Athanasopoulos, 2012). Statistical models can be used (see, for overviews Hendry and Clements, 1998, 2001), while judgemental methods rely more on the judgement of agents involved (see, for example Lawrence et al., 2006; Goodwin and Wright, 1993). Agents in various common situations generate forecasts based on some combination of these two approaches; for example during political elections pollsters produce forecasts regarding vote shares by sampling the population, and in sporting events bookmakers produce forecasts regarding outcomes. Bookmakers are one vehicle through which agents can trade based on predictions; assuming a classical framework agents will bet with a bookmaker up to the point where their subjective beliefs, or forecasts, regarding outcomes differ from those posted publicly by bookmakers. In recent years prediction markets such as Betfair (see www.betfair.com) have allowed agents to trade contracts with each other, contracts whose value is contingent on the outcome of some forthcoming event, based on their forecasts regarding those outcomes. Prediction markets differ fundamentally from bookmakers in that the prices offered are those offered by other willing bettors, rather than by a bookmaker. Furthermore, as a prediction market does not take one side of bets agreed on it, it has no exposure to losses caused by unexpected outcomes of events. Historically for political and sporting events, experts have pronounced opinions, often in newspapers, regarding outcomes, and these predictions have been analysed (Forrest and Simmons, 2000), in particular, regarding their use of information, both public and private. Finally, fora have always existed for agents to share predictions without taking financial positions based upon them; Servan-Schreiber et al. (2004) analyse one such forum, a market in which users trade based on virtual money for 'bragging rights' rather than financial return, but many other formal (e.g. fantasy leagues) and informal institutions (e.g. discussions amongst friendship groups)

---

[2]The data collected for this study was scraped using Python (see www.python.org/, and the econometric analysis carried out in R (R Core Team, 2014). Some of the tables were created using stargazer (Hlavac, 2014).

2

exist for this purpose, online (e.g. the *Hollywood Stock Exchange*) and offline (e.g. office sweepstakes).

In this paper a novel dataset is exploited in order to compare the forecasts from these three sources: bookmakers, prediction markets and tipsters. The website www.oddsportal.com lists prices published by bookmakers, and those available on prediction markets, immediately prior to the start of historical sporting events. Alongside these prices are the predictions of numerous website users for the events in question. The website, as its name suggests, collects odds for the information of those seeking to place bets on sporting outcomes. As such, we can reasonably presume that those users who post their own forecasts also bet on these match outcomes. The website has a 'community' in which users post forecasts, and other can comment on them, should they choose. For any given historical match, a summary of these user forecasts, alongside bookmaker and betting exchange odds, is provided, and as such we have a significant range of forecasts per match, differing in their processing of public (and potentaily private) information. The existence of a website such as this one is indicative of the level of competition that exists amongst bookmakers; this is the culmination of a dramatic change in sports betting markets operating in the UK; Forrest et al. (2005) discuss the onset of this change in the late 1990s, and it is likely that their dataset, covering 1998 to 2003, covers the greatest period of structural change as competitive pressure increased. Our dataset, collected for the period 2011–2014, comes after much of that structural and legislative change has taken place, and hence covers a more settled market setting, but one nonetheless that appears very highly competitive.

Through the types of matches we collect data on, we also have a range of characteristics of events being forecast. We collect data for the top five English football leagues; we take the almost complete current seasons, plus the previous two complete seasons, for the English Premier League (EPL), the English Championship (ECH), League One (LG1), League Two (LG2), and the Conference Premier Division (CPD). The EPL is the top professional football league in England, and is widely considered to be one of the top football leagues globally. The Championship is the league immediately below the Premiership, and within the ranks of second tier leagues globally, it ranks highest both in terms of product quality and quantity demanded.[3] LG1 and LG2 are the subsequent two professional (Football League) divisions and together with ECH they form the *Football League*, and the CPD is the fifth division of English football, where many teams are only semi-professional. European-style football leagues are distinct from North American-style ones by the presence of promotion and relegation — movement between a hierarchy of leagues based on the realised quality of teams competing. Hence, for example, there is direct movement between the ECH and the EPL; the three best team from the ECH move into the EPL (promotion), and the three worst teams from the EPL move into the ECH (relegation), each year. Promotion and relegation is possible between any two adjacent leagues in the pyramid system (hence also between ECH and LG1, though not directly from LG1 to EPL). As the EPL is the top league in England, there is no promotion from it to any higher league, whereas for all other divisions there is both promotion and relegation.[4] This feature likely con-

---

[3]For the former, see www.transfermarkt.de/en/2010/startseite/wettbewerbeDE.html?from=top_navi, and for the latter, see comparetheleagues.com/european-league-stats/second-divisions/.

[4]The pyramid structure of English football is such that it continues down to very sub-
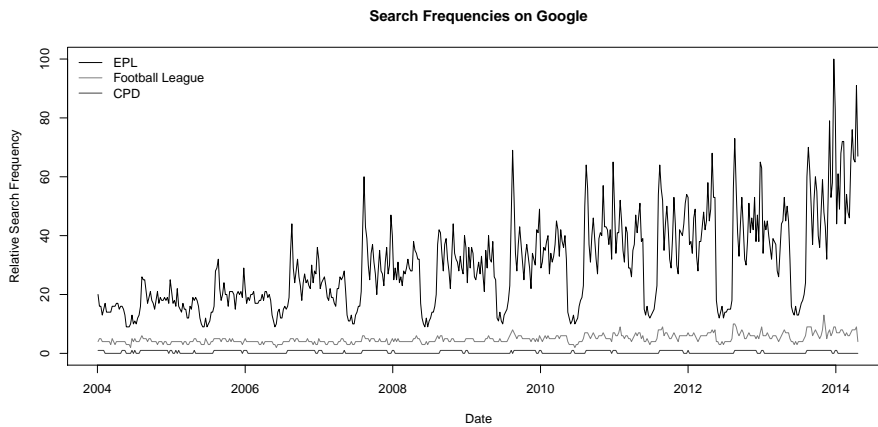
Search Frequencies on Google

Figure 1: Search frequencies for the different football leagues considered in our dataset, revealing the different levels of interest and hence information available on each. Source: Google Trends.

tributes to the quality differences observed between teams in the EPL relative to the two other leagues; the best teams remain each year rather than enter a higher league, and as such the distribution of team quality might be expected to be more skewed, on average, as a result. The statistics appear to support this; for the 2013/14 season, the excess skewness for EPL average team quality was almost twice that of the ECH and around nine times that of the CPD. If there are larger quality differences between teams, this makes the events intrinsically more predictable.

The quality of forecasts must also be a function of available information. By information we refer to detailed information regarding teams as they approach each match, rather than simply the measured quality of teams competing. For example, whether a team is suffering from injuries, fatigue, other types of unrest, and whether conditions at a forthcoming match will favour one team over another. By and large, this kind of information is much more readily available for matches in the top leagues, hence EPL in particular, and ECH, as media interest is more focussed on these leagues. Figure 1 lends support to this information hypothesis, as it plots search frequencies on Google for the leagues our dataset covers. The EPL has, over the period our data covers (since 2011), about seven times as many searches as the entire Football League combined (ECH, LG1 and LG2), and more than a hundred times as many searches as the CPD has. Google searches can be viewed as a proxy for demands for information about these leagues and gives some sense of the differences between them.[5] As such, we propose that forecasts will be better for the EPL than the four leagues beneath it, as both more information is known about this league, and as mentioned, the quality differences between teams are greater also. We might

---

regional, amateur levels. The pyramid as a whole contains 120 leagues with 480 divisions (see http://en.m.wikipedia.org/wiki/English_football_league_system).

[5]It is likely that this comparison is favourable to the CPD as "football conference", the search term used, picks up a number of North American 'conference' leagues.

anticipate a steady deterioration in forecast quality as we move to lower leagues, and based on the information hypothesis regarding the favourite-longshot bias, we might also anticipate that this bias gets larger the lower the league we are considering, as the costs of obtaining important information are higher. That said, however, both Forrest et al. (2005) and Forrest and Simmons (2000) refer to the 'broken-leg cue', bits of information that may be picked upon more likely by bookmakers, bettors and tipsters if media coverage is greater, as potentially detrimental if rather than providing important information, it only provides noise. It may be that all the extra information available at higher levels, rather than providing actual signals about changes in relative team quality, actually provides only noise and thus makes forecasts worse.

It is naturally of interest to understand which forecast methods perform best; for bettors this may inform better strategies to make better returns, whilst of more academic interest principles of forecasting theory can be evaluated. In comparing such different sources of forecasts, this paper is similar in nature to Spann and Skiera (2009), who compare forecasts by the German state-run bookmaker *Oddset*, from a German virtual prediction market, and by a tipster at the leading German sports newspaper, *Sport Bild*.[6] They find that tipster forecasts are inferior to both their bookmaker and prediction market, and note that with lesser commission charges, a trading strategy building on prediction market forecasts could yield positive expected returns. Our study differs from Spann and Skiera as we consider multiple competing bookmakers, alongside commercial prediction markets that compete with bookmakers, and we look at a self-selected group of tipsters rather than professional ones employed by newspapers.

