

# *Polls to probabilities: comparing prediction markets and opinion polls*

Article

Accepted Version

Creative Commons: Attribution-Noncommercial-No Derivative Works 4.0

Reade, J. J. ORCID: <https://orcid.org/0000-0002-8610-530X> and Vaughan Williams, L. (2019) Polls to probabilities: comparing prediction markets and opinion polls. *International Journal of Forecasting*, 35 (1). pp. 336-350. ISSN 0169-2070 doi: <https://doi.org/10.1016/j.ijforecast.2018.04.001> Available at <https://centaur.reading.ac.uk/77452/>

It is advisable to refer to the publisher's version if you intend to cite from the work. See [Guidance on citing](#).

To link to this article DOI: <http://dx.doi.org/10.1016/j.ijforecast.2018.04.001>

Publisher: Elsevier

All outputs in CentAUR are protected by Intellectual Property Rights law, including copyright law. Copyright and IPR is retained by the creators or other copyright holders. Terms and conditions for use of this material are defined in the [End User Agreement](#).

[www.reading.ac.uk/centaur](http://www.reading.ac.uk/centaur)

**CentAUR**

Central Archive at the University of Reading

Reading's research outputs online



# Polls to Probabilities: Comparing Prediction Markets and Opinion Polls

J James Reade  
University of Reading and  
Programme for Economic Modelling (EMoD),  
Institute for New Economic Thinking at the Oxford Martin School

Leighton Vaughan Williams  
Nottingham Business School

May 18, 2018

## Abstract

The forecasting of election outcomes is a hugely popular activity, and not without reason: the outcomes can have significant economic impacts, for example on stock prices. As such, it is economically important, as well as of academic interest, to determine the forecasting methods that have historically performed best. However, the forecasts are often incompatible, as some are in terms of vote shares, while others are probabilistic outcome forecasts. This paper sets out an empirical method for transforming opinion poll vote shares into probabilistic forecasts, and then evaluates the performances of prediction markets and opinion polls. We make comparisons along two dimensions, bias and precision, and find that converted opinion polls perform well in terms of bias, while prediction markets are good for precision.

*Keywords:* Forecasting models, Information and knowledge, Elections, Voting behavior, Prediction markets.

## 1 Introduction

Electoral outcomes can have significant economic impacts, and as such, it is economically important, as well as of academic interest, to evaluate forecasting methods for elections. This article compares the forecasts of US election outcomes from a number of prediction markets and opinion polls.

Opinion polls are surveys of the voting intentions of a sample of voters, while prediction markets allow participants to trade contracts of which the value is contingent on some particular outcome occurring. The largest commercial prediction markets over this period produced predicted probabilities, whereas opinion polls produced projected vote shares. Thus, we begin by developing an empirical method for converting vote shares into outcome probabilities, then compare the two sources of forecasts by considering both the bias and the precision, which are reflected elegantly in the commonly-used Brier score, or mean squared error.

We consider all opinion polls from two common aggregators of polling information (*Real Clear Politics* and *Pollster*), and look at three well-known prediction markets: Intrade, Betfair and

Search frequency of Gallup, Intrade and Betfair on Google

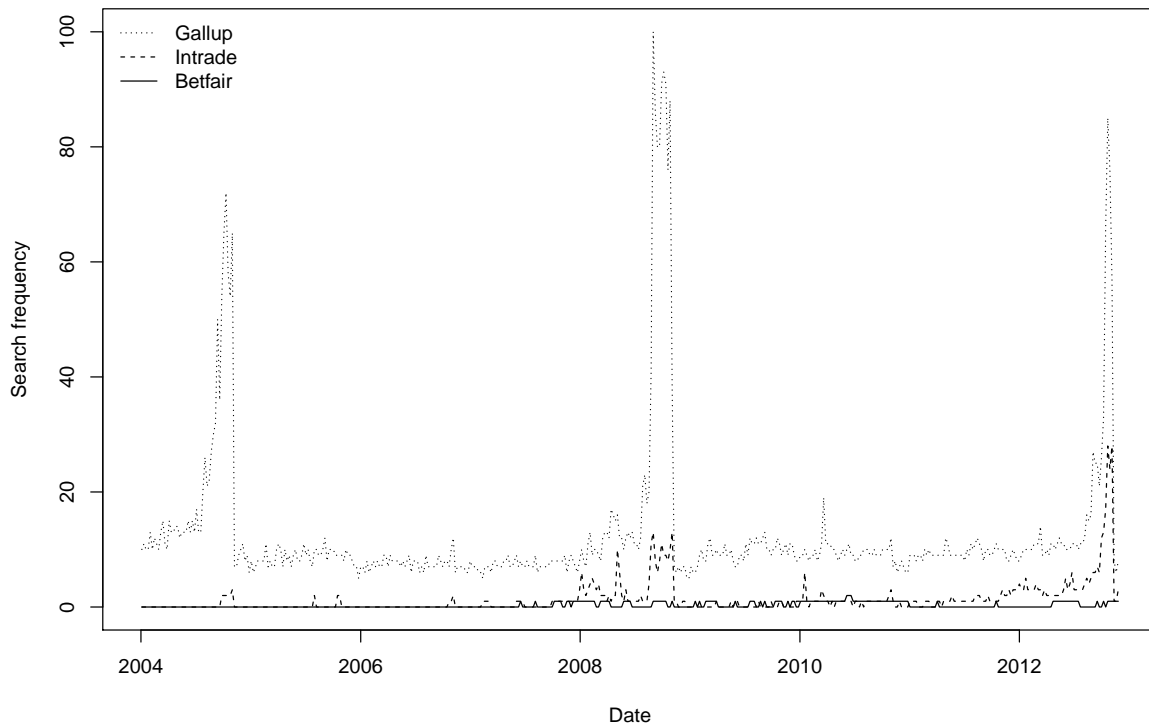


Figure 1: Relative frequencies of searches for Gallup, Intrade and Betfair on Google. 100 refers to the week with the most searches of Gallup, Intrade or Betfair since 2004. Source: Google Trends.

Iowa Electronic Markets (IEM, henceforth). There have already been several academic investigations into the performances of opinion polls and prediction markets (e.g. [1]), and we contribute to this growing body of evidence.

While prediction markets have a long and rich history [2], their internet-based electronic variants have been garnering increasing amounts of attention over recent years, with 2012 arguably marking the first year in which one of them, Intrade, was in the news regularly alongside traditional polling information. Figure 1 provides some background on this, as it reports the relative search frequencies on Google in the US for Gallup (a historic polling company), Intrade and Betfair (the two best known prediction markets). Google search information is available for the period since early 2004, and hence, three election cycles, those in 2004, 2008 and 2012, are distinctly visible from the spikes in search volumes, most pertinently for Gallup, but also for Intrade. Gallup registered more than sixty times as many searches as Intrade around the 2004 election, but only ten times as many in 2008, and three times as many in 2012.

The 2012 election, and to a lesser extent the 2004 and 2008 elections, also bore witness to a distinct divergence between the two most commonly known prediction markets, Intrade and Betfair, with Republican presidential candidates tending to be priced more favourably on Intrade than on Betfair. This divergence was noted by [3], who analysed the behaviour of one particular trader who lost around \$4m, apparently in manipulating the market price in favour of the Republican candidate, Mitt Romney. A clear implication of the [3] paper is that there

was a distinct bias on Intrade relative to Betfair (and also to the Iowa Electronic Markets); investigating the bias forms part of our description of each market.

Thus, this paper enables an enhanced comparison of opinion polls and prediction markets by providing an empirical method for converting polls into probabilities. We also build on previous studies that have compared opinion polls and prediction markets by considering a range of prediction markets, rather than a single one. Section ?? introduces our datasets, giving some details about the nature of each source of forecasts, and the information we have for each source. Section ?? discusses our methodology for transforming vote-share polls into probabilistic forecasts, and subsequently for appraising each forecast source based on its bias and precision. Section ?? provides the results, comparing our prediction markets to each other and to opinion polls, and Section ?? concludes.

## 2 Data

Data are fundamental to this investigation, and hence, we begin by introducing our data and sources, before discussing our methods. As our primary aim is to convert opinion polling data into a form that can be compared more readily to prediction market data, we start by introducing our opinion polling data in Section ?. We then introduce prediction market data from Betfair (Section ?), Intrade (Section ?) and IEM (Section ?). It is important to be clear at the outset: our selection of the period 2008–2012 is dictated by data availability. While polling information is available for much longer than the selected sample period, the particular form of data that we have from Betfair and Intrade is not.

### 2.1 Opinion polls

We collect all polls reported by Real Clear Politics (<http://www.realclearpolitics.com/>) for elections in 2010 and 2012, and augment nationwide presidential election polls with those reported by Pollster (<http://elections.huffingtonpost.com/pollster>).<sup>1</sup> This yields 271 elections, with the 2012 Presidential race yielding the greatest number of polls by some distance, at 1609, of which 1098 are at the state level. The data contain the polled vote shares for all candidates in an election, along with the date range the poll was conducted over, the sample size and the type of sample constructed (almost exclusively one of: adults, registered voters or likely voters), and the pollster (plus an indication of whether that pollster is known to favour Republicans or Democrats).

Figure ?? plots the absolute error for polls by the number of days until the election. For both types of candidates, 89% of polls are within ten points of the actual outcome, 38% are within three points, and 12% are within a single point.

The difference in the units on the vertical axis of Figure ?? relative to, say, Figure ??, reveals a problem with comparing prediction markets to opinion polls: opinion polls in general quote vote shares, whereas prediction markets in general report probabilistic forecasts of election outcomes. Nonetheless, we can view the distribution of opinion poll vote shares as a probability distribution for election outcomes, and generate probabilities of outcomes; Section ?? outlines our method for converting one to the other.

---

<sup>1</sup>For simplicity, we ignore primary polling, as this often relates to multiple-candidate elections, which adds complications.

