Polls to Probabilities: Comparing Prediction Markets and Opinion Polls

April 23, 2018

Abstract

Forecasting election outcomes is a hugely popular activity, and not without reason: 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, forecasts are often incompatible, as some are in terms of vote shares, and others are probabilistic outcome forecasts. In this paper we set out an empirical method for transforming opinion poll vote shares into probabilistic forecasts, and then evaluate the performance of prediction markets and opinion polls. We compare along two dimensions: bias and precision. We find that converted opinion polls perform well in terms of bias, and prediction markets on precision.

JEL Classification: C53, D83, D72.
Keywords: Forecasting Models, Information and Knowledge, Elections, Voting Behavior, Prediction Markets.

1 Introduction

Electoral outcomes can have significant economic impacts; as such, it is economically important, as well as of academic interest, to evaluate forecasting methods for elections. In this article we compare 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 whose value is contingent on some outcome occurring. The largest commercial prediction markets over this period produced predicted probabilities, whereas opinion polls produced projected vote shares. Thus we first develop an empirical method for converting vote shares into outcome probabilities. We then compare the two sources of forecasts by considering both bias and precision, which are both elegantly reflected 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 Iowa Electronic Markets (IEM, henceforth). There have already been academic investigations into the performance of opinion polls and prediction markets (e.g. Kou and Sobel, 2004; Leigh and Wolfers, 2006; Berg et al., 2008; Rothschild, 2009), and we contribute to this growing body of evidence.

While prediction markets have a long and rich history (Rhode and Strumpf, 2013), their internet-based electronic variants have been gathering increasing attention in recent years, and 2012 arguably

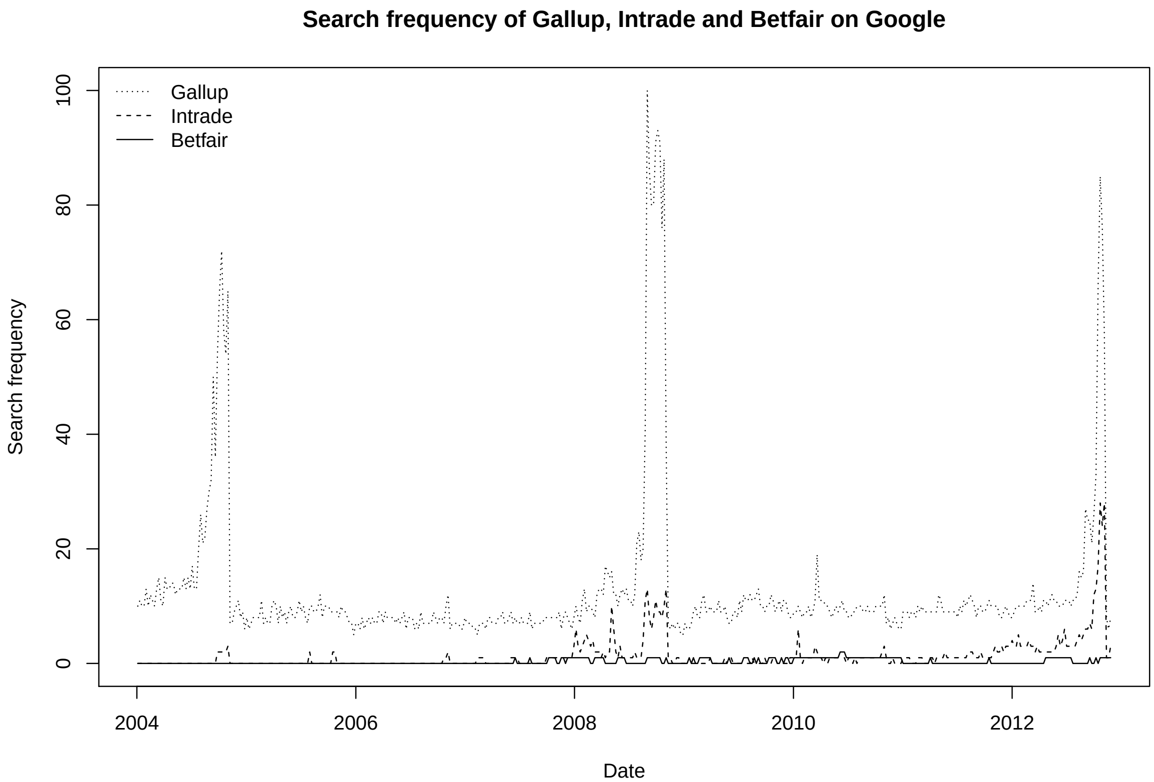

Figure 1: Relative frequency of searches for Gallup, Intrade and Betfair on Google. 100 refers to the week with most searches of Gallup, Intrade or Betfair since 2004. Source: Google Trends.

marked the year where one of them, Intrade, was regularly in the news 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 since early 2004, and hence three election cycles are distinctly visible, in 2004, 2008 and 2012, by the spikes in search volume for most pertinently Gallup, but also Intrade. Gallup registers more than sixty times more searches around the 2004 election than Intrade, but only ten times more in 2008, and three times more 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, as Republican presidential candidates tended to be priced more favourably on Intrade than Betfair. This divergence is noted in Rothschild and Sethi (2016), who analyse the behaviour of one particular trader who lost around $4m apparently manipulating the market price in favour of the Republican candidate, Mitt Romney. A clear implication of the Rothschild and Sethi (2016) paper is that there was a distinct bias on Intrade relative to Betfair (and also the Iowa Electronic Markets); investigating bias forms part of our description of each market.

We thus enable enhanced comparison between opinion polls and prediction markets in this paper by providing an empirical method for converting polls into probabilities. We also build on previous studies comparing opinion polls and prediction markets by considering a range of prediction markets, rather than a single one. In Section 2 we introduce our datasets, giving some detail about the nature of each source of forecasts, and the information we have for each source. In Section 3 we discuss our methodology for transforming vote-share polls into probabilistic forecasts, and for subsequently appraising each forecast source based on bias and precision. In Section 4 we provide the results, comparing our prediction markets to each other and to opinion polls, and in Section 5 we conclude.

2 Data

Data are fundamental to this investigation, and hence we firstly introduce our data and sources before discussing our methods. As our primary aim is to convert opinion polling data into a form more readily comparable to prediction markets, we first introduce our opinion polling data in Section 2.1. We then introduce prediction market data from Betfair (Section 2.2.1), Intrade (Section 2.2.2) and IEM (Section 2.2.3). It is important at the outset to be clear; 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 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).[1] This yields 271 elections, with the 2012 Presidential race yielding by some distance the greatest number of polls at 1609, of which 1098 are at the state level. The data contains 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), the pollster (plus an indication of whether that pollster is known to favour Republicans or Democrats).

In Figure 2 we plot the absolute error for polls by the number of days until the election. For both types of candidate, 89% of polls are within ten points of the actual outcome, 38% are within three points, and 12% are within a single point.

The units on the vertical axis of Figure 2 relative to, say, Figure 5, 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. We can, nonetheless, view the distribution of opinion poll vote shares as a probability distribution for election outcomes, and generate probabilities of outcomes; in Section 3 we outline our method for converting one to the other.