Before describing our methods for comparing these forecast sources, we introduce our notation. We refer to outcome k of football match i as y_{ik} such that:

$$y_{ik} = \begin{cases} 1 & \text{if event } k \text{ occurs in match } i, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

The existence of a tied outcome, or draw in more common parlance, distinguishes football from numerous other sports, particularly those of North American flavour. While adjustments could be made to factor in the tied outcome, for our purposes it is sufficient simply to consider the drawn match outcome as another possible outcome of a match. Hence in (1) k refers to three possible outcomes: the home team wins, a draw, or the away (visiting) team wins. We observe forecasts of these three outcomes from three different sources: bookmakers, prediction markets and tipsters, and refer to such forecasts as $\widehat{y}_{iks}$ for match i, event k, and forecast source s. Forecasts $\widehat{y}_{iks}$ are bounded between zero and one; we convert all forecasts into probabilistic forecasts, as both bookmakers and prediction markets report their forecasts in terms of decimal odds, denoted D_{iks}. We convert decimal odds into probabilities by inverting them, hence $\widehat{y}_{iks} = 1/D_{iks}$. We must correct bookmaker and prediction market odds for the overround; if the raw probabilities taken from the reciprocal of the decimal odds are summed up for all, their total is usually greater than one. The overround is considered to be the transactions cost of a bet, in that it reveals

---

[6]Virtual in the sense that no real money is traded on outcomes; instead some form of virtual money is traded.

the cost to a bettor of securing a guaranteed return. Amongst bookmakers it is generally argued that the overround reflects the bookmaker's margin. On prediction markets, where it has been shown to be smaller (Vaughan Williams, 1999), it still reflects the margin of those trading, since all prediction markets levy some explicit commission charge on net winnings.

In an early investigation of market efficiency of bookmakers, Pope and Peel (1989) consider bookmaker prices which are posted some time in advance of the sporting events in question, English football matches. They found some evidence for inefficiency, but an absence of profitable trading strategies. As already mentioned, such early studies consider a very different betting industry to what exists now, and as such our results may differ somewhat to those found under a much more heavily regulated industry.

Forrest et al. (2005) consider five bookmakers in the midst of the extensive deregulation of the betting industry, of which three are found in our dataset of 58 bookmakers. With such a number of bookmakers, it is not practical to consider the differences between individual bookmakers, but we do in passing make reference to different relative performances by different bookmakers, not least because during their sample period (1997–2002) the authors note that competitive pressures were increasing, and also the performance of bookmakers improved. As with Pope and Peel (1989), in the sample period covered by Forrest et al. (2005) bookmakers posted their prices up to five days before a match began, and generally did not change them. As we detail in Section 3, this is generally not the case now with in-play betting, and constantly updating prediction markets forcing bookmakers to be constantly updating their prices even during matches. Almost overlapping, Deschamps and Gergaud (2012) consider forecasts of English football matches between 2002 and 2006 for five bookmakers (two in common, William Hill and Ladbrokes) in a study of market efficiency, and uncover a number of biases.

Forrest and Simmons (2000) appraise newspaper tipsters forecasting English football matches from all four professional divisions, focussing in particular on three newspapers. They find that while tipster predictions are an improvement on those generated at random, tipsters do not appear to make use of all available information when constructing forecasts, and only one appeared to make use of private information. Finally, they find that a combination of all three tipsters provides better predictions than each forecast individually. This study differs from ours in that our tipsters are self-selected users of a website, rather than being employed by national newspapers to provide predictions. Furthermore, in this paper we only make use of the average of a group of tipsters, rather than each individual's forecast, and as such we already consider forecast combinations in the context of tipsters, rather than each individual prediction.

Other investigators (Croxson and Reade, 2011; Franck et al., 2010) have found prediction markets to outperform bookmakers in terms of odds offered and also liquidity available, although not necessarily in terms of forecast accuracy. This investigation to some extent mirrors forecast comparisons of opinion polls and prediction markets for political elections (Kou and Sobel, 2004; Leigh and Wolfers, 2006; Berg et al., 2008; Rothschild, 2009).

We now describe the various methods we use to compare forecasts.

2.1 Percentage of Correct Forecasts

A very basic method for appraising forecasts is to consider how often these forecasts are correct. The percentage of times a forecast is correct is sometimes referred to as the *hit rate*. While such a method is straightforward when we consider a specific event with some numerical value, it is a little trickier when we are considering probabilistic forecasts of multi-outcome events such as football matches. We must first determine what event has been predicted in any given match, and then determine whether that event occurred. As a simple rule, we take the event in a given match for a forecaster that has the highest probability. For tipsters this is the event that the simple majority of tipsters predicts will occur. We then calculate how often that predicted event actually occurs.

2.2 Mean Forecast Error

We have forecasts $\widehat{y}_{iks}$ of event k in match i by forecaster s, and actual match outcomes y_{ik}. The latter is binary, but considering its average, we get some sense of whether forecasters predict particular events, on average, the right number of times. The forecast error for event k in match i for forecast source s is:

$$\widehat{e}_{iks} = y_{ik} - \widehat{y}_{iks}. \tag{2}$$

A simple measure of forecast accuracy is to take the average of these forecast errors across matches (i) and sources (s):

$$\mathsf{ME}_k = \frac{1}{N} \sum_i \sum_s \widehat{e}_{iks}, \tag{3}$$

where N denotes the total sample size for event k. We consider mean errors separately for each match outcome, as all previous studies noted in Section 2 have noted distinct differences between these events. The mean error (ME) provides a measure of the bias of a forecast source. If we observe $\mathsf{ME} > 0$ this implies that that on average the event k occurs more frequently than source s predicts; source s forecasts are biased down. Conversely, if we observe $\mathsf{ME} > 0$ this suggests forecasts are biased upwards, as events occur less frequently than their forecasts would suggest.

2.2.1 Favourite-Longshot Bias

The simple bias as reflected in (3) is not the only bias of interested when considering forecasts. Deschamps and Gergaud (2012) consider biases in bookmaker prices, and refer in particular to the favourite-longshot bias. This bias, noted in many contexts (see for example Vaughan Williams and Paton, 1997; Woodland and Woodland, 1994), reflects that favourites are priced such that they win more often than their odds imply, and outsiders win less often. To investigate this, one looks at either observed outcomes or theoretical returns as a function of implied probabilities. Given that tipster predictions are not bets, it makes less sense to think in terms of returns in this context. Outcomes ought to be a positive function of implied forecast probabilities, and this relationship ought to be one-for-one, in the absence of any bias: $\mathsf{E}(y_{ik}) = \mathsf{E}(\widehat{y}_{iks})$. If some events

occur more often than their forecast implies, then $\mathsf{E}\left(y_{i'k'}\right) > \mathsf{E}\left(\widehat{y}_{i'k's}\right)$ for that subset of events $i' \in i, k' \in k$, and positive returns could be made. Specifically if outcomes that are favourites occur more often, then $\mathsf{E}\left(y_i'k'\right) > \mathsf{E}\left(\widehat{y}_{i'k's}\right)$ for events that are most likely, while $\mathsf{E}\left(y_i'k'\right) < \mathsf{E}\left(\widehat{y}_{i'k's}\right)$ for all other types of events; this is the favourite-longshot bias. One method to test its existence is to run the OLS regression of y_{ik} on $\widehat{y}_{iks}$:

$$y_{ik} = \alpha + \beta\widehat{y}_{iks} + e_{is}, \tag{4}$$

and test whether $\alpha = 0$ and $\beta = 1$ such that $\mathsf{E}\left(y_{ik}\right) = \mathsf{E}\left(\widehat{y}_{iks}\right)$. If $\beta > 1$ then the favourite-longshot bias exists, as more probable events (y_{ik}, a probability, nearer to one) occur more often than their implied probability dictates they should, while less probable events (y_{ik} nearer to zero) occur less often than their odds suggest.

It is of interest to consider the favourite-longshot bias regression (4) along a number of dimensions in our dataset, as many explanations for the bias exist. The two main explanations are risk love and misperceptions (Snowberg and Wolfers, 2010). The bias suggests that bets with low odds and hence high risk have lower payoffs, and the only way to reconcile this within the neoclassical paradigm is to assume risk loving agents. On the other hand, it may simply be that agents are unable to distinguish between the likelihood of events with small probabilities, moving towards the behavioural economic paradigm. Another hypothesis is that information costs contribute to the bias being priced into bookmaker odds (Hurley and McDonough, 1995); while bookmakers are equipped to collect information regarding likely outcomes, bettors are not. In the limit, bettors are totally uninformed and would back all events equally, and as such it is in the interest of bookmakers to price favourites more attractively, and outsiders less so, in order to encourage bettors to back the favourite over the outsiders. Supporting this hypothesis, Vaughan Williams and Paton (1997) found the favourite-longshot bias to be stronger in lower grade horse races where information is more costly to obtain, and in finding similar patterns on prediction markets, (see Smith et al., 2006) offers further support for the information costs hypothesis.

We can check this hypothesis using our dataset as we collect data from five different English football leagues, all differing in the volume of information available for; in this sense our work mirrors that of Vaughan Williams and Paton (1997). By looking at tipster forecasts, we can provide new insight into the bias; clearly tipsters bear no risk in making their predictions, hence bias observed cannot be explained by this. Conversely, if the explanation is misperceptions, then the bias amongst tipster predictions ought to be similar to that of prediction markets; there seems little reason to believe that when making forecasts upon which money is staked that misperceptions should be systematically different to those where this risk is absent, particularly since those making forecasts on a website such as *OddsPortal* likely bet themselves. Finally, tipster forecasts allow us to appraise how individual bettors process information and thus distinguish favourites from longshots, a key component of the information costs explanation. Hence, assuming both predictions on a prediction market and on a tipster forum are in all other respects identical, considering both ought to allow adjudication between these two explanations.