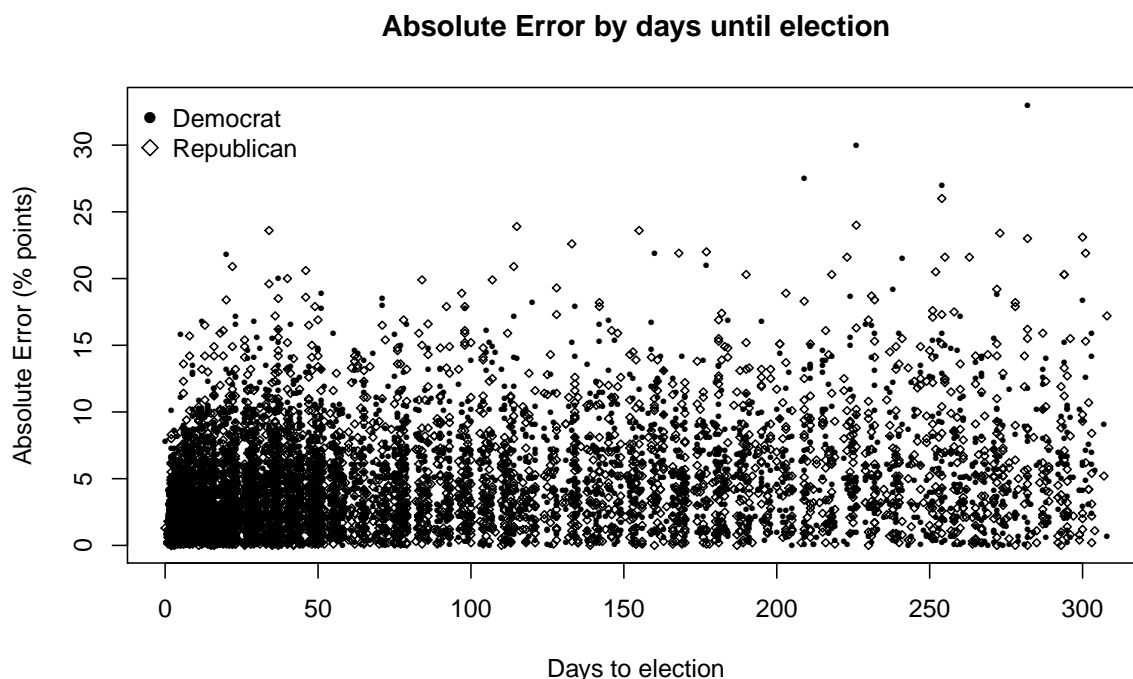


Figure 2: Absolute difference between each polled vote share and each outcome in 2010 and 2012, where Democrat and Republican errors are distinguished by different points. Source: Pollster and Real Clear Politics.

When considering opinion polls as forecasts, we note that polls only report (imperfectly) the intended voting behaviour at a particular point in time, usually some time before the election actually occurs.<sup>2</sup> Nonetheless, opinion polls are indeed forecasts in the sense that they provide information, in advance, on a variable to be revealed at a future point in time (election day), and we treat them as such.

## 2.2 Prediction markets

### 2.2.1 Betfair

Betfair is a UK-based betting exchange, or prediction market. It enables willing bettors to both back events to happen (buy contracts whose payout is contingent on the outcome of events) and lay them (sell contracts). It specialises in common sporting markets like horse racing and football, but also offers markets for political events. Betfair has claimed to have over four million customers and a turnover of £50m a week.<sup>3</sup> It matches those willing to back events with those willing to lay them at particular prices, and charges a commission of between 2 and 5% on net winnings.<sup>4</sup> Bets on Betfair are denominated in decimal odds, which represent the multiple of

<sup>2</sup>? considers an alternative, less commonly publicised type of poll: vote expectation surveys, which ask voters which candidate they believe will win the election, rather than which candidate they will vote for.

<sup>3</sup>See “Betfair moves operations to Gibraltar”, *The Telegraph*, 9 March 2011 ([goo.gl/JQS0Er](http://goo.gl/JQS0Er)).

<sup>4</sup>The commission is between 2 and 5% for UK-based traders, but higher for traders based in a number of other countries (see <http://www.betangel.com/forum/viewtopic.php?f=15&t=8856>).

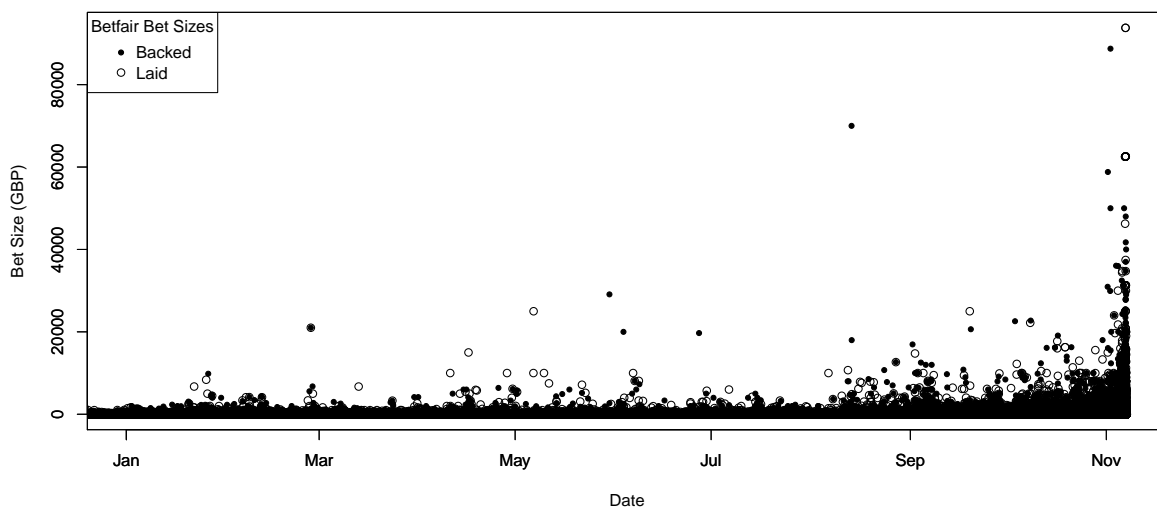


Figure 3: Sizes of bets matched on Betfair by type: backing an event or laying it.

the amount bet that must be exchanged if the event occurs. The reciprocal of decimal odds is commonly interpreted as the implied probability of occurrence of the event being bet.

Our dataset includes every single trade on Betfair in the four-year election cycle between the Presidential elections in 2008 and 2012, and includes a number of markets, not just in the US, but more widely. We have information on the precise second at which a bet was placed, the average price at which it was placed, how much was bet, whether it was to back or lay, and which event. This amounts to a total of 249,756 observations over the cycle, and £40.7m traded (gross) over 80 different contracts. We are not aware of any previous study of US election prediction markets and polls that has included data from Betfair, and hence we provide a unique contribution to the literature in this regard.

The largest single transaction was for £93,796.87, placed at 23:45:51 UTC on election day, 6th November 2012, laying Obama to win. The second largest transaction came less than an hour earlier and similarly was to lay Obama for sixteen pounds less than the largest trade.<sup>5</sup> Both of these trades were at an average price of 1.24, which implies a probability of that event occurring of about 80%. The largest transaction to back a bet came four days previously on 2nd November at 07:04:29, and was for £88,733.56, backing Obama to win. These transactions are shown in context in Figure ??; the hollow circles are bets laid and the solid circles are bets backed.

There are often multiple markets for similar outcomes on Betfair. For example, for the election, there are often markets for both the next president and the winning party; thus, once the nominations for the two major parties are known, these two markets are essentially identical.<sup>6</sup> Figure ?? shows that these two markets are essentially identical on election day.

Of the 249,756 trades that we have information on, 51.4% of bets are laying the event to occur, while the remainder are backing the event. Accordingly, the average bet to back an event is slightly larger, at £167.87 (median £15), than that to lay the event (£158.46, median £9.99). Bets laid are marginally less focussed on the Presidential election, with 53.4% of them being

<sup>5</sup>As UTC is five hours ahead of Eastern Time, these trades did occur while voting was still open.

<sup>6</sup>Except for the negligible risk that a candidate may be forced to withdraw before election day.

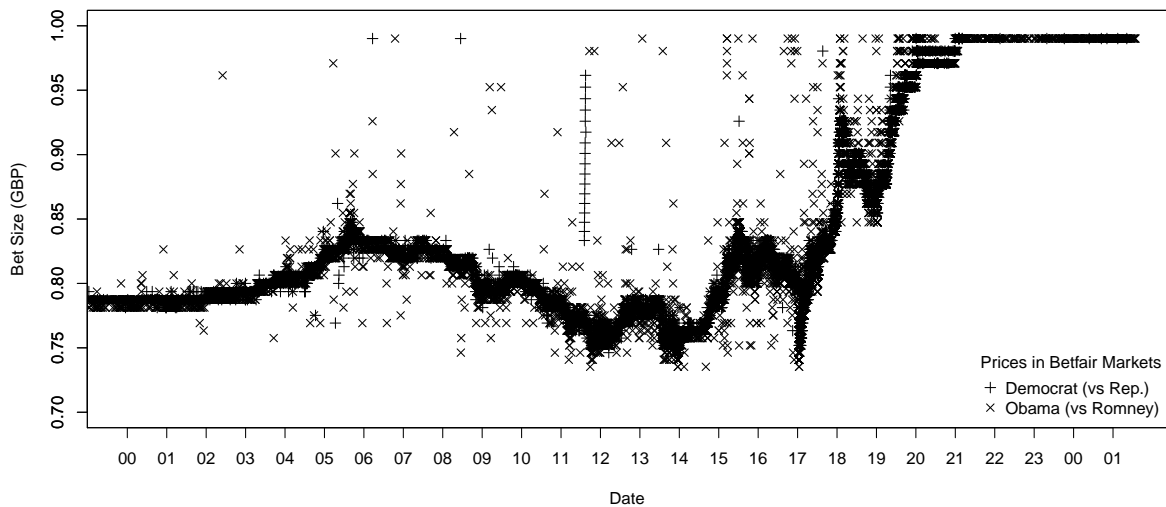


Figure 4: Different prices for bets matched on election day in two markets on Betfair: the markets for the winning party and the next president.

focused on the candidate or party markets, compared to 59.5% of bets backed. The mean odds that events are backed at is 19.87 (implied probability 5%), although the median is just 1.93 (implied probability 51.8%), while the mean odds that events are laid at is 28.69 (median 2.4).<sup>7</sup>

### 2.2.2 Intrade

Intrade was a betting exchange, or prediction market. Based in Ireland, its focus was on US political events, and its trading mechanism differed somewhat from that of Betfair. On Intrade, willing participants agreed to trade contracts denominated in US dollars, paying out \$10 to the buyer if the event in question occurred, and zero otherwise. These contracts are priced between one cent and \$10, with the price commonly being interpreted as the implied probability of the event occurring.