When considering opinion polls as forecasts, we note that polls only report (imperfectly) intended voting behaviour at a particular point in time, usually some time before the election actually occurs.[2] In the sense that opinion polls provide information, in advance, on a variable to be revealed at a future point in time (election day), they are nonetheless forecasts, and we treat them this way.

[1] For simplicity we ignore primary polling as this often relates to multiple-candidate elections which adds complications.

[2] Graefe (2013) 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.

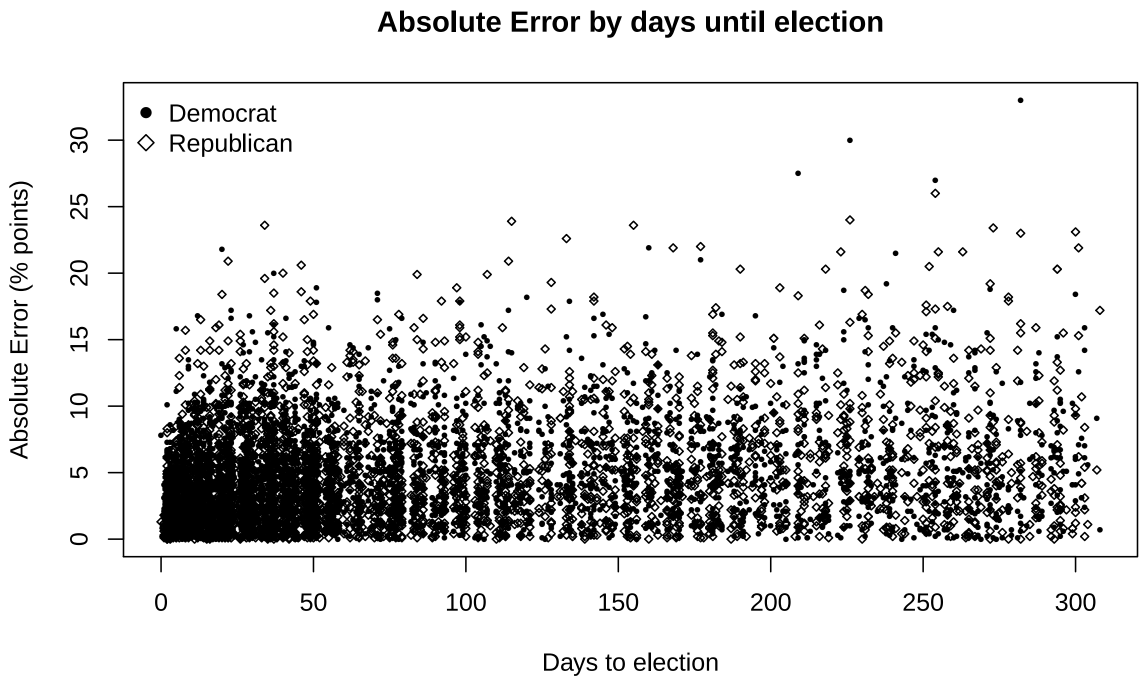

Figure 2: Absolute difference between each polled vote share and each outcome in 2010 and 2012, where Democrat and Republican errors are distinguished by different points. Source: Pollster and Real Clear Politics.

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), or lay them (sell contracts). It specialises in common sporting markets like horse racing and football, but also offers markets for political events. Betfair claimed to have over 4 million customers and a turnover of £50m a week.[3] Betfair matches those willing to back events with those willing to lay them at particular prices, and charges a commission of betweeen 2 and 5% on net winnings.[4] Bets on Betfair are denominated in decimal odds, which represent the multiple of the amount bet that must be exchanged if the event occurs. The reciprocal of decimals odds is commonly interpreted as the implied probability of the event being bet on occurring.

Our dataset includes every single trade on Betfair in the four-year election cycle between the Presidential elections in 2008 and 2012. It 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. In total over the cycle this

[3] See "Betfair moves operations to Gibraltar", *The Telegraph*, 9 March 2011 (goo.gl/JQS0Er).

[4] 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).

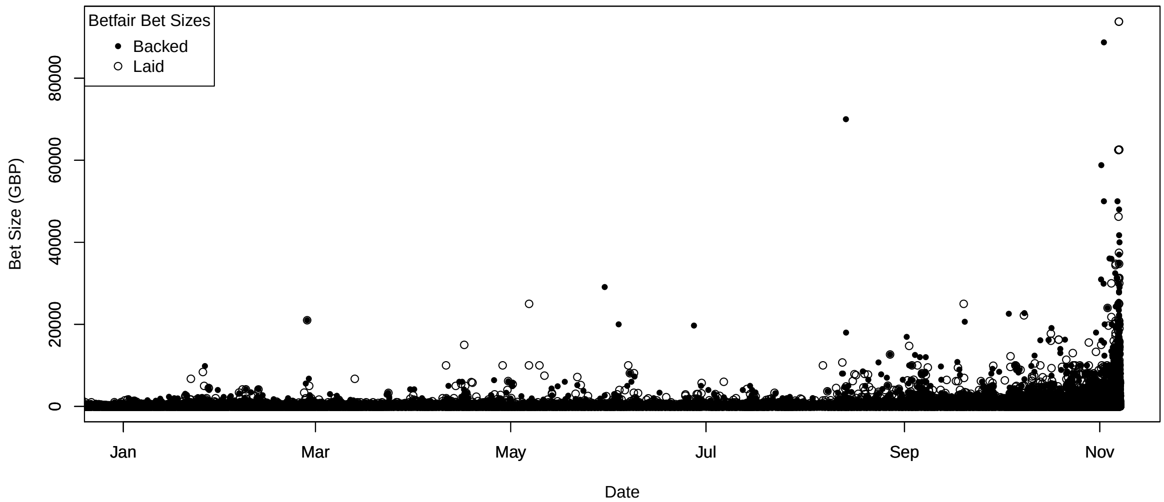

Figure 3: Size of bets matched on Betfair by type: Backing an event or laying it.

amounts to 249,756 observations, 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 includes data from Betfair, and hence we provide a unique contribution to the literature in this regard.

The largest single transaction is for £93,796.87, placed at 23:45:51 UTC on the 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.[5] 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 3; the hollow circles are bets laid and the solid circles are bets backed.

There are often multiple markets for similar outcomes on Betfair; for the election, there is a market for the next president, and also the winning party, and as such once the nomination is known for each of the two major parties, these two markets are essentially identical.[6] Figure 4 shows that on election day these two markets are essentially identical.

Of the 249,756 trades we have information on, 51.4% of bets are laying the event to occur, with the remainder backing the event. Accordingly, the average bet to back an event is slightly larger at £167.87 (median £15) than it is to lay the event (£158.46, median £9.99). Bets laid are marginally less focussed on the Presidential election, with 53.4% of them focussed 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%), and the mean odds that events are laid at is 28.69 (median 2.4).[7]

[5]As UTC is five hours ahead of Eastern Time, these trades do occur while voting was still open.

[6]Save for the negligible risk that a candidate may be forced to withdraw before election day.