2.3 Mean absolute forecast error

The mean absolute forecast error is:

$$\mathsf{MAE} = \frac{1}{N} \sum_i \sum_s |\widehat{e}_{is}|, \tag{5}$$

where N is the total number of forecasts, i refers to the event being forecast, and s to the forecast source. The MAE gives a measure of absolute forecast error size by ignoring the sign of the forecast error (of importance for determining bias). Thus rather than allowing positive and negative errors to cancel out as in the mean forecast error, the mean absolute error gives a sense of how large each forecast error is, in absolute terms.

2.4 Mean squared forecast error

The mean squared forecast error is:

$$\mathsf{MSE} = \frac{1}{N} \sum_i \sum_s \widehat{e}_{is}^2, \tag{6}$$

and as with the mean absolute error, provides a measure of the size of errors independent of their sign. Additionally, by squaring each error, we get an approximation to the variance of forecasts, assuming they are unbiased. This may be of interest when considering our signal vs. noise information hypothesis; if extra information for higher leagues contains a significant proportion of noise, this would be expected to only increase the variance of forecasts rather than their accuracy.

The mean squared forecast error is often referred to as the Brier score (Brier, 1950) in the context of binary outcomes like those we consider in this paper. The MSE can differ from the MAE, and the usual explanation for this is that there may be a number of particularly large forecast errors; squaring affords them a greater weight than taking absolute values.

2.5 Profitability of betting strategies

Finally we attempt to add some economic significance to results by considering hypothetical trading strategies. The strategies are hypothetical in that they were not actually carried out, and had they been carried out bookmaker and prediction market prices may have differed to those observed in our dataset. Nonetheless if we assume sufficiently small bets were placed (say a pound in each case) then we can be reasonably confident that this would not have influenced market outcomes.[7] At least two ideas underly this appraisal of betting markets; the first is that any positive expected returns suggests the market is inefficient in that trading had not taken place to arbitrage away any such profitable differences in prices. The second suggests that if by looking at combinations of bookmaker and prediction market prices significantly different returns can

---

[7]As will be seen later, there is some range in the amount of liquidity available on prediction markets, and a tiny number of markets have such small amounts of liquidity available that this strategy might have had some impact, but this is a handful of matches in a sample of many thousands.

be made, it suggests there are economically meaningful differences between the two markets. We also make use of tipster predictions to inform a betting strategy to again determine whether tipster forecasts are economically meaningfully different from bookmaker and prediction market forecasts.

3 Data

Our dataset is composed of 7622 matches, 1095 from the EPL, 1640 from each of the ECH and LG1, 1642 from LG2 and 1605 from the CPD.[8] We have forecasts from bookmakers, prediction markets and tipsters, and overall we have 425,688 observations, on average 55.8 observations per match. In our dataset, 42.9% of matches result in home wins, 26.6% as draws, and 30.5% as away wins. Similarly, the average prediction from all of our sources for a home win is 43.3%, for a draw 26.8% and for an away win 29.9%.

The observations we have per match are split into three types, as already indicated: bookmaker prices, prediction market prices, and tipster forecasts. We now discuss each in turn.

3.1 Bookmakers

Our dataset contains odds posted by 58 different bookmakers for all three events; we also have information on the percentage of bets that were paid out by each bookmaker for each match. None of these bookmakers report odds for every single matche in our dataset, but 53 provide prices on matches in all five leagues, and all but one provide prices on the top four leagues.[9] The bookmakers listed rank from traditional UK bookmakers like *Ladbrokes* and *William Hill* to more recent market entrants such as *FortunaWin* and *SportsBetting*. Many betting companies have different geographical focusses, with, for example, *NordicBet* focussing on the Nordic countries. Oddsportal provides a ranking of the different bookmakers it sources prices from, based on how often payouts are made on bets, the rating provided by www.BookmakersReview.com, and also the ratings that Oddsportal users provide.

As a panel dataset, our bookmakers thus form a reasonably balanced panel; over a period of three years there appears to be no market exits, and a small number of entrances (see Table 7 in the Appendix on page 26 for information), although it is to be noted that some of the bookmakers used in other studies cited earlier (e.g. StanleyBet, Oddset and GameBookers) do not appear in this dataset. It is not clear the basis upon which *OddsPortal* collects odds information from particular bookmakers, but nonetheless with on average 50.6 bookmakers per match we have a rich picture of bookmaker odds during our sample period. A final point to note is that some smaller bookmakers may rely on a single source of information for their prices, rather than each individually employing odds-setters. This may result in particularly high correlations between particular bookmakers that belong to a specific group. Most correlations

---

[8]The EPL is composed of 20 whereas the other divisions have 24 teams, meaning that there are 380 matches in the EPL per season, and 552 in the others (plus end-of-season playoff matches).

[9]BetOnline, Instant Action Sports, Intertops, SportsBetting and The Greek are the bookmakers who do not provide CPD prices.

between bookmakers are above 0.9, which could cause problems for interpreting standard errors, although our huge sample size reduces this problem somewhat.

3.2 Prediction Markets

Our dataset includes prices and available liquidity for two betting exchanges for all observations, Betfair and Betdaq, and an additional exchange for all our EPL and Football League observations, Matchbook. By available liquidity we mean amounts of money available for events to be either backed (bought) or laid (sold). Betfair and Betdaq have been in operation since 2000, while Matchbook only entered into European sports markets in 2012. Betfair is by some distance the largest prediction market offering sports bets, and this is reflected in the amount of money available to back bets in our dataset; in total Betfair has £68.1m available over the matches in our dataset, followed by Betdaq with £10.8m and Matchbook with £7.4m. Betfair charges between 2 and 5% commission, Betdaq charges 3% on 'most markets', and Matchbook charges 1%.[10]

Of particular interest from prediction markets are the back and lay prices. As both prediction markets reflect user activity, we can infer some degree of strength in predictions, as well as popularity of particular matches. The amount of money available to lay an event reflects the number of bettors willing to back that event, but who as yet have not been matched with another bettor willing to lay the event, and as such a large volume of money available to lay an event must reflect the strength of sentiment amongst bettors for that particular event. Other information we have with the back and lay prices is the betting equivalent of the bid-ask spread: the difference between the price to back the event, and the price to lay it. A larger spread might reflect a less liquid market (as would volume information), and may impinge on the quality of forecasts produced.

3.3 Tipsters

Our dataset includes the outcomes of tips made by website users; as of April 17 2014, the website listed 1967 users who had made at least 50 predictions on football matches alone. For any given match the number of users providing predictions varies (and is difficult to collect information on), but for EPL and CPD matches the number predicting appears to depend on the number of matches taking place elsewhere at a given time, and the interest level in that match (hence Manchester United matches attract more attention than Welling United matches), and the most interesting matches attract in excess of 300 forecasts in both these extreme leagues in our sample, whereas some of the less interesting matches can have fewer than 10 forecasts.

An important point related to these tipster forecasts is that they are not all by the same tipsters, and it is unlikely that one tipster has forecast all matches in our sample. In that respect this data more accurately resembles a pooled cross section rather than a panel of tipsters. Of the forecasters on *oddsportal*, there is a considerable range of quality of forecasts, as measured by ROI, which ranges from 291.9% to -72.4%. For each match the proportion of tipsters picking each

---

[10]The commission charges by Betfair may differ based on the country in which a trader is based.

event to occur is reported, hence if 50 people out of a total number of forecasters for a given match of 100 predict the home win to occur, 50% is reported.

Rather than being employed by newspapers to provide tips each week, the tipsters in our dataset are a self-selected sample of *OddsPortal* users. As with newspaper tips, their forecasts are made public, as the website resembles a social network site as tipsters have public profiles and can comment on the tips others make for sporting events. We can surmise that given these tipsters have become registered users of the *OddsPortal* website, they are seeking to bet on sporting outcomes, and hence will likely alongside making their tips place bets on events. While it is less reasonable to anticipate that these tipster forecasts have any impact on bookmaker prices, it seems plausible to imagine that the strengths of tips made on the website might reflect the strength of betting on prediction markets; the fundamental difference being that when posting tips regarding outcomes the only risk users face is that of social embarrassment, whereas when placing bets on betting exchanges or bookmakers (which of course is anonymous) there is financial risk without any immediate social risk.

4 Results

4.1 Percentage of Correct Forecasts

Considering first the percentage of correct forecasts, we take the outcome forecast as that with the highest probability, and calculate how often that event occurs. Overall, 47.6% of the time forecasts in our database are correct, although this differs between divisions, as might be expected based on the earlier discussion regarding information and predictability. Potentially following this hypothesis, 55.28% of EPL forecasts are correct, the highest proportion. Following this, ECH and LG1 matches are forecast correctly between 45 and 46% of the time, and in LG2 this falls to 43.88%. Countering the hypothesis, however, is CPD, our lowest league, where 49.11% of matches are forecast correctly. All these differences are statistically significant, countering our earlier hypothesis since a league in which information costs are higher, CPD, has a greater number of correct forecasts than all Football League divisions.

Breaking down the forecasts by season does not affect the rankings; in general, forecasts in 2014 are statistically significantly better than both previous years, and forecasts in 2013 were the worst overall, but in each year CPD forecasts were better than Footbal League ones. One alternative explanation for this unexpected finding is the mix of professional and semi-professional teams in the CPD; in the EPL and Football League all teams are professional, and as such it might be that the distinction between professional and amateur teams yields an extra element of predictability in the CPD.