For Intrade, we have data on all contracts bought and sold, with the buyer/seller ID (though no further characteristics), price and quantity, specific contract, and timestamp. <sup>?</sup> focused on the Next President market over the final 15 days of the election campaign, but over the entire election cycle since late 2008, the dataset contains 952,387 observations on 1,106 different contracts bought or sold.<sup>8</sup>

On average, each trade saw 26.5 contracts exchanged (although the median is just five contracts, suggesting a rather skewed distribution), at an average price of \$2.99, implying a 29.99% probability of occurrence of the event being traded (the median price is \$2.30). Of all of the 25

<sup>7</sup>Betfair lists prices in decimal odds, and caps prices at 1000 (implying a probability of 0.001). The 75th percentile of the odds distribution is 5.5, suggesting that the distribution is very skewed by bets placed at very high odds, which explains the large gap between the mean and median decimal odds.

<sup>8</sup>Our first observation is the first contract sold in the 2012 winning party market, which occurred prior even to the 2008 election on 25th September of that year. Note that, like Betfair, there are often multiple markets for the same event on Intrade; for example, markets for the winning party and the winning candidate at the US presidential election.



million contracts traded in our dataset, 30.4% of them were paid out (i.e., the event occurred), suggesting a reasonably calibrated market.

One controversy in 2012 was the allegation that the Intrade prices for Romney and Obama were manipulated in an attempt to aid the Republican candidate, Romney. The impact of such trading can be observed in Figure ??, where Intrade consistently had the lowest price of our three prediction markets for Obama throughout the second half of 2012. Thus, it might be anticipated that Intrade would perform worse than the other markets, and possibly than the polls too. However, it should be noted that the Obama/Romney market constituted only around 21% of all trades in our dataset, with the others being scattered across various other elections, including primaries. Furthermore, considering particularly large trades, only four of the 35 trades of 9,000 or more contracts are for either Obama (one) or Romney (three) to win the presidential election. We mention this not to rule out the possibility of politically motivated market manipulation, but to point out that our dataset for Intrade is much broader than the Obama/Romney market. ? used Intrade data corrected for a favourite-longshot bias in his comparison with bias-corrected opinion poll data for US elections over a similar time period to our comparison.

### 2.2.3 Iowa electronic markets

The Iowa electronic markets (IEMs) were one of the earliest electronic prediction markets, predating both Betfair and Intrade by going back to the late 1980s, but remain somewhat smaller in nature, potentially due at least in part to the restrictions placed on trading: the maximum initial balance a trader can deposit is \$500. Traders buy or sell contracts that are worth \$1 if the event occurs and zero otherwise, and these contracts trade at anywhere between a cent and a dollar. ? used IEMs for their comparison of prediction market and opinion poll forecasts, and found IEMs to outperform opinion polls, particularly over longer forecast horizons.

We have collected all of the IEMs between 2008 and 2012, to coincide with our data from Betfair and Intrade. This yields 19,068 observations and encompasses 12 markets: Congress, House and Senate for 2010 and 2012, the Republican Convention in 2012, the Iowa Caucus of 2012, the 2010 Florida Senate election, and the 2012 Presidential election.<sup>9</sup> IEM data are at a daily frequency and in summary format: high price, low price, average price and last price, along with the numbers of contracts exchanged and the dollar volume exchanged. This raises an important methodological point pertaining to aggregation; our IEM data are aggregated to summary daily information, whereas our Betfair and Intrade data remain fully disaggregated. As we wish to avoid reducing the information content in our datasets, we decided not to aggregate our Betfair or Intrade information; we note that the empirical evidence regarding forecasting using aggregate or disaggregate variables is mixed (?).

A total of just under two million contracts were traded on IEM over the electoral cycle, while the total dollar amount traded was just over half a million dollars, reflecting the fact that the average price of a contract is 29 cents. This tallies closely with the empirical likelihood of success of a contract over the period, which was 25%.

Figure ?? plots the evolution of the market price on IEM for the Obama to win market during the second half of 2012, over which period the IEM price fluctuates between the Betfair and

---

<sup>9</sup>We exclude the vote share markets as they do not yield a probability of election outcome interpretation as the other markets do. These markets correspond to 2,207 observations.

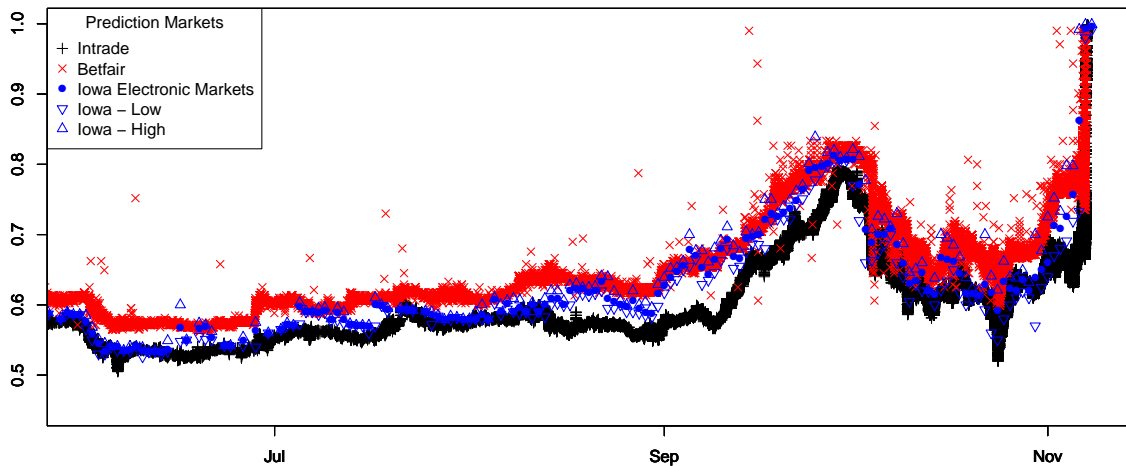


Figure 5: All three prediction markets for Obama to win, June–November 2012.

Intrade prices. It appears that IEM moves with Betfair in upswings in the Obama price, but is aligned with Intrade otherwise, suggesting that new information that enters the market is processed in a similar manner by IEM and Betfair. ? note research that has been conducted into potential geographic biases in the processing of information that could sustain the kinds of price differentials that they note between the UK-based Betfair and the US-focused Intrade, and the IEM price may lend partial credence to this view, as geographically a knowledge of IEM is limited largely to North America (see Figure ??), and the IEM price tracks the Intrade price for some periods.

### 3 Methodology

This section sets out our methodological approach to the comparison of opinion polls and prediction markets. We consider the bias and precision of the two methods, but need to convert opinion polls from vote share projections into probabilistic forecasts. As part of this, we bias-correct opinion polls, and hence, it is appropriate to also bias-correct prediction market forecasts in order to facilitate more insightful comparisons.

Ideally, opinion polls would be random samples of voting intentions; on the other hand, prediction markets are self-selected samples of agents trading contracts whose payoffs are contingent on voting outcomes. Strictly speaking, neither are forecasts, although both are used commonly as forecasts. The theory of enlightened voters (?) might suggest that opinion polls during election campaigns will display more variance than prediction markets, because opinion polls reflect voters who are becoming enlightened, while prediction markets tend to be self-selected samples of those who are already “enlightened” regarding the likely outcome of the election. Conversely, prediction markets may be influenced by ‘irrational exuberance’ and other behavioural traits that increase their volatility; we have already noted (in the introduction) the existence of market manipulation. However, it seems unlikely for many reasons that opinion polls are truly random samples, not least because a number of pollsters may have political leanings and because it

seems plausible that herding instincts and other behavioural biases mean that voting intentions may be influenced by a knowledge of the intentions of other voters.

Fundamentally, the model of information transmission for opinion polls is centralised with the polling company, whereas for prediction markets it is decentralised amongst market participants. Thus, we might assert that prediction markets ought to provide better forecasts than opinion polls by appealing to the so-called Hayek hypothesis (??), which suggests that markets can work efficiently even when the participants' knowledge of the environment or other participants is limited (see also ?). Of the prediction markets in our sample, Intrade and Betfair do not restrict the financial exposure that participants can accumulate, while IEM does; thus, if we assume that the prediction market mechanism allows information to be conveyed in both prices and quantities, it is possible that IEM restricts this process.

? argue against the superiority of prediction markets, suggesting that the information on election outcomes that is available from opinion polls is much broader than simply the polled vote share released, not least because of the enlightened voter idea. ? present evidence that polls, adjusted for such information by means of bias correction, provide more accurate forecasts of election outcomes than prediction markets. The findings of ? are in contrast to those of ?, who compare a richer set of prediction markets (the same set covered in this paper) to opinion polls and find that prediction markets perform better. However, ? do not adjust prediction market forecasts, suggesting that market prices do not need correction. This constitutes an assumption that prediction markets are efficient, but without testing this formally. ? does bias-correct Intrade prediction market data when comparing with opinion polls, and concludes in favour of Intrade against opinion polls.