[7]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, explaining the large gap between the mean and median decimal odds.

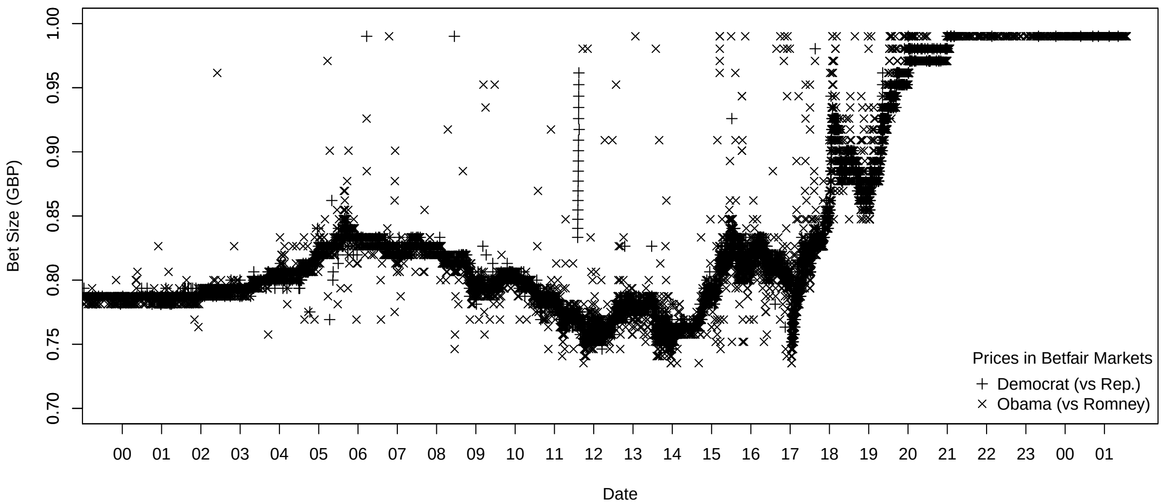

Figure 4: Different prices for bets matched on election day in two markets on Betfair: the market for the winning party and the next president.

2.2.2 Intrade

Intrade was a betting exchange, or prediction market. Based in Ireland, its focus was US political events, and its trading mechanism differed somewhat from 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 a cent and $10, with their price commonly interpreted as the implied probability of the event occurring.

For Intrade we have all contracts bought and sold, with buyer/seller ID (though no further characteristics), price and quantity, specific contract, and timestamp. Rothschild and Sethi (2016) focus 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 has 952,387 observations on 1,106 different contracts bought or sold.[8]

On average, each trade saw 26.5 contracts exchanged (although the median is just 5 contracts suggesting a rather skewed distribution), at an average price of $29.99, implying a 29.99% probability of the event being traded on occurring (the median price is 23). Of all the 25 million contracts traded in our dataset, 30.4% of them paid out (i.e. the event occurred), suggesting a reasonably calibrated market.

A controversy in 2012 was the allegation that 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 5, where of our three prediction markets Intrade has the lowest price for Obama consistently throughout the second half of 2012. It might thus be anticipated that Intrade would perform worse than other markets and possibly polls also. It might be noted nonetheless that the Obama/Romney market constituted only around 21% of all trades in our dataset, with the others scattered across various other elections including primaries. Furthermore considering particularly large trades, of the 35 trades of 9000

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

or more contracts, only four are for either of Obama (one) or Romney (three) to win the presidential election. We mention this not to rule out the possibility of politically motivated market manipulation, instead to point out that our dataset for Intrade is much broader than the Obama/Romney market. Rothschild (2009) uses Intrade data corrected for 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

Iowa Electronic Markets (IEMs) are one of the earliest electronic prediction markets, predating both Betfair and Intrade to the late 1980s, but remain somewhat smaller in nature, potentially at least in part due to the restrictions placed on trading: the maximum initial balance a trader can deposit is $500. Traders buy or sell contracts worth $1 if the event occurs, zero otherwise, and these contracts trade at anywhere between a cent and a dollar. Berg et al. (2008) use IEMs for their comparison of prediction market and opinion poll forecasts. They find that IEMs outperform opinion polls, particularly over longer forecast horizons.

We have collected all 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 2012, the Iowa Caucus of 2012, the 2010 Florida Senate election, alongside the 2012 Presidential election.[9] IEM data is at a daily frequency and in summary format: high price, low price, average price and last price, along with the number of contracts exchanged and the dollar volume exchanged. This raises an important methodological point pertaining to aggregation; our IEM data is aggregated to summary daily information whereas our Betfair and Intrade data remain fully disaggregated. As our intention is to avoid reducing the information content in our datasets, we do not aggregate our Betfair or Intrade information; we note that the empirical evidence regarding forecasting using aggregate or disaggregate variables is mixed (Hendry and Hubrich, 2011).

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

In Figure 5 the evolution of the market price on IEM is plotted for the Obama to win market during the second half for 2012; the IEM price fluctuates between the Betfair and Intrade prices throughout the period. It appears that in upswings in the Obama price, IEM moves with Betfair, but otherwise IEM is aligned with Intrade, suggesting that in times when information is entering the market that IEM and Betfair process that information in a similar manner. Rothschild and Pennock (2014) note research conducted into potential geographic biases in the processing of information that could sustain the kinds of price differentials they note between the UK-based Betfair and US-focused Intrade, and the IEM price may lend partial credence to this view as knowledge of IEM is reasonably geographically limited to North America (see Figure 12 on page 26), and for periods the IEM price tracks the Intrade price.

3 Methodology

In this section we set out our methodological approach to comparing opinion polls and prediction markets. We consider the bias and precision of both methods, yet we need to convert opinion polls from vote share

[9] 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.

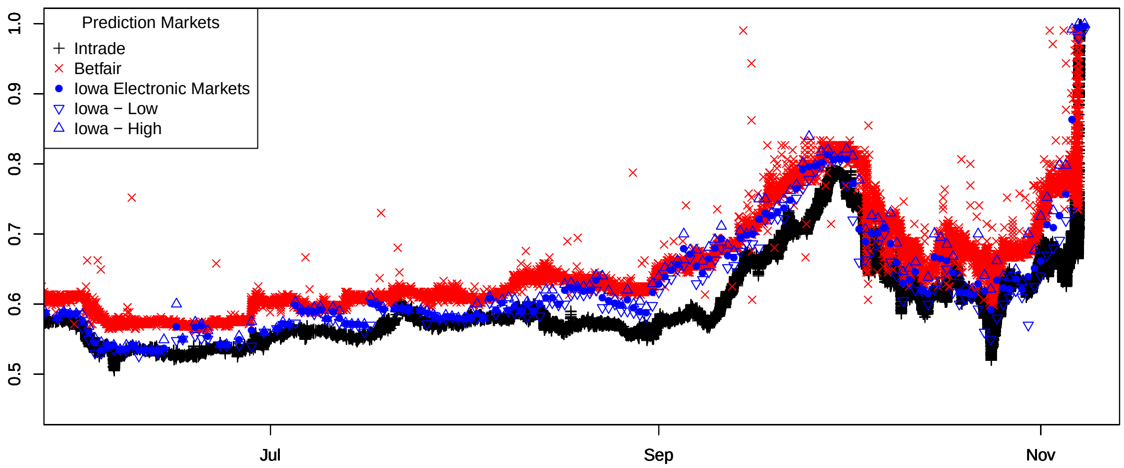

Figure 5: All three prediction markets for Obama to win, June–November 2012.

projections into probabilistic forecasts. In the process of doing this, we bias correct opinion polls, and hence it is appropriate to also bias correct prediction market forecasts in order to facilitate a more insightful comparison.