Considering our different types of forecast, in Table 1, we find that in all divisions bookmakers and prediction markets outpredict tipsters by a statistically significant margin (p-value 4.4%, not reported elsewhere), whilst the difference between the two is statistically insignificant. This result is consistent with Spann and Skiera (2009). In the CPD tipsters forecast correctly 46.24% of the time, while bookmakers and prediction markets forecast correctly 49.08% and 49.16% of the time, respectively. In the EPL all percentages are higher, with tipsters forecasting correctly 52.69% of the time, and bookmakers and prediction

	EPL	ECH	LG1	LG2	CPD
Betting exchanges	55.91	45.39	46.94	44.29	49.08
Bookmakers	55.26	45.34	46.79	43.91	49.16
User predictions	52.69	43.46	43.60	40.16	46.24

Table 1: Percentage of forecasts that turn out to be correct, by type of forecast and league.

markets predicting correctly 55.26% and 55.91% of the time, respectively.

When considering individual prediction markets and bookmakers, all are (individually) significantly better than tipsters at a 5% level of significance; however, reflecting the high correlation between bookmaker and prediction market prices, jointly the prices are insignificant.[11] Of the prediction markets, Matchbook's, The Greek's and Instant Action Sports forecasts are all 5 percentage points better than those of tipsters.

Finally considering prediction markets, the amount of available liquidity for the home win market does increase the accuracy of predictions, and although the effect is statistically significant, it is small; an extra thousand pounds of liquidity only increases accuracy by a third of a percentage point. An extra thousand pounds of liquidity in the away win market has a similar effect, although this is only borderline significant. An extra thousand pounds in the draw market reduces the accuracy of forecasts by a slightly larger amount, but again this effect is statistically borderline, and further there are few markets in our entire dataset where more than a thousand pounds is available to back or lay the draw.[12] These results may give some credence to the so-called 'wisdom of crowds' hypothesis (Surowiecki, 2004), as they suggest that the bigger are markets, the better is information processed in forming forecasts (at least for home and away wins).

4.2   Mean Forecast Errors

Turning to mean forecast errors, these are presented in the first panel of Table 2 by type of forecast, league and event (home win, draw or away win). In terms of magnitude, the errors from tipsters are always larger than bookmakers or prediction markets, but often the signs are different. Bookmaker and prediction market errors are never statistically significantly different from each other. Tipster forecasts are biased upwards (negative sign) for home and away wins (except in the ECH), but biased down for the draw; hence tipsters under-predict the frequency with which draws occur, and over-predict how often outright results occur. There are no such clear patterns for either betting exchanges or bookmakers. Although small, the errors are all statistically significantly different from zero.

A commonly noted bias in forecasts produced by bookmakers and prediction markets is the favourite-longshot bias: favourite outcomes occur more often than their odds (hence forecast) implies, while unfavoured outcomes occur less often. Within our dataset we can identify favourites (events with highest implied probability) and outsiders (lowest implied probability) for each event and then assess

---

[11] For details see Table 6 the Appendix.

[12] About 14% of observations have more than a thousand pounds available.

		error					abs error			squared error		
		home	draw	away	fav	out	home	draws	away	home	draw	away
betting exchanges	EPL	-0.004	-0.004	0.008	0.011	0.012	0.414	0.364	0.354	0.206	0.180	0.178
bookmakers	EPL	0.001	-0.004	0.003	0.018	0.007	0.419	0.369	0.360	0.206	0.182	0.179
user predictions	EPL	-0.073	0.092	-0.020	-0.201	0.130	0.388	0.317	0.323	0.242	0.203	0.203
betting exchanges	ECH	-0.011	0.004	0.007	-0.015	0.013	0.468	0.396	0.396	0.234	0.199	0.199
bookmakers	ECH	-0.004	0.000	0.004	-0.005	0.009	0.470	0.397	0.399	0.235	0.199	0.200
user predictions	ECH	-0.115	0.103	0.012	-0.253	0.193	0.456	0.357	0.375	0.297	0.242	0.239
betting exchanges	LG1	-0.009	0.002	0.007	-0.002	-0.015	0.459	0.395	0.399	0.229	0.199	0.201
bookmakers	LG1	-0.003	-0.001	0.004	0.008	-0.013	0.462	0.396	0.401	0.229	0.198	0.200
user predictions	LG1	-0.089	0.125	-0.036	-0.288	0.213	0.448	0.338	0.403	0.303	0.248	0.270
betting exchanges	LG2	-0.015	0.002	0.013	-0.012	0.014	0.467	0.399	0.411	0.232	0.201	0.209
bookmakers	LG2	-0.013	0.003	0.009	-0.005	0.015	0.468	0.401	0.413	0.232	0.201	0.209
user predictions	LG2	-0.093	0.116	-0.024	-0.302	0.221	0.461	0.353	0.409	0.309	0.252	0.268
betting exchanges	CPD	0.007	-0.015	0.008	0.006	0.007	0.454	0.380	0.397	0.227	0.187	0.200
bookmakers	CPD	0.011	-0.018	0.007	0.019	-0.008	0.458	0.381	0.400	0.228	0.186	0.200
user predictions	CPD	-0.079	0.120	-0.041	-0.287	0.202	0.447	0.295	0.394	0.315	0.221	0.267

Table 2: Mean forecast errors, squared forecast errors and absolute forecast errors by type of forecast and league.

forecast errors and hence bias for these particular types of events. The fourth and fifth columns contain the forecast errors in these cases. For favourites, in general tipsters show a much greater bias, and in all cases upwards (indicating that favourites are under-predicted, the reverse favourite-longshot bias). For prediction markets and bookmakers for favourites in the EPL and CPD forecasts are biased down (favourite-longshot bias), but for the Football League generally forecasts are biased up (suggesting reverse favourite-longshot bias). For outsiders, tipsters again display bias consistent with the reverse favourite-longshot bias, forecasting outsiders to win more often than they actually do, but interestingly with the exception of LG1, bookmakers and prediction markets also display bias in this direction, suggesting that underlying the favourite-longshot bias is an over-prediction of favourites to win rather than an under-prediction of outsiders to win.

However, it should be noted that this is not the standard method for detecting the existence, or otherwise, of favourite-longshot bias. The numbers in Table 2 may be distorted by events for which all outcomes are similarly likely; more common methods for detecting the bias described in Section 2.2.1 will now be employed.

Considering (4) on our dataset as a whole, column (1) of Table 3 reports this; we see that $\alpha = 0.006$ and $\beta = 0.983$, and the F-test that $\alpha = 0$ and $\beta = 1$ is rejected. The β coefficient less than unity suggests that, overall, there is negative favourite-longshot bias in our sample; this, however, mixes together prediction markets, tipsters and bookmakers. The second column of Table 3 adds dummy variables to (4) in the following manner:

$$y_i = \alpha(1, D_j) + \beta \widehat{y}_{is}(1, D_j) + e_{is}, \tag{7}$$

where D_k contains dummy variables, in the first instance for the type of forecast, and both α and β are vectors of coefficients. Such a construction enables the entire regression model to vary for bookmakers, tipsters and prediction markets, and allows us, via the testing of the appropriate elements of the α and β vectors, whether the differences are statistically significant.

We note from column (2) that for the reference group in the regression, bookmakers, $\alpha = -0.019$ and $\beta = 1.057$, indicative of favourite-longshot bias, as might be anticipated, amongst bookmakers. The dummies for prediction

14

markets are both statistically significant suggesting prediction markets are different to bookmakers, and the implied coefficients are $\alpha = 0.0079$ and $\beta = 0.976$, indicative of reverse favourite-longshot bias. This finding is in contrast with that of Vaughan Williams (1999) who found the bias to be merely smaller on prediction markets rather than of the opposite sign. Finally, the coefficients on tipsters are much larger, and indicate that $\alpha = 0.2344$ and $\beta = 0.2969$, implying a substantial reverse favourite-longshot bias, as was indicated by the earlier results in Table 2. These results would appear to support the Hurley and McDonough (1995) information costs explanation for the favourite-longshot bias in the sense that when bettors are not betting, they display an inability to determine favourites (in terms of objective probability) consistent with high information costs, and hence bookmakers must over-price favourites in order to attract bettors to back them, and equivalently discourage bettors from backing outsiders by posting unattractive odds. That bookmakers do not appear to exploit this opportunity need not be evidence against the information costs hypothesis, but may instead indicate that bookmakers simply do not have sufficient price-setting power to act, consistent with the increased competitiveness in the industry as a whole since the late 1990s. Considering the risk-preferences interpretation, the evidence appears less supportive; prediction markets are driven by bettor activity that has financial risk, whereas tipster predictions incur no financial risk; as such, the absence of strong favourite-longshot bias in prediction markets suggests that risk loving preferences on the part of bettors cannot explain the bias. The strong reverse favourite-longshot bias of tipster predictions suggests a willingness amongst bettors to back the underdog, albeit without any reference to the risk associated with such an action on a prediction market.

It may be that a graphical representation of (4) is helpful. In Figure 2 we provide this; we group all forecasts in a particular probability range, and consider how often these events occur. If a forecast source is unbiased, we should expect that, for example, events forecast to occur at a probability of 50% should happen, on average, 50% of the time. If this was the case, then if we plot forecasts in particular probability ranges against the frequency with which those forecast events occur, the resulting plots should appear along a 45-degree line. In Figure 2 we do this for bookmakers, prediction markets and tipsters separately, and plot a 45-degree line. The solid black circles are tipster forecasts, the empty diamonds are prediction market forecasts and the crosses are bookmaker forecasts. With bookmakers and prediction markets we observe that forecasts for very likely events (those forecast to occur with probabilities 60% or more) all tend to actually occur more often than they are forecast to, while those less likely events (for example those forecast to occur with probability 20%) seem to occur less frequently than forecasts imply. This is consistent with the favourite-longshot bias already mentioned. We can also note what was hinted at in Table 2 — that it is events forecast to very likely occur (those with probabilities nearer to 100%, favourites) where the bias is much stronger, relative to events unlikely to occur (outsiders).