### 3.1 Conversion and bias correction

In practice, the distinction between opinion polls and prediction markets is the units that forecasts are produced in; opinion polls report projected vote shares for candidates, whereas prediction markets report probabilities of particular outcomes.

The outcome of election  $i$  can be described in terms of the vote share  $V_{ij}$  for candidate  $j$ , where  $\sum_j V_{ij} = 1$ , or alternatively as a probability of victory of candidate  $j$ , and hence  $P_{ij} = P(\max_k (V_{ik}) = V_{ij})$ . As we convert poll shares into implied probabilities of an election victory, we refer to  $P_{ij}$  exclusively. We denote the forecast probability from forecast source  $s$  as  $P_{sij} \in [0, 1]$ , where  $s$  is one of the prediction markets or opinion polls. Our forecast appraisal compares  $P_{sij}$  to the observed election outcome, denoted as  $O_{ij}$ :

$$O_{ij} = \begin{cases} 1 & \text{if } \max_k (V_{ik}) = V_{ij}, \\ 0 & \text{otherwise.} \end{cases} \quad (1)$$

Hence,  $O_{ij}$  is one if candidate  $j$  wins election  $i$ , and zero otherwise.

To be able to compare opinion poll outcomes, which are in vote shares, with prediction market outcomes, which are prices, expressed most commonly in terms of probabilities, we need to transform the opinion poll outcomes into probabilities regarding the outcome of the election. As we express the vote shares for candidate  $j$  in election  $i$  as  $V_{ij}$ , we can denote the projected vote share for that candidate in an opinion poll conducted at time  $h$  as  $\widehat{V}_{ij|h}$ . A conversion to outcome probabilities involves estimating the probability of election  $i$  being won by candidate  $j$ :  $P(\max_k V_{ik} = V_{ij} | \mathcal{I}_{ij})$ , where  $\mathcal{I}_{ij}$  is the information available regarding candidate  $j$  in election  $i$  that is *contained in a reported opinion poll*.

Various methods exist for turning opinion poll vote shares into outcome probabilities; ? develops a theoretical parametric approach, while another alternative is non-parametric density estimation using historical poll outcomes. In recent elections, Nate Silver’s 538 website has been converting polling information into probabilistic forecasts, although details regarding the methods employed are undisclosed, as would be expected with a commercial endeavour.<sup>10</sup> We follow an empirical approach, employing a regression method for creating probabilities; regression methods as simple as ordinary least squares (OLS) allow the estimation of the probabilities associated with observed binary outcomes such as election outcomes ( $O_{ij}$ ). However, OLS estimation with a binary dependent variable can produce estimated probabilities that lie outside the unit interval, and hence, we employ an alternative model that mitigates this problem, the probit model.

A probit regression model with  $O_{ij}$  as dependent variable estimates the probability of election victory conditional on a number of explanatory factors, which we denote by the vector  $X_{ij}$ . We write the model as:

$$\mathbb{P}\left(V_{ij} = \max_k V_{ik} \mid X_{ij}\right) = \Phi(\beta X_{ij}), \quad (2)$$

where  $\beta$  is a vector of coefficients and  $\Phi$  is the cumulative density function of the standard normal distribution.  $X_{ij}$  includes information from each poll for candidate  $j$  in election  $i$ . We include candidate  $j$ ’s polled vote share in election  $i$ , but we also include the opinion poll vote shares of other candidates in the election, as well as information on whether the organisation conducting the poll is identified as sympathising with one or another political party, the sample size and nature of sample (likely voters, registered voters, adults), and the number of days until an election. It seems reasonable to assume that such additional aspects of pollsters will provide systematic information on both the likelihood of victory of a candidate being polled and the polled vote share, and indeed, this tallies with the method proposed by ? for correcting poll outcomes prior to a comparison with prediction markets. ? do not need to convert to probabilities, as they consider vote-share IEMs (about 10% of all IEMs over our time period; we exclude these markets from our study).

We then use the estimated coefficients from Eq. (??) to construct fitted, or predicted values. We calculate  $\mathbb{P}(V_{ij} = \widehat{\max_k V_{ik}} \mid X_{ij})$ , which are the estimated implied probabilities of election outcomes from our opinion polls. We limit ourselves to only Republican and Democrat candidates, and run regressions separately for each party. We run separate regressions for Republican and Democrat candidates, which makes it easy for us to allow coefficients to vary between the two types of candidate.

In summary, our method replaces the election outcome as dependent variable in the bias correction linear regression methods in Eq. (??) with an indicator variable that takes a value of one if that candidate wins,  $O_{ij}$ . This allows us to generate an estimate of the probability of a candidate’s victory, which is the form in which prediction market forecasts are produced, using the vote share predictions produced by opinion polls.

A linear regression with the outcome poll share as the dependent variable and the opinion poll share as an explanatory variable is a bias correction method, and can be written as:

$$V_{ij} = \gamma \widehat{V}_{ij|h} + \beta X_{ij} + u_{ij}. \quad (3)$$

This kind of regression of outcomes on forecasts is referred to commonly as a Mincer-Zarnowitz regression (?), and is used to evaluate forecasts. In Eq. (??),  $\gamma$  captures the direct relationship

---

<sup>10</sup>See for example <http://goo.gl/gv7xQ1>, last accessed 16 October 2014.

between opinion poll vote shares and actual vote shares, while  $X_{ij}$  is a matrix of other explanatory variables that might capture biases in opinion polls, and  $\beta$  is a parameter that reflects their potential influences on outcomes that are not reflected in the opinion poll vote share  $\widehat{V}_{ij|h}$ . Section ?? sets out the explanatory variables that make up  $X_{ij}$ , including things like the poll sample size and type, the length of time before an election, and the potential partisan nature of a pollster.

The conventional statistical measure of the bias, for some estimator  $\widehat{\theta}$  for a parameter  $\theta$ , is  $E(\widehat{\theta}) - \theta$ ; in expectation, an unbiased estimator is equal to the true value of the parameter, and so  $E(\widehat{\theta}) - \theta = 0$ . In terms of forecasts, the question is whether a forecast  $\widehat{P}_{ij}$  made for outcome  $O_{ij}$  in advance of its occurring is equal in expectation to the true value:  $E(\widehat{P}_{ij}) - O_{ij}$ .

If opinion polls are unbiased and reflect all available information at time  $h$ , then  $\gamma = 1$  and  $\beta = 0$  in Eq. (??), so that  $E(V_{ij}) = \widehat{V}_{ij|h}$ .

We can then define bias-corrected poll vote shares as  $\widetilde{V}_{ij}$ , and calculate them according to:

$$\widetilde{V}_{ij} = \widehat{\gamma}\widehat{V}_{ij|h} + \widehat{\beta}X_{ij}, \quad (4)$$

where  $\widehat{\gamma}$  and  $\widehat{\beta}$  are ordinary least squares estimates of the parameters  $\gamma$  and  $\beta$ . Hence, as such, we can think of the opinion poll probabilities that we produce using Eq. (??) as being bias corrected, as well as simply converted into appropriate units for comparisons with prediction market probabilistic forecasts.

One problem when evaluating the bias is the specification of an alternative hypothesis in order to quantify any bias that may exist. While unbiasedness implies the functional form  $E(\widehat{P}_{sij}) = P_{ij}$ , biasedness could take many forms, and hence, formal testing may be affected by the specification of an alternative. We begin by considering graphical expressions in order to detect any unusual patterns of bias, then consider regression methods in order to quantify any bias.

A common method of determining the existence of bias in academic analyses of the betting industry involves running a regression of the form:

$$O_{ij} = \alpha + \beta P_{ijt} + e_{ijt}, \quad (5)$$

where  $P_{ijt}$  refers to a prediction at time  $t$  for candidate  $j$  in election  $i$ . This is another Mincer-Zarnowitz regression, similarly to Eq. (??), with the difference that here the forecast is a probabilistic forecast, and the outcome is a binary variable that takes a value of one if the event in question occurred. Hence, estimated using least squares, Eq. (??), is a linear probability model. The null hypothesis is that the forecasts are unbiased (and efficient), and hence  $E(O_{ij}) = E(P_{sij})$ . In order for such to be the case, we need  $\alpha = 1 - \beta = 0$ . After conducting the regression model described in Eq. (??), we can test this hypothesis using an  $F$ -test.

Thus, bias correction is possible, as with opinion polls. We can define a bias-corrected probabilistic forecast to be  $\widetilde{P}_{ijt}$ :

$$\widetilde{P}_{ijt} = \widehat{\alpha} + \widehat{\beta}P_{ijt}. \quad (6)$$

The alternative hypothesis here is that the bias is linear in form, which is clearly restrictive. Nonetheless, it is likely that the null of unbiasedness will still be rejected if the actual bias is non-linear, particularly for large sample sizes like ours. We will consider the remaining bias

in bias-corrected polls when comparing the corrected prediction market probabilities to the converted (and hence corrected) opinion poll projections.

When considering forecasts, the bias is conventionally measured using the forecast error, which is defined in terms of the forecast and the actual, or true, value:

$$\hat{e}_{sij} = O_{ij} - \hat{P}_{sij}. \quad (7)$$

A forecast with a zero error in expectation is unbiased.

As we consider a very large number of forecasts, we must consider methods for summarising the errors. At the same time, such a large number of forecasts enables a better evaluation of the expected value of forecasts from particular sources, relative to the actual outturns. For prediction source  $s \in \{\text{bet fair, intrade, iem, polls}\}$ , the mean forecast error is:

$$ME = \hat{e}_{sij} = O_{ij} - \hat{P}_{sij}, \quad (8)$$

Thus, the mean error of all forecasts made by each source  $s$  is a measure of the bias.