Opinion polls, ideally, are random samples of voting intentions; prediction markets are self-selected samples of agents trading contracts whose payoffs are contingent on voting outcomes. Neither are forecasts, strictly speaking, although both are commonly used as forecasts. The theory of enlightened voters (Gelman and King, 1993) might suggest that opinion polls during election campaigns would display more variance than prediction markets. This is because opinion polls reflect voters who are becoming enlightened, while prediction markets may be self-selected samples of the 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 already noted the presence of market manipulation in the introduction. It seems unlikely, however, that opinion polls do conduct truly random samples for many reasons; 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 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. We might thus assert that prediction markets ought to provide better forecasts than opinion polls by appealing to the so-called Hayek hypothesis (Hayek, 1945; Smith, 1982) which suggests that markets can work efficiently even when participants have a limited knowledge of the environment or other participants (see also Hurley and McDonough, 1995). Of the prediction markets in our sample, Intrade and Betfair do not restrict the financial exposure participants can accumulate, while IEM does; assuming that the prediction market mechanism allows information to be conveyed in both prices and quantities, it may thus be that IEM restricts this process.

Erikson and Wlezien (2008) argue against the superiority of prediction markets, suggesting that information on election outcomes from opinion polls is much broader than simply the polled vote share released, not least because of the enlightened voter idea. Erikson and Wlezien (2008) present evidence that polls, adjusted for such information by a method of bias correction, provide more accurate forecasts of election outcomes than prediction markets. The findings of Erikson and Wlezien (2008) are in contrast to Vaughan Williams and Reade (2016) who compare a richer set of prediction markets (the same set covered in this paper) to opinion polls, finding that prediction markets perform better. Erikson and Wlezien (2008) do not adjust prediction markets forecasts, however, suggesting that market prices do not need correction. This constitutes an assumption that prediction markets are efficient, although this is not formally tested. Rothschild (2009) 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

Practically, 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 vote share, V_{ij} for candidate j, where $\sum_j V_{ij} = 1$, or alternatively as a probability of victory of candidate j, hence $P_{ij} = P(\max_k (V_{ik}) = V_{ij})$. As we convert poll shares into implied probabilities of 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}:

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

Hence O_{ij} is 1 if candidate j wins election i, zero otherwise.

We need to transform opinion poll outcomes, which are in vote shares, into probabilities regarding the outcome of the election in order to compare them to predicton market outcomes, which are prices most commonly expressed in terms of probabilities. As we express 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 $\widehat{V}_{ij|h}$. Converting 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 *contained in a reported opinion poll.*

Methods exist to turn opinion poll vote shares into outcome probabilities; Page (2008) develops a theoretical parametric approach, while alternatives include 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 as would be expected with a commercial endeavour, details regarding the methods employed are undisclosed.[10] We follow an empirical approach, employing a regression method to create probabilities; regression methods as simple as ordinary least squares (OLS) allow the estimation of the probability associated with observing binary outcomes such as election outcomes (O_{ij}). OLS estimation with a binary dependent variable can produce estimated probabilities lying outside the unit interval, and hence we employ an alternative model that mitigates this problem, the probit model.

[10]See, for example, http://goo.gl/gv7xQ1, last accessed 16 October 2014.

A probit regression model using O_{ij} as dependent variable estimates, using historical data, 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:

$$\mathsf{P}\left(V_{ij} = \max_k V_{ik} \middle| X_{ij} \right) = \Phi\left(\beta X_{ij}\right), \tag{2}$$

where β is a vector of coefficients, and Φ is the cumulative density function of the standard normal distribution. In X_{ij} we include 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 also the number of days until an election. It seems reasonable to assume that such additional aspects of pollsters provides systematic information on the likelihood of victory of a candidate polled in addition to the polled vote share, and indeed this tallies with the method proposed by Erikson and Wlezien (2008) of correcting poll outcomes before comparing with prediction markets. Erikson and Wlezien (2008) 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).

Using the estimated coefficients from (2) we then construct fitted, or predicted values: We calculate $\widehat{\mathsf{P}\left(V_{ij} = \max_k V_{ik} \left| X_{ij}\right.\right)}$, 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 enables us to easily allow coefficients to vary between the two types of candidate.

In sum, our method replaces in (2) the election outcome as dependent variable in bias correction linear regression methods with an indicator variable taking the value 1 if that candidate wins, O_{ij}. In doing so, we can 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 that opinion polls produce.

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

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

This kind of regression, of outcomes on forecasts, is commonly referred to as a Mincer-Zarnowitz regression (Mincer and Zarnowitz, 1969), and is used to evaluate forecasts. In (3), γ would capture the direct relationship 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 β is a parameter that reflects their potential influence on outcomes not reflected in the opinion poll vote share $\widehat{V}_{ij|h}$. We set out in Section 4.1 the explanatory variables that make up X_{ij}, which includes things like the poll sample size and type, length of time before an election, and the potential partisan nature of a pollster.

The conventional statistical measure of bias, for some estimator $\widehat{\theta}$ for a parameter θ, is $\mathsf{E}\left(\widehat{\theta}\right) - \theta$; in expectation, an unbiased estimator is equal to the true value of the parameter, and so $\mathsf{E}\left(\widehat{\theta}\right) - \theta = 0$. 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: $\mathsf{E}\left(\widehat{P}_{ij}\right) - O_{ij}$.

If opinion polls are unbiased and reflect all available information at time h, then $\gamma = 1$ and $\beta = 0$ in (3), so that $\mathsf{E}(V_{ij}) = \widehat{V}_{ij|h}$.

We can then define bias corrected poll vote shares as $\tilde{V}_{ij}$, and they are calculated according to:

$$\tilde{V}_{ij} = \widehat{\gamma}\widehat{V}_{ij|h} + \widehat{\beta}X_{ij}, \tag{4}$$

where $\widehat{\gamma}$ and $\widehat{\beta}$ are ordinary least squares estimates of the parameters γ and β. Hence, as such, we can think about the opinion poll probabilities that we produce using (2) as bias corrected, as well as simply converted into appropriate units to compare to prediction market probabilistic forecasts.

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

A common method of determining the existence of bias in academic analysis of the betting industry is to run a regression of the form:

$$O_{ij} = \alpha + \beta P_{ijt} + e_{ijt}, \tag{5}$$

where P_{ijt} refers to a prediction at time t for candidate j in election i. This is another Mincer-Zarnowitz regression, similar to (3), the difference being that the forecast is a probabilistic forecast here, and the outcome a binary variable taking the value one if the event in question occurred. Hence estimated using least squares, (5), is a linear probability model. The null hypothesis is that forecasts are unbiased (and efficient), hence $\mathsf{E}(O_{ij}) = \mathsf{E}(P_{sij})$. In order for that to be the case we need that $\alpha = 1 - \beta = 0$. We can test this hypothesis using an F test having conducted the regression model described in (5).