Tipster forecasts show, very strongly, the opposite of this, or the reverse-favourite-longshot bias; events forecast to be unlikely to happen, outsiders, occur more often than those forecast imply; events forecast with probabilities less than 30% almost always occur more often than their forecasts imply, and events forecast to occur with probabilities above around 40% (reflecting that there are three events in a football match) seem to occur less often than their forecasts

Bias of Forecasts
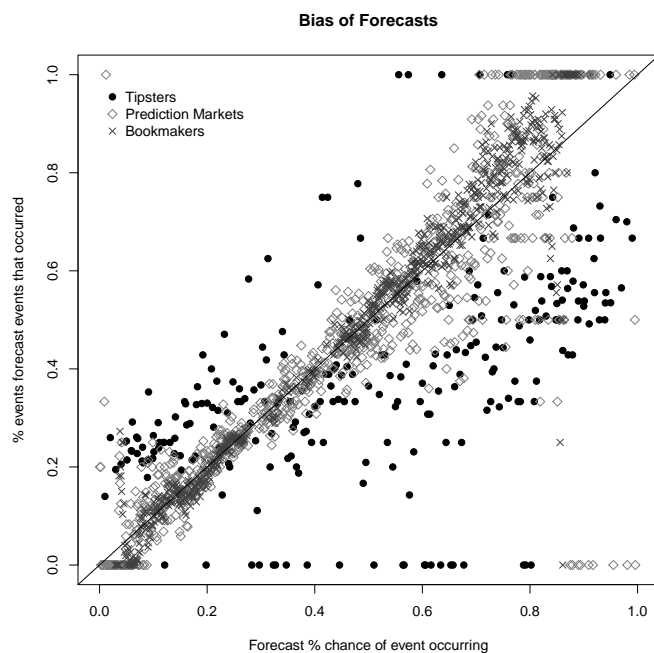
Figure 2: Graphical representation of bias from predictions from bookmakers, prediction markets and tipsters.

suggest they should.

In column (3) of Table 3 we additionally add dummy variables to allow the regression model to be different for each of our leagues as well as for each type of forecast, in order to investigate the impact of information costs. The reference group is bookmakers for the EPL, and hence the line "Forecast, β" gives the actual β coefficient in that circumstance, which is $\beta = 1.078$, suggesting a favourite-longshot bias, as indicated earlier on. If we then run down the leagues (the ECH, LG1, LG2 and CPD lines), we can learn something of the impact of higher informational costs when posting odds. Rather than being uniform as might be expected when moving down to increasingly obscure leagues, there is no clear pattern; for ECH the dummy is sufficiently large that it creates a slight reverse favourite-longshot bias, for LG1 there is no significant difference from $\beta = 1.078$, for LG2 a stronger reverse favourite-longshot bias results, while for CPD there is a slightly smaller favourite-longshot bias. As such there is no clear pattern in terms of information costs, nor predictability, for the favourite-longshot bias in bookmaker forecasts. This is in contrast to the results of Vaughan Williams and Paton (1997) who found that for more obscure horse races the bias was larger.

Turning to prediction markets, we must first add in the prediction markets dummy, and then consider the league dummies in turn. For prediction markets the favourite-longshot bias is greatly mitigated in the EPL but still exists, and in addition to the league dummies (e.g. the line ECH), we must also add in

16

the effect of the interaction dummy between the league and prediction markets, e.g. "Prediction Markets, EPL". Apart from for the CPD, this results in a reverse favourite-longshot bias for all other leagues (hence ECH, LG1 and LG2). Again this gives no simple pattern as we move to leagues for which there are greater information costs. As with simple bias earlier, however, it might be that the professional/amateur distinction in the CPD renders the league much more predictable than any of the Football League divisions, and hence we see favourite-longshot bias, but when we move into markets for which it is harder to predict (high costs of accumulating information), in the Football League, we find a reverse favourite-longshot bias. This would appear to be evidence contradicting the information hypothesis for the favourite-longshot bias.

Finally, for tipsters, we see that for each of the non-EPL divisions, the reverse favourite-longshot bias increases, and is worst for LG1 and CPD, where the β coefficients are smaller than 0.3, though remaining significantly greater than zero. These results, it might be argued, reflect the information costs individual agents face relative to bookmakers distinctly, as agents are simply unable to forecast outcomes particularly well. What is perhaps confounding for the information-costs explanation of the favourite-longshot bias is that bookmakers do not appear to exploit this more strongly, since the bias is weaker than the EPL or negative beneath the EPL.

While it might be hypothesised that information costs are greater for individual bettors than for bookmakers, this does not explain the substantial difference between tipster forecasts and prediction market forecasts, which are also, in principle, based on the actions of individual bettors. These results may thus suggest that the more sophisticated bettors on prediction markets (not those placing their tips on the website) are able to exploit the uninformed bettors, and the result is that there is little, if any bias on the prediction markets, while by and large the favourite-longshot bias still remains for bookmakers to induce bettors to back favourites rather than outsiders. It is worth noting though that for two of our five divisions a reverse favourite-longshot bias exists, suggesting perhaps that in those markets competitive pressure are such that bookmakers have diminished price setting power.

4.3 Mean Absolute Forecast Errors

Turning to mean absolute errors (MAEs, see equation (5), we are considering the absolute magnitude of errors and ignoring the direction of errors. As such we focus more on just how big errors are, rather than any particular biases that exist. A striking result for MAEs is that we find that tipsters forecast marginally, but statistically significantly, better than bookmakers and prediction markets for all match events in all divisions, apart from for the away win in LG1. Prediction markets have significantly smaller absolute errors than bookmakers for the home and away win markets, but for the draw market the difference is only significant at a 10.5% level of significance. Absolute errors are smaller for EPL matches than CPD matches, but for the home win and draw markets, ECH errors are larger than CPD matches.

The result in favour of tipsters is at odds with Spann and Skiera, who found that tipsters forecast significantly worse than prediction markets and bookmakers, and Forrest and Simmons (2000) also noted a poor performance of tipsters. However, Spann and Skiera do not consider MAEs in their study.

Table 3: Favourite-Longshot Bias Regression Results

	Dependent variable:		
	outcome		
	(1)	(2)	(3)
Constant (α)	0.006***	−0.019***	−0.026***
	(0.001)	(0.001)	(0.002)
Prediction Markets		0.027***	0.020**
		(0.004)	(0.008)
Tipsters		0.253***	0.182***
		(0.005)	(0.013)
CPD			−0.016***
			(0.004)
ECH			0.029***
			(0.004)
LG1			−0.001
			(0.004)
LG2			0.038***
			(0.004)
Prediction Markets, ECH			0.003
			(0.013)
Prediction Markets, LG1			0.019
			(0.013)
Prediction Markets, LG2			0.004
			(0.014)
Prediction Markets, CPD			0.013
			(0.013)
Tipsters, ECH			0.054***
			(0.017)
Tipsters, LG1			0.098***
			(0.017)
Tipsters, LG2			0.072***
			(0.018)
Tipsters, CPD			0.095***
			(0.017)
Forecast (β)	0.983***	1.057***	1.078***
	(0.003)	(0.004)	(0.006)
Prediction Markets		−0.081***	−0.059***
		(0.012)	(0.022)
Tipsters		−0.760***	−0.545***
		(0.012)	(0.029)
ECH			−0.087***
			(0.011)
LG1			0.003
			(0.011)
LG2			−0.114***
			(0.012)
CPD			0.047***
			(0.010)
Prediction Markets, ECH			−0.010
			(0.035)
Prediction Markets, LG1			−0.056
			(0.036)
Prediction Markets, LG2			−0.013
			(0.039)
Prediction Markets, CPD			−0.038
			(0.034)
Tipsters, ECH			−0.162***
			(0.039)
Tipsters, LG1			−0.295***
			(0.038)
Tipsters, LG2			−0.217***
			(0.039)
Tipsters, CPD			−0.284***
			(0.038)
Observations	1,269,219	1,269,219	1,269,219
R^2	0.067	0.070	0.070
Adjusted R^2	0.067	0.070	0.070
Residual Std. Error	0.455 (df = 1269217)	0.455 (df = 1269213)	0.455 (df = 1269189)
F Statistic	90,537.880*** (df = 1; 1269217)	19,032.190*** (df = 5; 1269213)	3,294.345*** (df = 29; 1269189)

Note: *p<0.1; **p<0.05; ***p<0.01

With MAEs, absolute errors more closely approximate the pattern we might expect to observe when running down through the divisions; all forecast errors outside the EPL are significantly larger than inside the EPL, although once outside the EPL there is no obvious deterioration once moving down through the divisions.

4.4 Mean Squared Forecast Errors

Considering mean squared forecast errors (MSEs, see equation (6), again we consider the size of errors irrespective of their direction, and furthermore get an approximation to forecast variances. We find here that tipsters forecast statistically significantly worst, suggesting that they are more prone than bookmakers to large individual errors, given the absolute error outcomes observed in the previous section. There is no difference between bookmakers and prediction markets, with the largest t-statistic at -0.4 across the three events. These results are in line with Spann and Skiera (2009), who also observed similar performance between prediction markets and bookmakers, but inferior performance for tipsters.