### 3.2 Precision

Precision refers to the spread of forecasts, independent of the true value or outcome. As such, the variance of forecasts is a measure. Conventionally, the reciprocal of the variance is used as the precision of a measure. This has the property that the measure of precision falls as the forecasts become less precise (and hence, the variance increases). In the limit, a measure with an infinite (undefined) variance has zero precision, and point forecasts with a zero variance have infinite precision, loosely speaking.

Thus, for precision, we calculate the reciprocal of the variance of the forecasts from each forecast source  $s$  for a given election  $i$  and candidate  $j$ :

$$PR_s = \frac{1}{\sum_i \sum_j \left( \hat{P}_{ij} - \mathbb{E} \left( \hat{P}_{ij} \right) \right)^2}. \quad (9)$$

Note that this measure is independent of the true values or outcomes of forecast events. We calculate sample variances for each contract (candidate and election) for each prediction market and for opinion polls, and take the average of these (hence the bar notation in Eq. (??)).

### 3.3 Bias and precision

The bias and precision in a forecast are both summarised conveniently in the mean squared error measure of forecast errors:

$$MSE_s = \sum_i \sum_j (\hat{e}_{sij}^2), \quad (10)$$

$$= ME_s^2 + \sum_i \sum_j \left( \hat{P}_{ij} - \mathbb{E} \left( \hat{P}_{ij} \right) \right)^2. \quad (11)$$

The mean squared error is the sum of the squared bias and the variance of the forecast, which forms the precision via Eq. (??).

In binary outcome events like elections, the mean square forecast is often referred to as the Brier score  $\beta$ , which is a probability scoring rule. Unlike other probability scoring rules such as the logarithmic scoring rule, the Brier score should be minimised by the best forecast model.

Hence, we consider the mean error, Mincer-Zarnowitz regressions, and graphical methods for evaluating the forecast bias, and the variance of forecasts for considering the precision. We then consider the mean squared forecast error as a summary measure of both.

## 4 Results

### 4.1 Converting polls to probabilities

The first two columns of Table ?? report the results of the probit regression that converts the polled vote share into implied probabilities of an election victory for the candidates (Eq. (??)). Each column corresponds to a particular candidate, Republican or Democrat, respectively. The first row of the table is the constant term, while the subsequent three rows show the impact on the likelihood of victory of additional percentage points for Democratic, Republican and Independent candidates in a released opinion poll. The coefficients show the anticipated signs: an extra point for a Republican (or Democratic) candidate increases their likelihood of winning; conversely, an extra point for an Democratic (Republican) candidate reduces the likelihood of a Republican (Democratic) candidate winning. The effects are similar in magnitude for each candidate, but the impact of an independent candidate is not symmetric across the two types of candidate: an increased polling share for an independent candidate reduces the likelihood of a Republican candidate winning, but not that of a Democratic candidate winning. The impact of an independent candidate's vote share is about three fifths of that of a Democratic candidate's vote share.

Beneath the impact of candidate vote shares are a number of other explanatory variables that may capture biases in opinion polls. The coefficients differ with the type of sample, which may be likely voters, registered voters, or simply adults. Polls of both likely and registered voters increase the likelihood of a Republican victory relative to polls of adults, whereas the opposite is true for the probability of a Democratic victory. All four coefficients are borderline significant, but may potentially signify biases in the construction of opinion polls; thus, including terms such as these in the process of turning opinion polls into probabilities, bias-corrects polls. These coefficients suggest that the closer to election day that a poll is conducted, the larger (smaller) is the probability of a Democratic (Republican) victory. Finally, the impact of pollsters being aligned to a particular party is shown by the final two rows of the table: the likelihood of a Republican victory is greater if we are observing a poll from a partisan pollster, whereas a Democrat victory is less likely. Both effects are independent of the identified political leaning of the pollster itself. They imply a bias in partisan pollsters, but not necessarily towards the party they are identified with.

It is difficult to interpret these coefficients due to the dependent variable being the likelihood of victory. Thus, we attempt to get a better sense of the biases that our probit models correct whilst generating probabilities of victory by also reporting in the third and fourth columns of Table ?? a linear regression of the polled vote shares on all the explanatory variables that are used to generate probabilities; that is, estimates of Eq. (??).

Fewer coefficients are significant in these regressions than in the probit regressions. The constant term is the average vote share, with all other explanatory variables being set to zero, and

Table 1: Probit regressions for creating probabilities for election outcomes.

	<i>Dependent variable</i>			
	Republican win	Democrat win	Republican vote share	Democrat vote share
	<i>probit</i>	<i>probit</i>	<i>OLS</i>	<i>OLS</i>
	(1)	(2)	(3)	(4)
Constant	3.550*** (0.625)	-4.548*** (0.637)	49.437*** (1.063)	32.582*** (1.074)
Republican opinion poll vote share	0.091*** (0.006)	-0.082*** (0.007)	0.447*** (0.012)	-0.286*** (0.012)
Democrat opinion poll vote share	-0.196*** (0.007)	0.213*** (0.007)	-0.443*** (0.011)	0.656*** (0.011)
Independent opinion poll vote share	-0.082*** (0.007)	0.001 (0.010)	-0.309*** (0.015)	-0.467*** (0.015)
Days until election	-0.001*** (0.0002)	0.001*** (0.0002)	0.0004 (0.0005)	0.001* (0.0005)
Sample size	-0.0005*** (0.0001)	0.0004*** (0.0001)	-0.0001 (0.0001)	0.0001 (0.0001)
Likely voters	0.850* (0.473)	-0.994** (0.474)	-1.865*** (0.584)	0.198 (0.590)
Registered voters	0.965** (0.472)	-1.088** (0.475)	-1.235** (0.577)	0.526 (0.583)
Identified Democratic pollster	0.638*** (0.072)	-0.715*** (0.074)	0.767*** (0.161)	-1.294*** (0.163)
Identified Republican pollster	0.484*** (0.143)	-0.541*** (0.147)	0.073 (0.352)	-0.535 (0.356)
Observations	4,025	4,025	4,025	4,025
$R^2$			0.666	0.781
Adjusted $R^2$			0.665	0.781
Log likelihood	-1,297.573	-1,250.410		
Akaike information criterion	2,615.146	2,520.819		
Residual std. error (df = 4015)			3.773	3.811
$F$ -statistic (df = 9; 4015)			887.808***	1,593.581***

Note: \* $p < 0.1$ ; \*\* $p < 0.05$ ; \*\*\* $p < 0.01$ .



as such is not particularly informative, although it becomes a little more meaningful when taken in conjunction with the two subsequent rows, the polled vote shares for the candidates. The Republican coefficient in the Republican (Democrat) regression, here 0.447 ( $-0.286$ ), is the marginal effect of an extra Republican polling point, whereas the Democrat coefficient in the Republican (Democrat) regression, here 0.443 ( $0.656$ ), is the marginal effect of an extra Democrat polling point. Taking the example of a 50/50 poll, with the constant term (but all other variables set to zero), this would suggest that the corrected polled vote shares should be 51.1% for the Democrat candidate, and 49.6% for the Republican candidate.

If a poll is constructed of likely (registered) voters, this reduces the corrected Republican vote share by almost two (one) points, suggesting that these two types of polls over-estimate the likely Republican vote but have no impact on the Democrat vote. A pollster who is identified as a Democrat partisan underweights the Republican candidate by 0.7 points and overweights the Democrat candidate by 1.3 points, while there is no significant impact of a Republican pollster (potentially reflecting the smaller number of these pollsters in our dataset).

Figure ?? provides a graphical representation of both regression models. The top row shows the conversion of vote shares into probabilities for Republican (left) and Democratic (right) candidates; the vertical axis is the probability produced by the probit model, and on the horizontal axis is the polled vote share. For both types of candidate, a polled vote share below 30% implies essentially a zero probability of winning, while a polled vote share above 60% implies almost certain victory. In between these points there is a steep slope from zero to one.

We can now revisit the 2012 presidential campaign using these probabilities, and Figure ?? plots all opinion poll vote shares for the Republican (crosses) and Democrat (diamonds) candidates on the top panel, and our model's implied probabilities of election outcome on the bottom panel.<sup>11</sup> Although the polled vote shares are always within the range of 37 to 54%, with some crossover between the candidates, the implied probabilities from the model suggest that the Democratic candidate was always the strong favourite to win the election, with the lowest probability of winning being 54%, and the average probability at 90%. This suggests that the additional information contained within each released poll, that can be included in regression models such as those in Table ??, adds significantly to the consideration of likely poll outcomes, and that hence bias correction has value. It suggests that the polled vote share of a candidate provides limited forecast information on likely election outcomes.

The bottom row of plots in Figure ?? shows the corrections for various factors that bias the polled vote shares, where the horizontal axis is the original polled vote share and the vertical axis is the corrected vote share. These plots are equivalent to plotting the actual values against the fitted values for our vote share linear regression model in Eq. (??), as reported in Table ?. Both plots include a 45-degree line that shows the extent to which vote shares are impacted by our regression method. Both plots show similar impacts of the correction of polls, namely a reversion to the centre: polled vote shares above 50% are pulled back towards 50%, while those below 50% are pushed up towards it. This would appear to suggest that raw opinion polls are somewhat too extreme in their prediction of likely outcomes. This conclusion certainly seems to be in line with the findings of ?, who suggest that the use of raw opinion poll numbers is unwise, due to such known biases as those shown by our regressions here. The value added by our method's inclusion of additional explanatory variables enables further biases to be corrected rather than simply remaining in the residuals, as would be the case in a more basic bias correction method.

<sup>11</sup>Where the probability of winning the election is the probability of having the largest vote share.