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

$$\tilde{P}_{ijt} = \widehat{\alpha} + \widehat{\beta}P_{ijt}. \tag{6}$$

The alternative hypothesis here is that the bias is linear in form, which is clearly restrictive. Nonetheless, it is likely that if the actual bias is non-linear, the null of unbiasedness will still be rejected, particularly for large sample sizes like those we have. 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, their bias is conventionally measured using the forecast error, which is defined in terms of the forecast and the actual, or true, value:

$$\widehat{e}_{sij} = O_{ij} - \widehat{P}_{sij}. \tag{7}$$

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 each error. At the same time, such a large number of forecasts enables better evaluation of the expected value of forecasts from particular sources, relative to actual outturns. For prediction source $s \in$

$\{betfair, intrade, iem, polls\}$, the mean forecast error is:

$$ME = \widehat{e}_{sij} = O_{ij} - \widehat{P}_{sij}, \tag{8}$$

The mean of all forecasts made by each source s, the mean error, is thus a measure of 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 as forecasts become less precise (hence variance increases), the measure of precision also falls. In the limit, a measure with infinite (undefined) variance has zero precision, and point forecasts with zero variance have infinite precision, loosely speaking.

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

$$PR_s = \frac{1}{\sum_i \sum_j \left(\widehat{P}_{ij} - \mathsf{E}\left(\widehat{P}_{ij}\right)\right)^2}. \tag{9}$$

Note that this measure is independent of the true value 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 (9)).

3.3 Bias and Precision

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

$$MSE_s = \sum_i \sum_j \left(\widehat{e}^2_{sij}\right), \tag{10}$$

$$= ME^2_s + \sum_i \sum_j \left(\widehat{P}_{ij} - \mathsf{E}\left(\widehat{P}_{ij}\right)\right)^2. \tag{11}$$

The mean squared error is the sum of the squared bias, and the variance of the forecast, which forms the precision via (9).

The mean square forecast is often referred to, in binary outcome events like elections, as the Brier score Brier (1950). The Brier score is a probability scoring rule. Conversely to 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, to evaluate the bias of forecasts, and the variance of forecasts to consider the precision. We then consider the mean squared forecast error as a summary measure of both.

4 Results

4.1 Converting Polls to Probabilities

In Table 1 the results of the probit regression converting polled vote shared into implied probabilities of election victory for candidates, equation (2), are reported in the first two columns. Each column corresponds to a particular candidate, Republican or Democrat, respectively. The first row of the table is the constant term, and beneath that the subsequent three rows show the impact on the likelihood of victory of additional percentage points for a Democratic, Republican and Independence candidate 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, and 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. The impact of an independent candidate is not symmetric across the two types of candidate, however; an increased polling share for an independent candidate reduces the likelihood of a Republican candidate winning, but not a Democratic candidate. 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 votes shares are a number of other explanatory variables which may capture biases in opinion polls. The coefficients differ for the type of sample, which can be composed of likely voters, registered voters, or simply adults. 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 they potentially signify biases in the construction of opinion polls; by including terms such as these, in the process of turning opinion polls into probabilities, we bias correct polls. The closer to election day a poll is conducted, the larger (smaller) is the probability of a Democratic (Republican) victory. Finally the impact of pollsters aligned to particular parties is shown by the final two coefficients; the likelihood of a Republican victory is greater given we observe a poll from a partisan pollster, whereas a Democrat victory is less likely. Both effects are independent of the political leaning of the pollster and suggest that all partisan polls induce biases.

It is difficult to interpret these coefficients due to the dependent variable being likelihood of victory; in order to get a better sense of the biases that our probit models correct whilst generating probabilities of victory, we also report in the third and fourth columns of Table 1 a linear regression of the polled vote shares on all the explanatory variables used to generate probabilities; that is, estimates of (3).

Fewer coefficients are significant in these regressions relative to the probit regressions. The constant term is the average vote share with all other explanatory variables set to zero, and as such is not particularly informative, although taken in conjunction with the subsequent two rows, the polled vote shares for the candidates, it becomes a little more meaningful. The Republican coefficient in the Republican (Democrat) regression, at 0.447 (-0.286), is the marginal effect of an extra Republican polling point, whereas the Democrat coefficient in the Republican (Democrat) regression, 0.443 (0.656), is the marginal effect of an extra Democrat polling point. Taking the example of a 50/50 poll, this would suggest, taking the constant term (but setting all other variables to zero), 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, while there is no impact on the Democrat vote. In identified partisan Democrat pollster underweights the Republican candidate by 0.7 points, and overweights the Democrat candidate by 1.3 points, while

	Dependent variable:			
	Republican win	Democrat win	Republican Vote Share	Democrat Vote Share
	probit	*probit*	*OLS*	*OLS*
	(1)	(2)	(3)	(4)
Constant	3.550***	−4.548***	49.437***	32.582***
	(0.625)	(0.637)	(1.063)	(1.074)
Republican Opinion Poll Vote Share	0.091***	−0.082***	0.447***	−0.286***
	(0.006)	(0.007)	(0.012)	(0.012)
Democrat Opinion Poll Vote Share	−0.196***	0.213***	−0.443***	0.656***
	(0.007)	(0.007)	(0.011)	(0.011)
Independent Opinion Poll Vote Share	−0.082***	0.001	−0.309***	−0.467***
	(0.007)	(0.010)	(0.015)	(0.015)
Days until election	−0.001***	0.001***	0.0004	0.001*
	(0.0002)	(0.0002)	(0.0005)	(0.0005)
Sample Size	−0.0005***	0.0004***	−0.0001	0.0001
	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Likely Voters	0.850*	−0.994**	−1.865***	0.198
	(0.473)	(0.474)	(0.584)	(0.590)
Registered Voters	0.965**	−1.088**	−1.235**	0.526
	(0.472)	(0.475)	(0.577)	(0.583)
Identified Democratic Pollster	0.638***	−0.715***	0.767***	−1.294***
	(0.072)	(0.074)	(0.161)	(0.163)
Identified Republican Pollster	0.484***	−0.541***	0.073	−0.535
	(0.143)	(0.147)	(0.352)	(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 Inf. Crit.	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

Table 1: Probit regressions to create probabilities for election outcomes.

there is no significant impact of a Republican pollster (potentially reflecting the lower number of these pollsters in our dataset).

Figure 6 graphically represents both regression models. The top row shows the conversion of vote shares into probabilities for Republican (left) and Democratic (right) polls; with Republican polls there is much greater spread between 30 and 60 spread across the range of probabilities, whereas for Democrats the spread of poll shares is smaller. The variation around the curve represents the impact of other explanatory variables. Using these probabilities we can revisit the 2012 presidential campaign; in Figure 7 we plot the associated probabilities to the polls. The implied probabilities show the importance of considering polls in context; only a small number of polls imply the race was particularly close, with most yielding an 80-20 split in terms of probabilities.