As with absolute errors, when considering squared errors we find that the EPL has the smallest errors consistently, with all other divisions showing higher errors, LG2 much larger than the other three divisions, with ECH the smallest. Hence our hypothesis regarding information costs appears to be upheld with MSEs. This also helps shed light on whether the extra information provides signals regarding changes to relative team qualities in matches, or simply represents noise. If it only represented noise, we would expect EPL and ECH forecasts to have the largest MSEs, but as this is not the case and indeed the league with much greater media coverage (the EPL) has much smaller MSEs, it suggests that the extra information is helpful when making forecasts.

Finally, we can update Forrest et al. (2005) to the present day when considering the MSE. Forrest et al. report this for William Hill, Ladbrokes and Coral for the five years of their study (1998–2002), and we can report this for our three years, 2012, 2013 and 2014. In Table 7 in the Appendix on page 26 we report scores for individual bookmakers and prediction markets, and highlight in boldface the three bookmakers in the Forrest et al. study. For home wins Forrest et al. (2005) found Brier scores of 0.23 and higher, for draws 0.195 or higher, and for away wins 0.184 or higher. We find that for home wins scores in the range 0.225 to 0.232, for draws scores in the range 0.193 to 0.195, and for away wins scores in the range 0.194 to 0.208, suggesting that there has been a small further improvement since 2002 for home wins and the draw, and a deterioration in performance for forecasting away wins. Such a mixed performance suggests that much of the improvement in forecast performance took place during the sample of Forrest et al. when the bookmaking industry was undergoing its most severe period of deregulation and restructuring.

4.5 Profitability of Trading Strategies

The differences noted thus far have, for the most part, been remarkably small; could such small statistical differences have any economic meaning? In order to ascertain this, we conduct an exercise similar to that of Forrest et al. (2005)

and Spann and Skiera (2009), namely to appraise trading strategies based on the advertised odds.

Firstly we consider the returns that would have been made by backing each event to occur with the same bookmaker.[13] The return from such a strategy is equivalent to the overround (though of opposite sign), on average, for each bookmaker. The average (median) overround is 7.5% (mean 7.1%), and breaking down by forecast provider the smallest overround is provided by three bookmakers (5Dimes, Island Casino and Pinnacle Sports) at around 2.4–2.5%. Hence a strategy of using all bookmakers and backing all outcomes in each match would yield a negative return on investment of 7.1%. Of the prediction markets Matchbook offers an overround of 2.8% while Betfair has an average overround of 4.1%, and Betdaq offers the largest average overround at 13.1%. Betdaq's average is clearly driven by some outliers however as its median overround is just 5.8%, leaving only bookmakers offering particularly large overrounds (Interwetten, Redbet, bet-at-home and betCRIS all offer median overrounds of more than 10%). The overround is often interpreted as the transactions costs for a bookmaker, and as such is an areaon which bookmakers may compete. As a result it is perhaps unsurprising that the average overround is smaller in the EPL (5.4%) than the ECH (6.8%) or the CPD (8.6%), reflecting the more competitive betting markets.

Then we consider what returns could have been made by betting on each outcome at the best listed prices for bookmakers. With, on average, 50.6 bookmakers per match, the likelihood is higher that a positive return could be made following a strategy such as this. Forrest et al. (2005) find mainly negative returns from following a strategy similar to this amongst their five bookmakers, but in our sample such a strategy would yield a 0.37% positive return. This should be cautioned against the trading barriers in place; in order to execute a strategy such as this, accounts with positive balance would be required at all 58 bookmakers in the sample, and would need to be monitored since losses would be incurred in a number of these accounts whilst others would make profits. As a number of these betting companies are located in various parts of the world, and furthermore many refuse to payout on a number of occasions (on average in our dataset 93% of bets are paid out on by bookmakers), meaning that this strategy is likely unprofitable. Nonetheless, the difference between this and the overall overround gives a clear indication of economic differences between bookmakers.

Next, we add in the possibility of backing events on a prediction market. Allowing for this, the return on investment is slightly larger at 0.51%, although net winnings on our prediction markets are charged commission, complicating the process. As an interesting aside, we note that 94% of the time bookmakers, rather than prediction markets, offer the best odds for the home and away win, and 95% of the time for the draw. These findings are sharp contrast with Croxson and Reade (2011), who consider in-match prices for football's European Championships in 2008.

Finally, a betting strategy could involve laying events. Bettors could lay an event on a prediction market, having backed the event using a bookmaker. For

---

[13]For all bookmakers in our sample, we only use prices from bookmakers that paid out on bets for that particular match, because an absence of payout suggests that bookmakers stopped providing prices for that match at some point and cancelled all bets already placed. Hence using odds with zero payouts would give a false calculation of available returns.

example, in the Everton vs Crystal Palace match, a bettor could have backed Everton at 1.40 on BetGun, thus potentially gaining $1.40 if Everton win, for a $1 stake (lost if Everton do not win). At the same time, the bettor could have taken $1 of the liquidity available to lay Everton on Matchbook at odds of 1.39, hence risking paying out $1.39 if Everton win (a 1 cent profit), but potentially gaining $1 if Everton do not win (breaking even). For such a strategy to be profitable thus, we need there to be available liquidity to lay an event at a lower price than the odds offered by a bookmaker. It turns out that these occasions are few and far between; they occur about 2.5% of the time for the home and away win markets, and less than 1% of the time for the draw market. A strategy that traded only when such opportunities existed would net a 36% net return, but conducting this strategy at all times (backing the highest price, laying the lowest price) would have produced a net loss on investment of 57%.

Differences in returns from these strategies indicate that the differences between bookmakers and prediction markets are economically meaningful.

Next, we explore trading strategies based on the predicted outcome of our tipsters, placing a bet on whichever is the outcome forecast to be most likely. We hypothetically place our bet with the bookmaker offering the best odds, and in doing so for all matches in our sample, we find that a 0.18% loss on investment would be realised, although this number is statistically indistinguishable from zero. It remains though that positive returns could not be made.

Finally, we explore sentiment via liquidity on betting exchanges. Forrest and Simmons (2008) investigate whether odds set in the Spanish La Liga are affected by the number of fans a team has. As a proxy of fan interest in a match, we consider the total liquidity available for a match on our prediction markets. We argue this is a good proxy because if we assume a fixed proportion of all football fans wish to gamble on match outcomes, then matches involving teams with more fans will have more people seeking to place bets.[14] Furthermore, matches between teams with greater numbers of fans will attract more media interest, and hence likely more betting interest. We regress the odds reported by bookmakers on the liquidity available on the prediction markets in what is essentially a panel estimation.[15] These regressions, for home win, draw and away win odds, are reported in Table 4; the units for total liquidity is thousands of pounds.[16] The regression results suggest that for all outcomes, listed prices do respond to the amount of liquidity available. Odds for the home win increase by 0.002 for each extra thousand pounds available, for the draw by 0.009, and for the away win by 0.023; all of these effects are strongly statistically significant.

Hence the more interest there is in a match, as represented by the amount of liquidity available on prediction markets, the more favourable odds are offered by bookmakers, particularly on the away win outcome. This may reflect the impact of sentiment as considered by Forrest and Simmons (2008); for matches

---

[14]Our data is a snapshot of availability at kick-off, yet we argue this is representative of more general interest in a match.

[15]Hence it is possible that our regression results suffer from unobserved heterogeneity bias due, especially as unobserved characteristics of football matches will likely influence how much money is bet upon them. With a sample size of 385,555 however, we appeal to large sample results.

[16]The match with the most liquidity is a match between Manchester City and Cardiff City on January 18 2014, where £644,376 was available. On average £11,290 was available per match, although the distribution is skewed by the small number of matches with very large amounts available; the median available liquidity is £2,207.

Table 4: Sentiment Regression Results

	Dependent variable:		
	odds1	odds2	odds3
	(1)	(2)	(3)
Constant	2.344***	3.389***	3.378***
	(0.002)	(0.001)	(0.003)
Total Liquidity (000s)	0.002***	0.009***	0.023***
	(0.00004)	(0.00002)	(0.0001)
Observations	385,555	385,555	385,556
R^2	0.005	0.306	0.169
Adjusted R^2	0.005	0.306	0.169
Residual Std. Error	0.897 (df = 385553)	0.440 (df = 385553)	1.767 (df = 385554)
F Statistic	2,067.735*** (df = 1; 385553)	169,982.900*** (df = 1; 385553)	78,492.240*** (df = 1; 385554)

Note:			*p<0.1; **p<0.05; ***p<0.01

with more fans, more favourable odds are offered. Other explanations might be the impact of heavy trading on bookmakers; if liquidity on prediction markets proxies activity with bookmakers, this may be the response of bookmakers to high demand — to increase the attractiveness of odds on the less likely outcome (away wins happen much less frequently than home wins). It may also reflect endogeneity; anecdotally it is known that bookmakers use prediction markets (Betfair, in particular) to cover exposure that they might have in their own books; with bigger matches may arise larger exposures to outcomes and hence reason for bookmakers to trade on prediction markets.