## 4.2 Bias

Figures ??–?? provide graphical representations of the bias. In each plot, the horizontal axis represents the forecast outcome probability (and hence is on the unit interval), while the vertical axis represents the frequency of the outcomes (the proportion of all forecasts of a particular probability that turn out to occur). On such a plot, unbiased forecasts will be scattered around the  $45^\circ$  line, as this represents probabilistic forecasts occurring as frequently as predicted.<sup>12</sup> From each plot, we see that the points plotted are some distance from the  $45^\circ$  line, indicating some degree of bias. The red bars in Figures ?? and ?? (for Betfair and Intrade, respectively) reflect the numbers of bets agreed at that implied probability forecast, and the relative frequency can be found on the right-hand scale in both cases. These two plots also display distinct patterns for the apparent bias in prediction market forecasts: a flat relationship below about 40% and above 60%, and a steep relationship between 40 and 60%. This suggests that any event that either market identifies as more than about 60% likely to occur almost always occurs, while any event that either market identifies as being less than about 40% likely to occur very rarely occurs. This distinct relationship is not visible for either IEM (Figure ??) or opinion polls (Figure ??), with these two plots displaying more linear relationships around the  $45^\circ$  line, indicative of less bias.

---

<sup>12</sup>For example, 50% of forecasts that put the probability of an event occurring at 50% turn out to be correct.

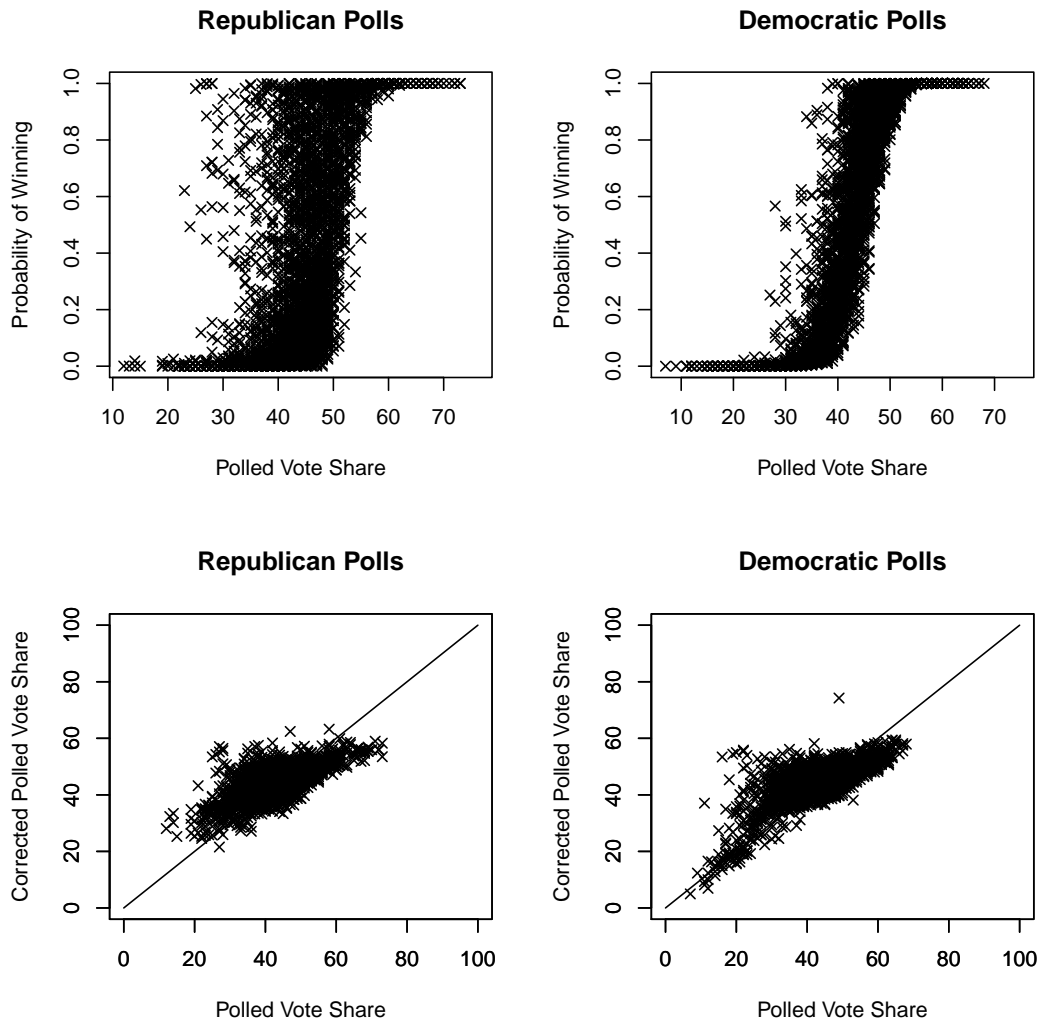


Figure 6: Plots related to opinion poll data. The top row represents the conversion of Republican (left) and Democrat (right) polls into probabilities of outcomes. The bottom row plots the polled vote shares against corrected polled vote shares for Republican (left) and Democrat (right) polls. Source: Real Clear Politics and Pollster.com.

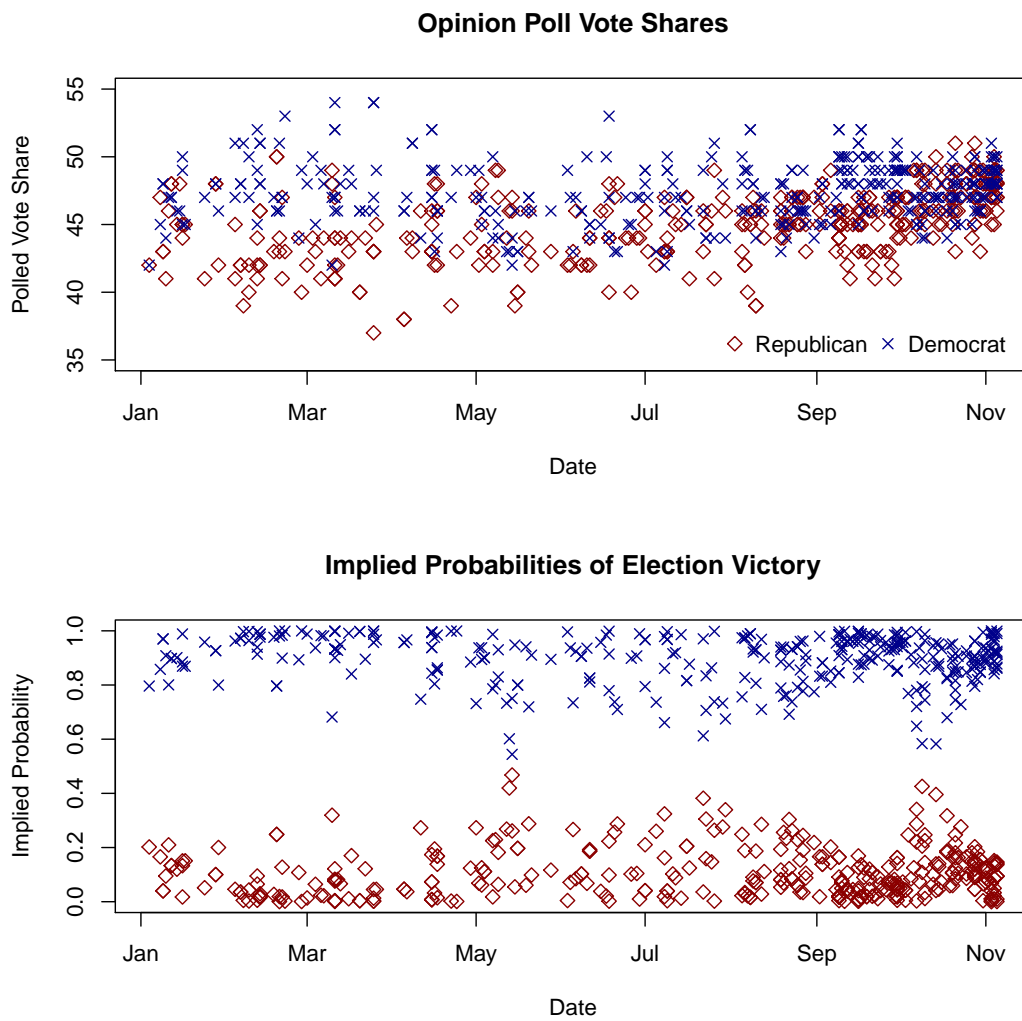


Figure 7: Probabilities implied from polls. Source: Pollster.

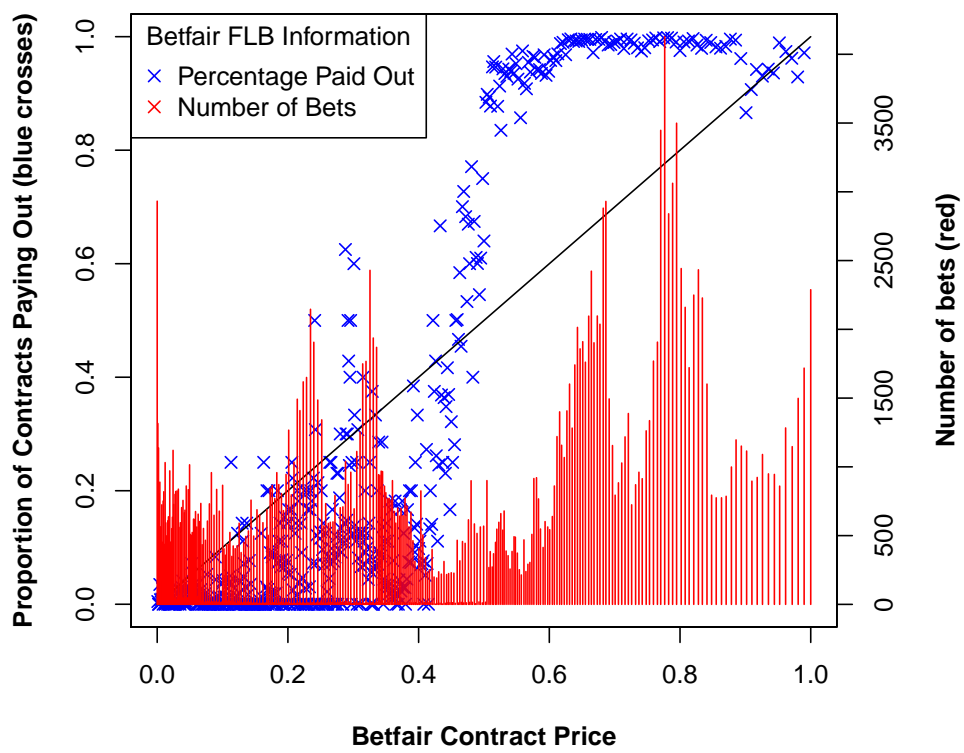


Figure 8: Betfair FLB plot.