The bottom row of plots in Figure 6 shows the correction for various factors that bias polled vote shares: 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 (3) as reported in Table 1. Both plots have a 45-degree line which shows the extent to which vote shares are impacted by our regression method. Both plots show a similar impact 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 would certainly seem in line with the findings of Erikson and Wlezien (2008) who suggest that using raw opinion poll numbers is unwise due to such known biases as those our regressions here have shown. The added value in our method, adding in further 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.

4.2 Bias

Figures 8–11 give a graphical representation of bias. In all plots, the horizontal axis represents the forecast outcome probability (hence is on the unit interval), while the vertical axis represents the frequency of outcomes; the proportion of all forecasts of a particular probability that turn out to occur. Unbiased forecasts would be scattered around the 45 degree line on such a plot, as this would represent probabilistic forecasts occurring as frequently as predicted.[11] From each plot, we see that the points plotted appear visually some distance from the 45 degree line, indicating some degree of bias. In Figures 8 and 9, Betfair and Intrade, the red bars reflect the number 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 a distinct pattern 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 to be more than about 60% likely to occur almost always occurs, and equally any event that either market identifies as being less than about 40% likely to occer very rarely occurs. This distinct relationship is not visible for either IEM (Figure 10) or opinion polls (Figure 11). These two plots display a more linear relationship around the 45-degree line, indicative of less bias.

[11] For example, 50% of forecasts that put the probability of an event occurring at 50% turn out to be correct.

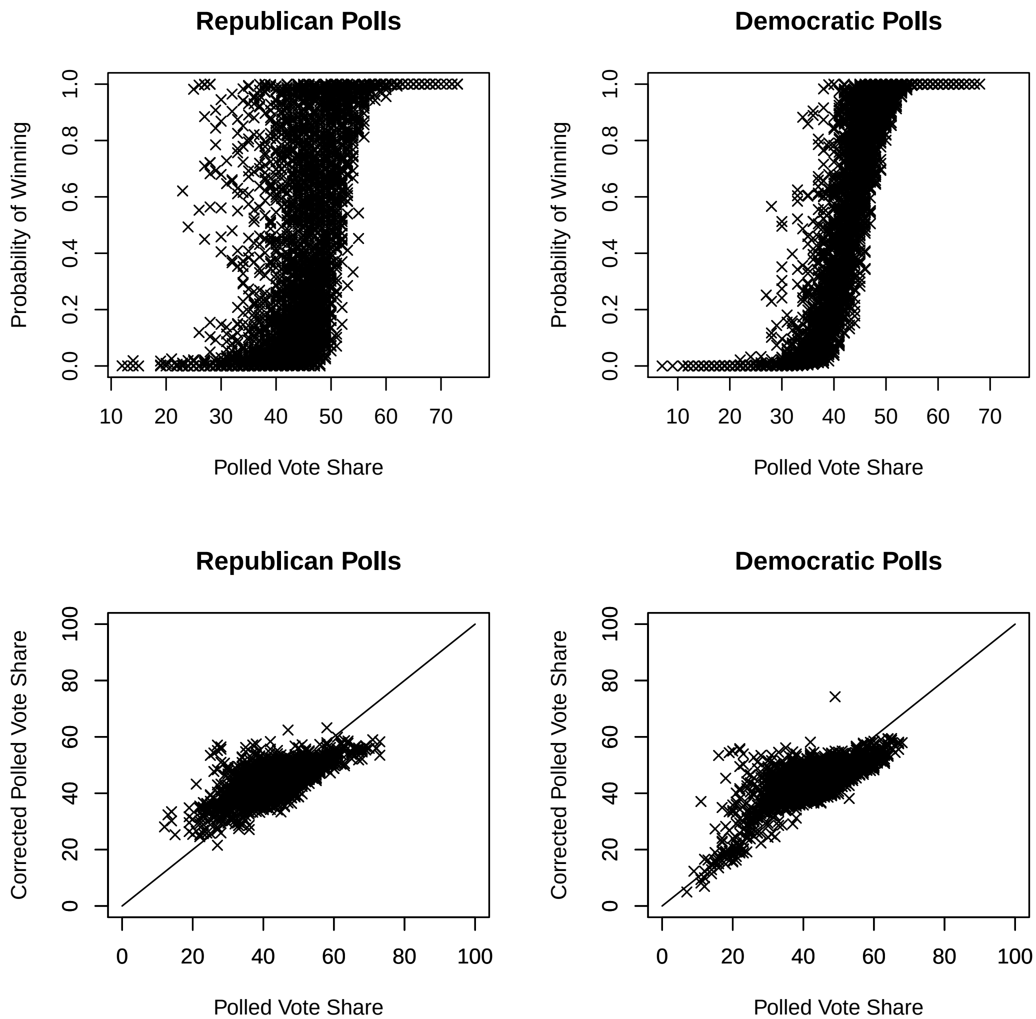

Figure 6: Plots related to opinion poll data. Top row graphically represents converting Republican (left) and Democrat (right) polls into probabilities of outcomes. Bottom row plots polled vote shares against corrected polled vote shares for Republican (left) and Democrat (right) polls. Source: Real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