Nonetheless, assuming that this result reflects that bookmakers adjust odds in response to higher levels of demand, it appears consistent with the results mentioned earlier on the overround; if all odds are becoming more favourable, then the overround must be falling in size. The overround can be viewed as the transactions costs imposed by bookmakers; the cost of securing a guaranteed return. It would seem natural that for matches where there is greater demand that competition is keener between bookmakers and hence overrounds are smaller. A question of greater interest would be whether individual bookmakers differ in this behaviour; our regression in Table 4 forces all bookmakers to act in the same way, yet it may be that some increase prices much more than others, yielding potentially more profitable trading strategies. While it is impractical to report a regression with dummy variables allowing the constant and slope terms to vary with bookmakers, we can at least note the outcome of an F test of the significance of those coefficients. The F test of the joint insignificance of dummies allowing the constant and slope to vary in the regression in Table 4 has a statistic of 10.816 for the home win, 166.66 for the draw, and 24.672 for the away win, which is compared to the F distribution with 114 degrees of freedom in the numerator and 385440 in the denominator, which gives a 95% critical value of 1.22, and a 0.1% critical value of 1.46, and as such the test is emphatically rejected, suggesting there are important differences between bookmakers in their response to high market activity. For the away win, 14 bookmakers increase prices by less than the average, while four by more than it, suggesting sufficient variation that could be exploitable in more sophisticated trading strategies that may yield returns higher than the 0.37% or 0.51% returns reported earlier.

5   Conclusions

In this paper we have used forecasts of sports events from bookmakers, prediction markets and tipsters to analyse hypotheses regarding forecast accuracy, information and predictability. Overall, our results seem consistent with the information hypothesis once we control for the professional/amateur distinction in the lowest league in our dataset, with increased costs having an impact on forecast quality, and further our results appear consistent with the idea that commercial pressures do drive better forecasts, as economic theory would predict. We find evidence consistent with the information-costs explanation for the favourite-longshot bias, at least on the part of bettor behaviour if not bookmaker pricing. We also find evidence which could be consistent with the 'wisdom of crowds' hypothesis. More specifically, we find in general that the forecasts from tipsters are markedly inferior to the forecasts embodied in odds set by bookmakers, and prices listed on prediction markets for the sporting outcomes we consider.

A   Appendix: Tables

References

J.E. Berg, F.D. Nelson, and T.A. Rietz. Prediction market accuracy in the long run. *International Journal of Forecasting*, 24(2):285–300, 2008.

G.W. Brier. Verification of forecasts expressed in terms of probability. *Monthly Weather Review*, 78(1):1–3, 1950.

K. Croxson and J.J. Reade. Exchange vs. Dealers: A High-Frequency Analysis of In-Play Betting Prices. Discussion Papers 11-19, Department of Economics, University of Birmingham, January 2011.

B. Deschamps and O. Gergaud. Efficiency in Betting Markets: Evidence From English Football. *The Journal of Prediction Markets*, 1(1):61–73, 2012.

D. Forrest and R. Simmons. Sentiment in the Betting Markets on Spanish Football. *Applied Economics*, 40:119–126, 2008.

D.K. Forrest and R. Simmons. Forecasting Sport: The Behaviour and Performance of Football Tipsters. *International Journal of Forecasting*, 16:317–331, 2000.

D.K. Forrest, J. Goddard, and R. Simmons. Odds-Setters As Forecasters: The Case of English Football. *International Journal of Forecasting*, 21:551–564, 2005.

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

P. Goodwin and G. Wright. Improving judgmental time series forecasting: A review of the guidance provided by research. *International Journal of Forecasting*, 9(2):147–161, 1993.

Name	Number of matches forecast
Tipsters	7622
10Bet	7596
12BET	7462
188BET	7611
32Red Bet	7567
5Dimes	7602
888sport	7606
Bestbet	7615
bet365	7610
bet-at-home	7604
Betclic	7481
BetCRIS	7563
Betdaq (back)	7608
Betdaq (lay)	7608
Betfair (back)	7620
Betfair (lay)	7620
Betfred	7593
Betgun	7581
Betinternet	6472
BetOnline	4712
Betrally	1292
BetRedKings	7567
Betsafe	7596
Betsson	7601
BetVictor	7613
Betway	7512
Boylesports	6737
bwin	7439
Coral	7554
Dafabet	7568
DOXXbet	7602
Expekt	7453
FortunaWin	5024
Instant Action Sports	2593
Intertops	4598
Interwetten	7577
Island Casino	7603
Jetbull	7615
Ladbrokes	7595
LEON Bets	7595
Luxbet	6409
MarathonBet	4912
Matchbook (back)	1027
Matchbook (lay)	1027
myBet	7599
NordicBet	7587
Noxwin	7026
Oddsring	2515
Paddy Power	7607
Paf	6137
Pinnacle Sports	7585
Redbet	7469
SBOBET	7604
Sportingbet	7585
SportsBetting	2700
Stan James	7589
The Greek	1712
Tipico	7005
Titanbet	7401
TonyBet	4911
totepool	7571
Unibet	7596
William Hill	7520
Winner	1764
youwin	6343

Table 5: Bookmakers and prediction markets in dataset, along with frequency of occurrence in dataset.

Bookmaker	Coefficient	Standard Error	t-statistic	p-value
10Bet	0.029006	0.008664	3.348	0.000815 ***
12BET	0.031660	0.008698	3.640	0.000273 ***
188BET	0.030171	0.008660	3.484	0.000494 ***
32Red Bet	0.030567	0.008671	3.525	0.000423 ***
5Dimes	0.029947	0.008663	3.457	0.000546 ***
888sport	0.029038	0.008662	3.352	0.000801 ***
Bestbet	0.028738	0.008659	3.319	0.000905 ***
bet365	0.029576	0.008661	3.415	0.000638 ***
bet-at-home	0.027190	0.008662	3.139	0.001696 **
Betclic	0.028170	0.008693	3.240	0.001194 **
BetCRIS	0.028174	0.008672	3.249	0.001160 **
Betdaq (back)	0.026546	0.008661	3.065	0.002178 **
Betdaq (lay)	0.029878	0.008663	3.449	0.000563 ***
Betfair (back)	0.029495	0.008662	3.405	0.000661 ***
Betfair (lay)	0.031372	0.008665	3.621	0.000294 ***
Betfred	0.028009	0.008665	3.232	0.001228 **
Betgun	0.028233	0.008668	3.257	0.001125 **
Betinternet	0.025768	0.008988	2.867	0.004143 **
BetOnline	0.028452	0.009756	2.916	0.003543 **
Betrally	0.048335	0.015338	3.151	0.001626 **
BetRedKings	0.030303	0.008671	3.495	0.000475 ***
Betsafe	0.029270	0.008664	3.378	0.000730 ***
Betsson	0.029351	0.008663	3.388	0.000704 ***
BetVictor	0.029126	0.008660	3.363	0.000770 ***
Betway	0.028738	0.008685	3.309	0.000937 ***
Boylesports	0.026171	0.008903	2.940	0.003286 **
bwin	0.028971	0.008704	3.328	0.000874 ***
Coral	0.028211	0.008675	3.252	0.001146 **
Dafabet	0.031826	0.008671	3.670	0.000242 ***
DOXXbet	0.028894	0.008663	3.335	0.000852 ***
Expekt	0.026465	0.008701	3.042	0.002352 **
FortunaWin	0.023204	0.009586	2.421	0.015494 *
Instant Action Sports	0.051050	0.011766	4.339	0.000014 ***
Intertops	0.023490	0.009823	2.391	0.016789 *
Interwetten	0.023598	0.008669	2.722	0.006486 **
Island Casino	0.029489	0.008662	3.404	0.000664 ***
Jetbull	0.028869	0.008659	3.334	0.000857 ***
Ladbrokes	0.026039	0.008664	3.005	0.002653 **
LEON Bets	0.029728	0.008664	3.431	0.000601 ***
Luxbet	0.026203	0.009009	2.909	0.003630 **
MarathonBet	0.026299	0.009644	2.727	0.006394 **
Matchbook (back)	0.064485	0.017410	3.704	0.000212 ***
Matchbook (lay)	0.100189	0.024030	4.169	0.000031 ***
myBet	0.028292	0.008663	3.266	0.001092 **
NordicBet	0.029176	0.008666	3.367	0.000761 ***
Noxwin	0.031190	0.008817	3.538	0.000404 ***
Oddsring	0.041240	0.011894	3.467	0.000526 ***
Paddy Power	0.028055	0.008661	3.239	0.001200 **
Paf	0.025994	0.009104	2.855	0.004301 **
Pinnacle Sports	0.030093	0.008667	3.472	0.000516 ***
Redbet	0.028264	0.008696	3.250	0.001154 **
SBOBET	0.031137	0.008662	3.595	0.000325 ***
Sportingbet	0.027587	0.008667	3.183	0.001458 **
SportsBetting	0.031482	0.011604	2.713	0.006670 **
Stan James	0.027732	0.008666	3.200	0.001374 **
The Greek	0.071389	0.013711	5.207	0.000000 ***
Tipico	0.025054	0.008823	2.840	0.004515 **
Titanbet	0.032227	0.008714	3.698	0.000217 ***
TonyBet	0.023747	0.009645	2.462	0.013812 *
totepool	0.027143	0.008670	3.131	0.001745 **
Unibet	0.027953	0.008664	3.226	0.001254 **
William Hill	0.029695	0.008683	3.420	0.000627 ***
Winner	0.048444	0.013550	3.575	0.000350 ***
youwin	0.028456	0.009031	3.151	0.001627 **

Table 6: Output from regression of correct forecasts on forecast type. Reference group is tipsters, hence all differences relative to tipsters.