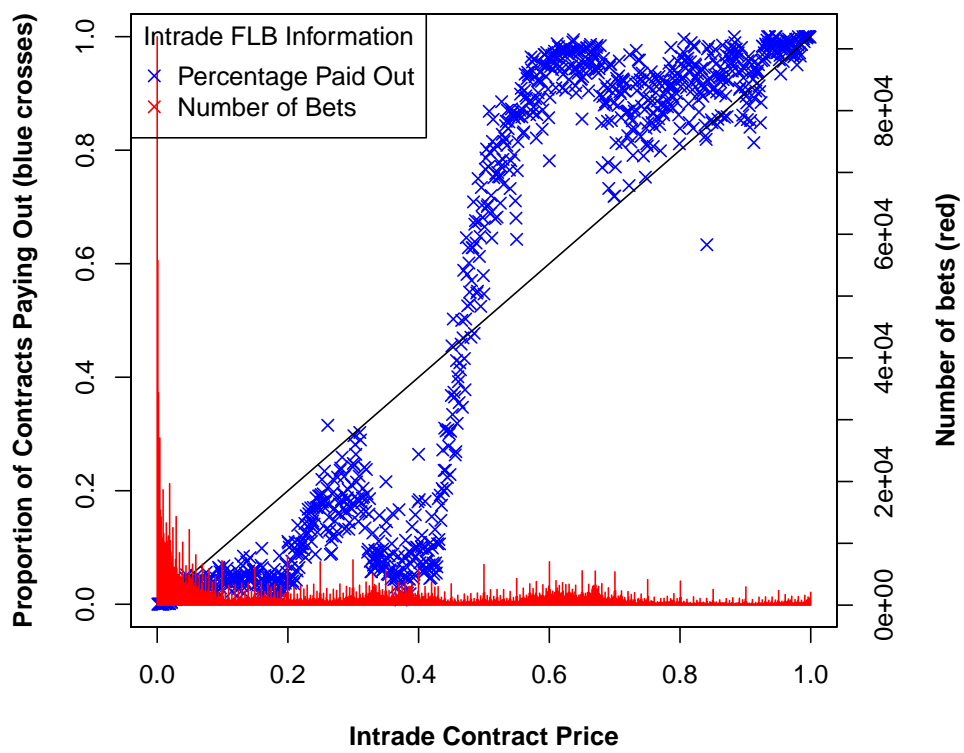


Figure 9: Intrade FLB plot.

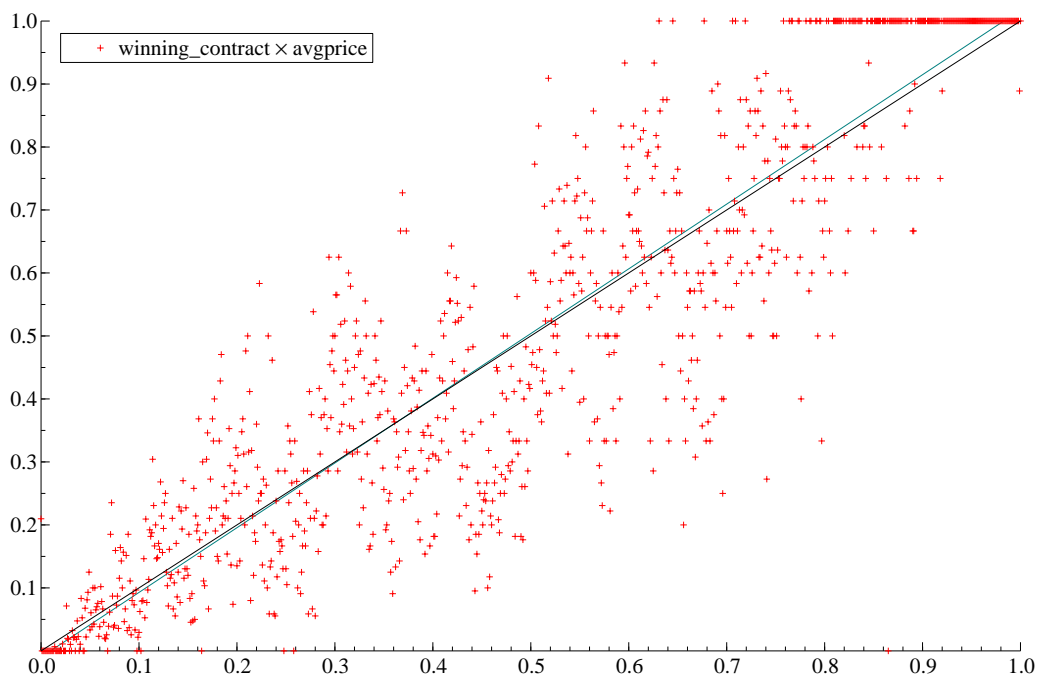


Figure 10: Iowa Electronic Markets FLB plot.

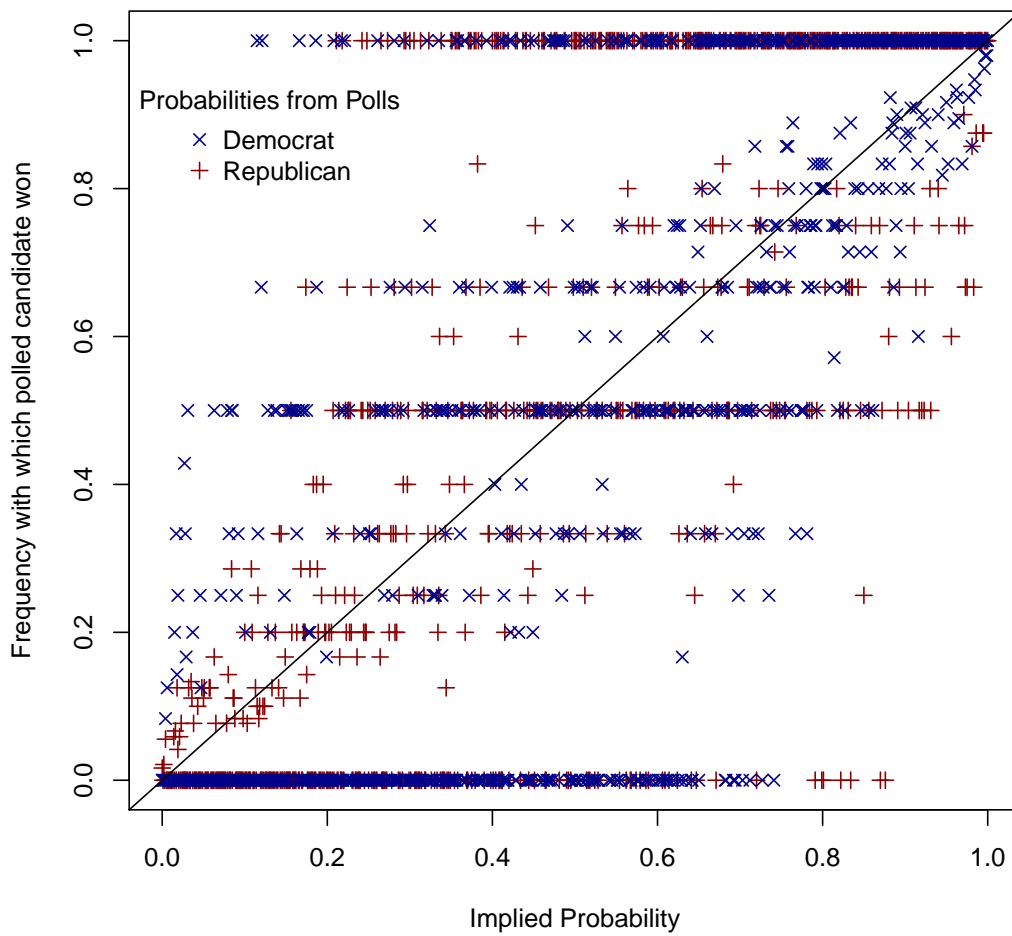


Figure 11: Polls FLB Plot.



Table 2: Bias and precision: various measures for evaluating forecasts.

Measure	Betfair	Betfair (corrected)	Intrade	Intrade (corrected)	IEM (low)	IEM (high)	IEM (mean)	IEM (last)	Polls (D)	Polls (R)
MSE	0.0875	0.0668	0.0902	0.0854	0.1725	0.1738	0.1304	0.1397	0.0933	0.0958
Bias	-0.0573	0.0000	-0.0041	0.0000	0.0718	0.0622	-0.0407	-0.0299	0.0012	-0.0024
Variance	0.0874	0.1830	0.0816	0.1217	0.0589	0.0625	0.0691	0.0668	0.1255	0.1181
Precision	11.444	5.4654	12.251	8.2145	16.989	16.002	14.462	14.979	7.9673	8.4685
Observations	249756	249756	949402	949402	19068	19068	11774	19068	4025	4025

Table 3: Linear probability models for the bias.