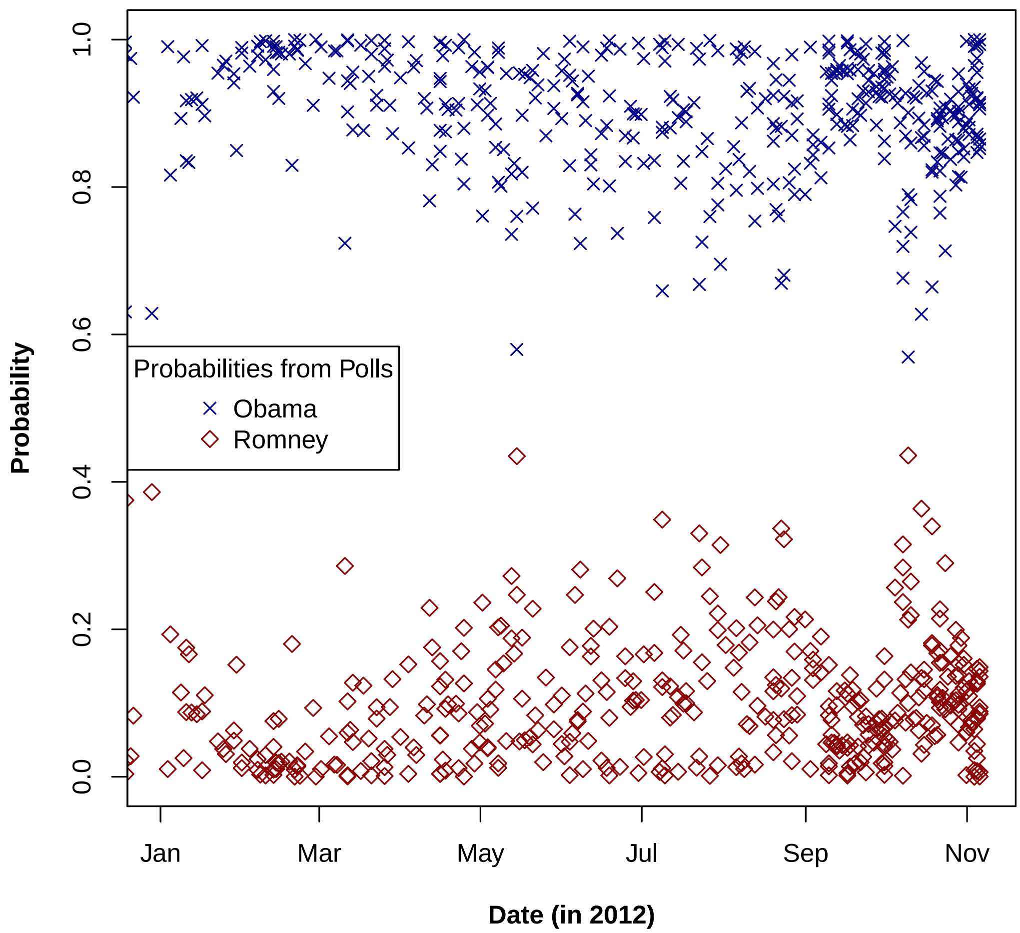

Figure 7: Probabilities implied from polls. Source: Pollster.

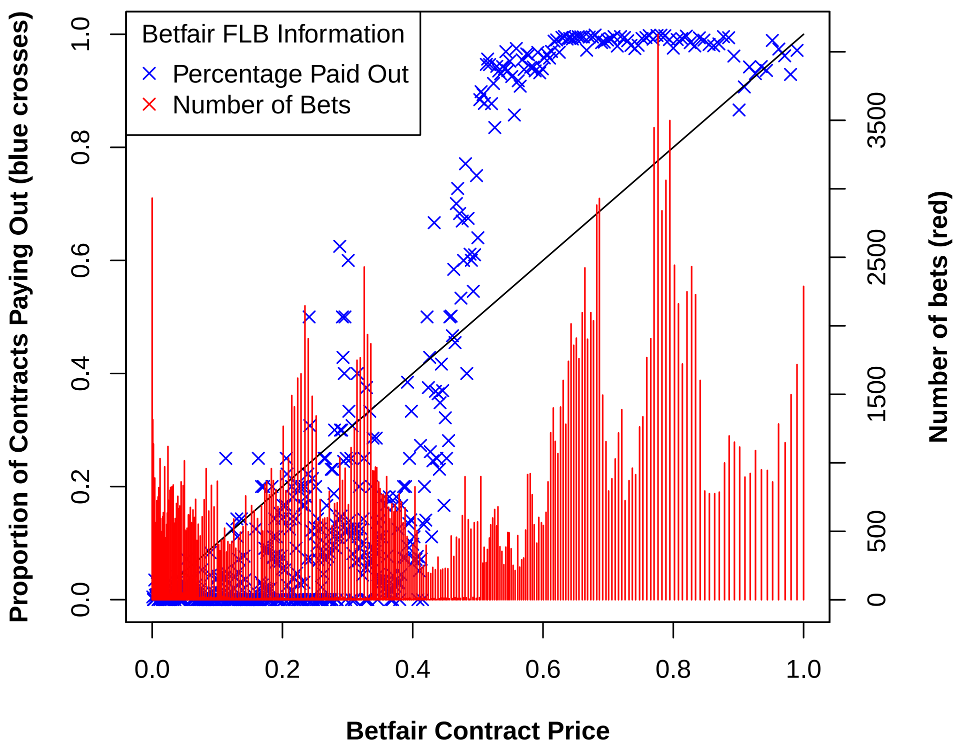

Figure 8: Betfair FLB plot.

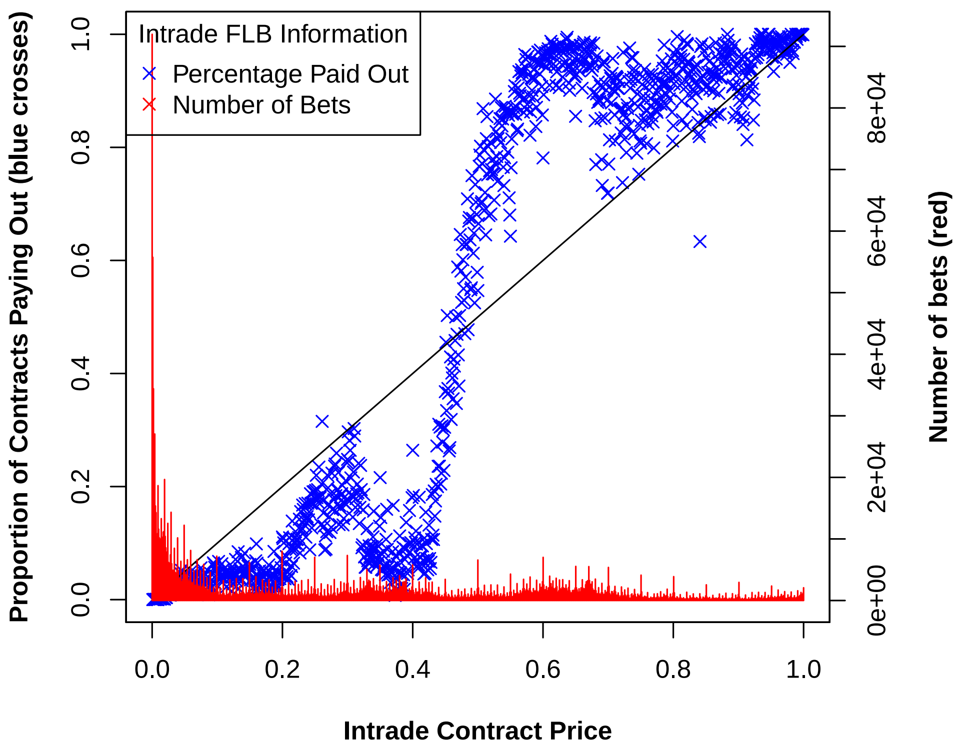

Figure 9: Intrade FLB plot.