	2012			2013			2014		
Forecast Source	Home win	Draw	Away win	Home win	Draw	Away win	Home win	Draw	Away win
Tipsters	0.329	0.268	0.281	0.298	0.231	0.258	0.283	0.228	0.236
10Bet	0.226	0.195	0.196	0.231	0.194	0.206	0.224	0.193	0.194
12BET	0.225	0.194	0.194	0.230	0.195	0.206	0.225	0.193	0.194
188BET	0.226	0.195	0.196	0.231	0.194	0.206	0.225	0.193	0.194
32Red Bet	0.225	0.194	0.196	0.231	0.194	0.206	0.225	0.193	0.194
5Dimes	0.225	0.194	0.196	0.231	0.194	0.206	0.224	0.193	0.194
888sport	0.226	0.195	0.197	0.231	0.194	0.207	0.225	0.193	0.194
Bestbet	0.226	0.195	0.196	0.231	0.195	0.206	0.225	0.193	0.194
bet365	0.226	0.195	0.196	0.231	0.194	0.206	0.224	0.193	0.193
bet-at-home	0.226	0.195	0.196	0.231	0.194	0.207	0.225	0.193	0.194
Betclic	0.226	0.195	0.197	0.231	0.195	0.206	0.224	0.193	0.194
BetCRIS	0.226	0.196	0.196	0.231	0.194	0.207	0.225	0.192	0.194
Betdaq (Back)	0.226	0.195	0.196	0.231	0.195	0.206	0.227	0.194	0.196
Betdaq (Lay)	0.226	0.195	0.196	0.232	0.195	0.206	0.225	0.193	0.195
Betfair (Back)	0.226	0.195	0.196	0.231	0.194	0.206	0.224	0.193	0.194
Betfair (Lay)	0.225	0.194	0.195	0.231	0.194	0.206	0.224	0.193	0.194
Betfred	0.226	0.195	0.197	0.231	0.194	0.206	0.225	0.193	0.194
Betgun	0.226	0.195	0.197	0.231	0.194	0.206	0.225	0.193	0.194
Betinternet	0.226	0.195	0.196	0.231	0.195	0.205	0.228	0.196	0.195
BetOnline	0.226	0.191	0.193	0.232	0.194	0.204	0.222	0.196	0.197
Betrally							0.220	0.194	0.193
BetRedKings	0.226	0.195	0.196	0.231	0.194	0.206	0.225	0.193	0.194
Betsafe	0.226	0.195	0.197	0.231	0.195	0.207	0.225	0.193	0.194
Betsson	0.226	0.195	0.197	0.231	0.195	0.207	0.225	0.193	0.194
BetVictor	0.225	0.195	0.195	0.231	0.194	0.206	0.224	0.193	0.194
Betway	0.227	0.194	0.196	0.231	0.194	0.206	0.225	0.193	0.194
Boylesports	0.225	0.195	0.196	0.231	0.195	0.206	0.228	0.199	0.193
bwin	0.227	0.195	0.196	0.232	0.197	0.204	0.225	0.193	0.194
Coral	0.225	0.195	0.197	0.231	0.194	0.207	0.225	0.193	0.194
Dafabet	0.225	0.195	0.196	0.230	0.195	0.206	0.225	0.193	0.194
DOXXbet	0.226	0.195	0.197	0.232	0.195	0.207	0.226	0.193	0.194
Expekt	0.226	0.195	0.197	0.231	0.195	0.206	0.224	0.193	0.194
FortunaWin				0.231	0.194	0.207	0.225	0.193	0.194
Instant Action Sports	0.226	0.193	0.195	0.228	0.195	0.189	0.213	0.187	0.187
Intertops	0.227	0.191	0.194	0.232	0.195	0.203	0.223	0.196	0.197
Interwetten	0.227	0.195	0.198	0.232	0.195	0.207	0.227	0.193	0.195
Island Casino	0.225	0.194	0.196	0.231	0.194	0.206	0.224	0.193	0.194
Jetbull	0.226	0.195	0.196	0.231	0.195	0.206	0.225	0.193	0.194
Ladbrokes	0.226	0.195	0.196	0.232	0.194	0.207	0.226	0.193	0.195
LEON Bets	0.225	0.194	0.196	0.231	0.194	0.206	0.224	0.193	0.194
Luxbet	0.228	0.197	0.191	0.231	0.194	0.206	0.225	0.194	0.194
MarathonBet				0.231	0.194	0.207	0.224	0.193	0.194
Matchbook (Back)				0.227	0.203	0.190	0.217	0.190	0.188
Matchbook (Lay)				NaN	NaN	NaN	0.210	0.188	0.185
myBet	0.226	0.195	0.197	0.231	0.195	0.206	0.225	0.193	0.194
NordicBet	0.226	0.195	0.197	0.231	0.195	0.206	0.225	0.193	0.194
Noxwin	0.225	0.195	0.196	0.232	0.195	0.204	0.225	0.193	0.194
Oddsring				0.217	0.190	0.210	0.224	0.193	0.194
Paddy Power	0.225	0.195	0.196	0.231	0.195	0.207	0.225	0.193	0.194
Paf	0.228	0.199	0.187	0.231	0.195	0.206	0.225	0.193	0.193
Pinnacle Sports	0.226	0.195	0.197	0.231	0.194	0.206	0.224	0.193	0.194
Redbet	0.226	0.196	0.196	0.230	0.195	0.206	0.225	0.193	0.194
SBOBET	0.226	0.195	0.196	0.231	0.194	0.206	0.224	0.193	0.194
Sportingbet	0.226	0.194	0.196	0.231	0.194	0.206	0.225	0.193	0.194
SportsBetting				0.238	0.194	0.195	0.222	0.196	0.197
Stan James	0.226	0.195	0.197	0.231	0.194	0.206	0.224	0.193	0.194
The Greek	0.226	0.191	0.193	0.208	0.198	0.174	0.205	0.166	0.180
Tipico	0.226	0.195	0.196	0.231	0.195	0.206	0.226	0.191	0.197
Titanbet	0.226	0.195	0.196	0.229	0.195	0.205	0.225	0.193	0.193
TonyBet				0.231	0.194	0.207	0.225	0.192	0.194
totepool	0.226	0.195	0.197	0.231	0.194	0.206	0.225	0.193	0.194
Unibet	0.226	0.195	0.196	0.232	0.195	0.206	0.225	0.193	0.194
William Hill	0.226	0.195	0.196	0.231	0.195	0.206	0.226	0.193	0.195
Winner							0.225	0.192	0.194
youwin	0.226	0.195	0.196	0.229	0.195	0.211	0.223	0.192	0.194

Table 7: Mean squared forecast errors, or Brier scores, for all forecast sources for each year of our sample.

D.F. Hendry and M.P. Clements. *Forecasting Economic Time Series*. Cambridge University Press, Cambridge, 1998.

D.F. Hendry and M.P. Clements. *Forecasting Non-stationary Economic Time Series*. MIT Press, Cambridge, Mass., 2001.

Marek Hlavac. *stargazer: LaTeX/HTML code and ASCII text for well-formatted regression and summary statistics tables*. Harvard University, Cambridge, USA, 2014. URL `http://CRAN.R-project.org/package=stargazer`. R package version 5.0.

W. Hurley and L. McDonough. A Note on the Hayek Hypothesis and the Favorite-Longshot Bias in Parimutuel Betting. *The American Economic Review*, 85(4):949–955, September 1995.

R.J. Hyndman and G. Athanasopoulos. *Forecasting: Principles and Practice*. www.otexts.com, 2012. URL `https://www.otexts.org/book/fpp`.

S. Kou and M.E. Sobel. Forecasting the Vote: A Theoretical Comparison of Election Markets and Public Opinion Polls. *Political Analysis*, 12(3):277–295, 2004.

M. Lawrence, P. Goodwin, M. O'Connor, and D. Önkal. Judgmental forecasting: A review of progress over the last 25 years. *International Journal of Forecasting*, 22(3):493–518, 2006. ISSN 0169-2070. doi: http://dx.doi.org/10.1016/j.ijforecast.2006.03.007.

A. Leigh and J. Wolfers. Competing Approaches to Forecasting Elections,: Economic Models, Opinion Polling and Prediction Markets. *Economic Record*, 82(258):325–340, September 2006.

P.F. Pope and D.A. Peel. Information, Prices and Efficiency in a Fixed-odds Betting Market. *Economica*, pages 323–341, 1989.

R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2014. URL `http://www.R-project.org/`.

D. Rothschild. Forecasting Elections. *Public Opinion Quarterly*, 73(5):895–916, 2009.

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

M. A. Smith, D. Paton, and L. Vaughan Williams. Market Efficiency in Person-to-Person Betting. *Economica*, 73:673–689, November 2006.

E. Snowberg and J. Wolfers. Explaining the Favorite-Longshot Bias: Is It Risk-Love or Misperceptions? *Journal of Political Economy*, 118(4):723–746, 2010.

M. Spann and B. Skiera. Sports forecasting: A comparison of the forecast accuracy of prediction markets, betting odds and tipsters. *Journal of Forecasting*, 28(1):55–72, 2009. ISSN 1099-131X. doi: 10.1002/for.1091. URL `http://dx.doi.org/10.1002/for.1091`.

J. Surowiecki. *The Wisdom of Crowds: Why the Many Are Smarter Than the Few and How Collective Wisdom Shapes Business, Economies, Societies and Nations*. Little, Brown, 2004. ISBN 0-316-86173-1.

L. Vaughan Williams. Information Efficiency in Betting Markets: A Survey. *Bulletin of Economic Research*, 51(1):1–30, 1999.

L. Vaughan Williams and D. Paton. Why Is There a Favourite-Longshot Bias in British Racetrack Betting Markets? *The Economic Journal*, pages 150–158, 1997.

L.M. Woodland and B.M. Woodland. Market Efficiency and the Favorite-Longshot Bias: The Baseball Betting Market. *Journal of Finance*, pages 269–279, 1994.