	<i>Dependent variable</i>							
	betfair (1)	intrade (2)	iem.low (3)	iem.high (4)	iem.avg (5)	iem.last (6)	polls.d (7)	polls.r (8)
price	1.447*** (0.002)	1.242*** (0.001)	0.635*** (0.012)	0.607*** (0.011)	0.894*** (0.012)	0.825*** (0.010)	1.034*** (0.014)	1.036*** (0.014)
Constant	-0.1479*** (0.001)	-0.069*** (0.0004)	0.134*** (0.004)	0.133*** (0.004)	-0.011** (0.005)	0.018*** (0.004)	-0.023** (0.010)	-0.011 (0.007)
Observations	249,756	949,065	19,068	19,068	11,774	19,067	4,060	4,060
$R^2$	0.733	0.596	0.133	0.129	0.305	0.256	0.587	0.583
Adjusted $R^2$	0.733	0.596	0.133	0.129	0.305	0.256	0.587	0.583
Residual s.d. error	0.2584 (df = 249754)	0.292 (df = 949063)	0.399 (df = 19066)	0.400 (df = 19066)	0.358 (df = 11772)	0.370 (df = 19065)	0.307 (df = 4058)	0.307 (df = 4058)

Note: \*  $p < 0.1$ ; \*\*  $p < 0.05$ ; \*\*\*  $p < 0.01$ .

Table ?? reports the output of Mincer-Zarnowitz regressions for determining the forecast bias (see Eq. (??)), with each column corresponding to a different regression on a different market. The leftmost columns are Betfair and Intrade, regressed over every bet matched in those markets, followed by four columns for the summary daily information reported by IEM, and Democrat and Republican polls on the right.

The coefficients of the regression models relate to the bias, and specifically to the *favourite-longshot bias*: relatively, forecast outcomes occur too frequently for favourites, and too infrequently for outsiders. The favourite-longshot bias manifests itself in a *beta* coefficient that is over one, and thus, we can conclude that there is a strong favourite-longshot bias on both Betfair ( $\hat{\beta} = 1.449$ ) and Intrade ( $\hat{\beta} = 1.24$ ), a strong reverse favourite-longshot bias with all IEM prices ( $\hat{\beta}$  coefficients between 0.61 and 0.89), and a small favourite-longshot bias for polls ( $\hat{\beta} = 1.034, 1.036$ ).

These regression results are consistent with the graphical representations of bias in Figures ??–??, since Betfair and Intrade have the most pronounced departures from the 45° line. This is of interest in our context because it raises something of a paradox: if forecasts are biased upwards for favourites but downwards for outsiders, this could be argued to be a good forecast performance, as favourites are identified correctly and predicted strongly to succeed. In the extreme case, we might anticipate seeing forecasts above 50% occurring 100% of the time, and forecasts below 50% occurring 0% of the time. This performance would rank poorly in terms of the measures we consider here, but would imply that that particular forecast was very effective from a practical perspective. Based on Figures ??–??, we might conclude that Betfair and Intrade perform best in this latter regard, with the forecasts either side of 50% being around either 100% or 0%; both exhibit a strong favourite-longshot bias. Furthermore, we might argue that Betfair is more decisive in this regard than Intrade, as is reflected in the size of the departure of the  $\beta$  coefficients from one. Hence, we may conclude that Betfair and Intrade are more likely to yield correct forecasts even if they exhibit more bias than (corrected) polls, and we see, for favourites in particular, that Betfair is very decisive, with identified favourites being up to 25 percentage points more likely to win.

Considering the bias more generally, the second row of Table ?? reports mean errors. Opinion polls report the smallest biases, with tiny biases of 0.1 percentage points either way being the biases for Democratic (negative) and Republican (positive) polls. It should perhaps be noted that we removed the biases induced by known Democratic or Republican leaning pollsters, and other measurable biases, when constructing these probabilities (see Section ??).<sup>13</sup> Following the polls, Intrade reports the smallest bias, at 0.4 (positive) percentage points, IEM reports biases of between three and seven percentage points, and Betfair’s bias is five percentage points. Thus, in terms of the bias on individual forecasts relative to the binary outcome of that election, we can say that corrected polls display the least bias, followed by Intrade, then Betfair and IEM. We report Betfair and Intrade probabilities corrected using the linear regressions in Table ??, and naturally discover that bias-corrected forecasts have no bias.

Thus, opinion polls perform best in terms of the bias, although they are bias-corrected, and there is no obvious best performer of the prediction markets, although Intrade does appear to display slightly less bias.

---

<sup>13</sup>Note also that the (D) and (R) in the tables refer to polls for Democrat or Republican candidates, rather than probabilities from pollsters with known political leanings.

### 4.3 Precision

We measure the precision of forecasts by considering the variance of the forecasts from each source, and its reciprocal. The variance is listed in the third row of Table ??, and the precision in the fourth row. A lower variance is a desirable property, and therefore, by construction, a higher value of the precision is too.

The three prediction markets all have lower variances than the opinion polls, and hence, higher precisions. IEM shows the lowest variance, which cannot be attributed solely to the daily frequency of the data, as the Betfair and Intrade prices still display greater variances than IEM if we aggregate them to the daily frequency.

The bias-corrected Betfair and Intrade probabilities display considerably less precision and greater variance, which probably reflects the linear bias-correction for what Figures ?? and ?? reveal to be a heavily non-linear bias.

We can also consider the mean squared error, or Brier score, from Eq. (??), which is provided in the first row of Table ?? and combines bias and precision. We find that Betfair and Intrade perform slightly better than the bias-corrected polls, and considerably better than the (uncorrected) IEM polls, while the bias-corrected Betfair and Intrade predictions provide further improvements in terms of prediction.

Thus, the picture appears somewhat mixed. As was noted in Section ??, the Hayek hypothesis suggests that decentralised prediction markets ought to provide better forecasts of outcomes than the more centralised opinion polls. Nonetheless, ? proposed that the performances of corrected polls can be comparable. In line with this, we find that our corrected polls do exhibit less bias, but generally seem less precise. On the other hand, prediction markets appear to show small but correctable biases, and are relatively precise, though there are differences between the prediction markets, with IEM showing a greater precision but also greater bias than the two larger markets, Betfair and Intrade.

## 5 Conclusions

Election outcomes matter for economic outcomes, and as such, it is important to determine effective ways of forecasting electoral outcomes. This paper provides an empirical method of transforming opinion poll vote shares into outcome probabilities in order to allow a comparison between prediction markets and polls around the 2008–2012 US election cycle. We consider a richer range of prediction markets than most previous studies, evaluating Intrade, Betfair and Iowa Electronic Markets, as well as all polls available for elections during this period. In the process of converting vote shares to probability forecasts in order to allow direct comparisons to prediction markets, we correct the polls both for a range of known biases and for unknown biases, and we find that these corrected forecasts exhibit little bias but an unfavourable precision relative to prediction markets.

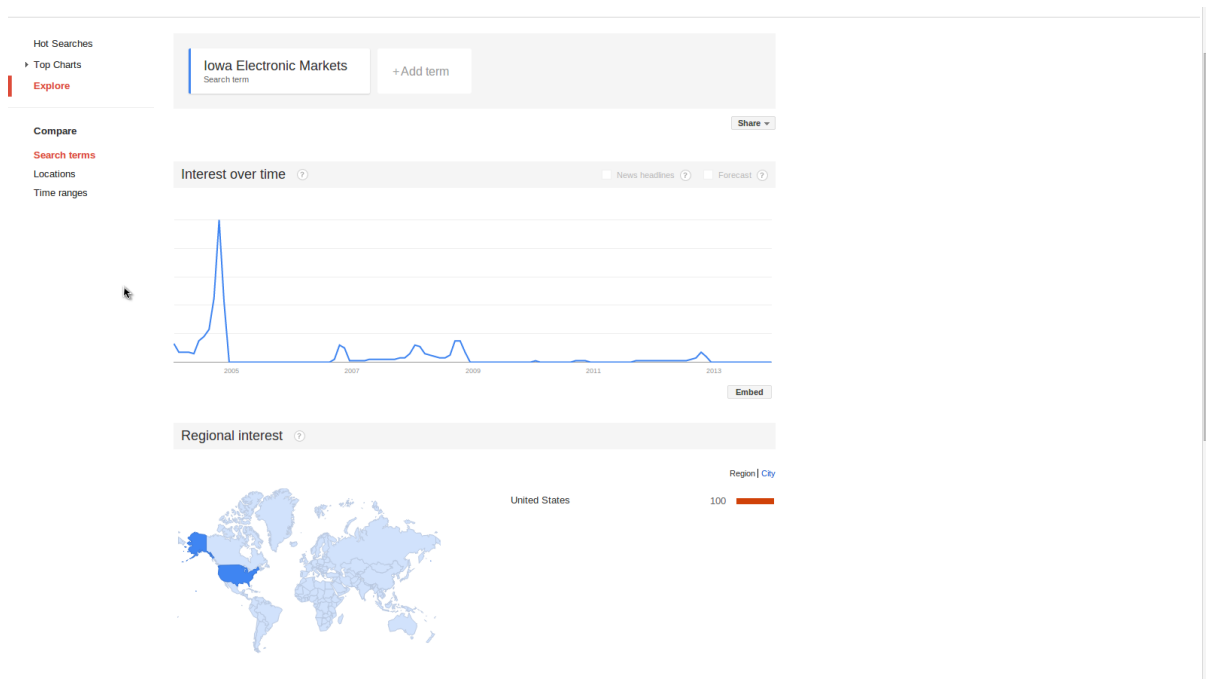


Figure 12: Google Trends information on searches for “Iowa Electronic Markets” over time and geography (available at [goo.gl/NwfGhf](http://goo.gl/NwfGhf)).