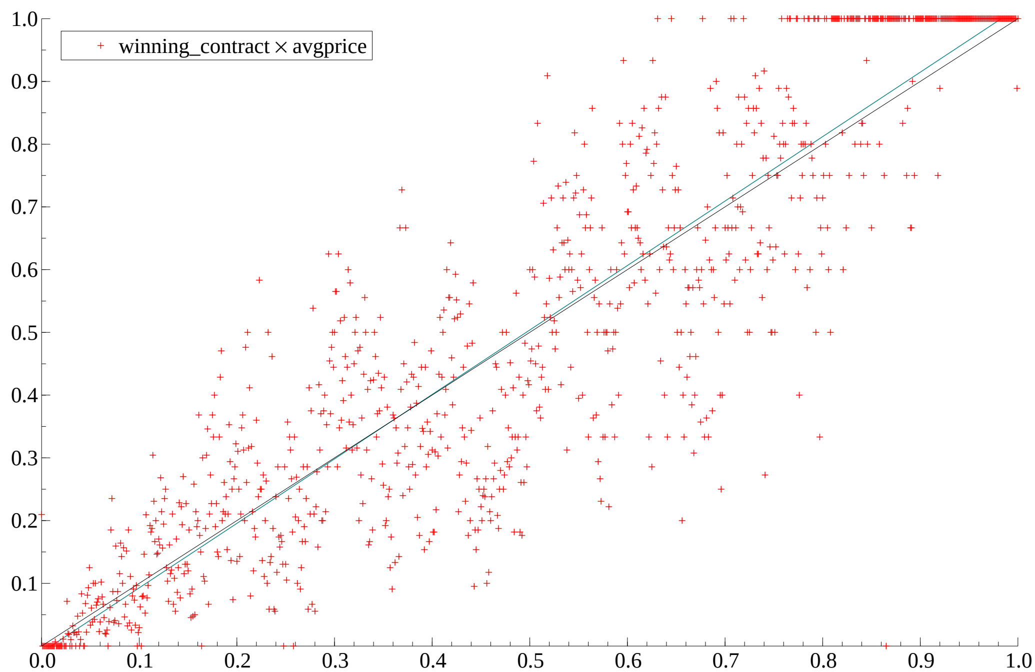

Figure 10: Iowa Electronic Markets FLB plot.

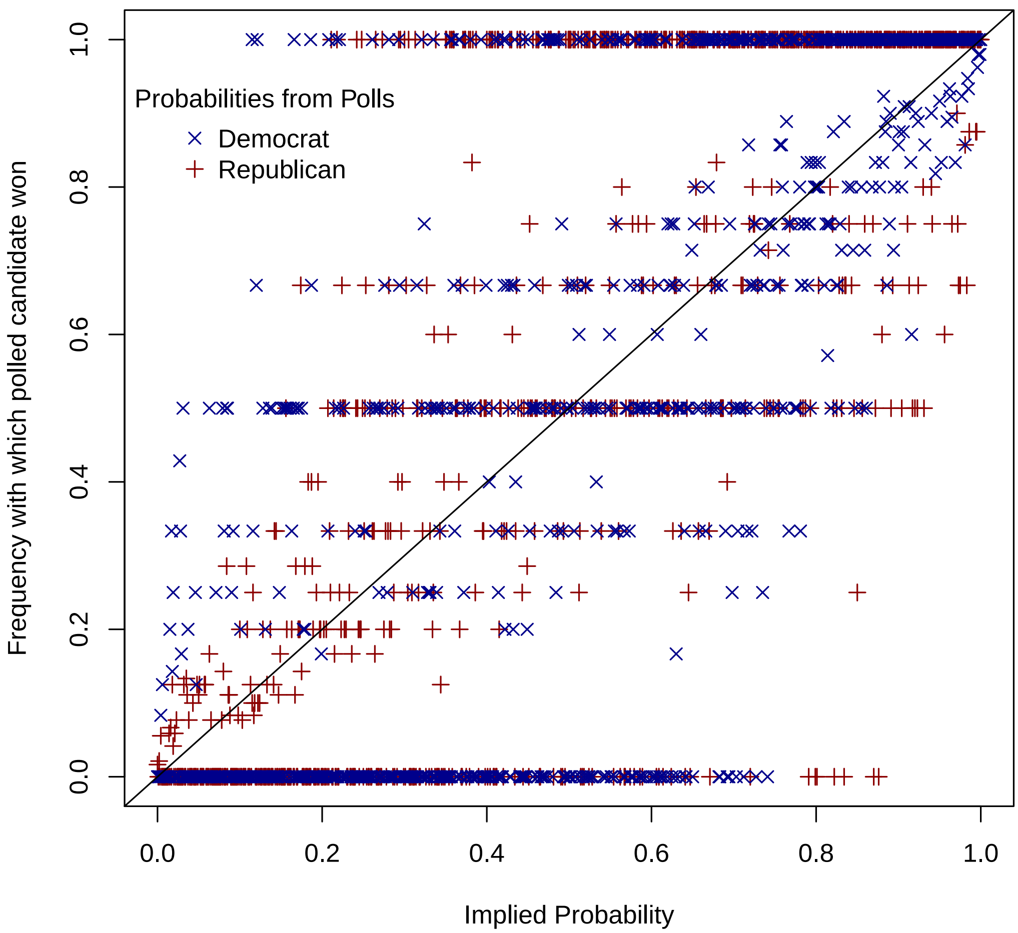

Figure 11: Polls FLB Plot.

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 2: Bias and Precision: Various measures for evaluating forecasts.

	Dependent variable:							
	betfair	intrade	iem.low	iem.high	iem.avg	iem.last	polls.d	polls.r
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
price	1.447***	1.242***	0.635***	0.607***	0.894***	0.825***	1.034***	1.036***
	(0.002)	(0.001)	(0.012)	(0.011)	(0.012)	(0.010)	(0.014)	(0.014)
Constant	−0.1479***	−0.069***	0.134***	0.133***	−0.011**	0.018***	−0.023**	−0.011
	(0.001)	(0.0004)	(0.004)	(0.004)	(0.005)	(0.004)	(0.010)	(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 Std. 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 3: Linear Probability Models for Bias

We report the output of Mincer-Zarnowitz regressions to determine forecast bias (see (5)) in Table 3; each column corresponds to a different regression on a different market. The leftmost columns are Betfair and Intrade, regressed over every bet matched in those markets, before there are four columns for the summary daily information reported by IEM, and finally on the right there are Democrat and Republican polls.

The coefficients of the regression models relate to bias and specifically the *favourite-longshot bias*: Forecast outcomes occur too frequently for favourites, and too infrequently for outsiders, relatively. The favourite-longshot bias manifests itself in a *beta* coefficient above 1, and thus we can conclude that there is a strong favourite-longshot bias on both Betfair ($\widehat{\beta} = 1.449$) and Intrade ($\widehat{\beta} = 1.24$), a strong reverse favourite-longshot bias with all IEM prices ($\widehat{\beta}$ coefficients between 0.61 and 0.89), and a small favourite-longshot bias for polls ($\widehat{\beta} = 1.034, 1.036$).

These regression results are consistent with the graphic representations of bias in Figures 8–11, since Betfair and Intrade have the most pronounced departures from the 45-degree line. This is interesting in our context as it raises something of a paradox; if forecasts are biased upwards for favourites but downwards for outsiders, then this could be argued to be a good forecast performance: favourites are correctly identified and strongly predicted to succeed. In the extreme, 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 our measures considered here, yet from a practical purpose would imply that that particular forecast was very effective. Based on Figures 8–11, we might conclude that Betfair and Intrade perform best in this latter regard, with forecasts either side of 50% being either around 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 reflected in the size of the depature of β coefficients from one. Hence we might conclude that Betfair and Intrade are more likely to yield correct forecasts even if they exhibit more bias than (corrected) polls, and specifically for favourites we see that Betfair is very decisive, with identified favourites more likely to win by up to 25 percentage points.

Considering bias more generally, we report mean errors in Table 2 on the second row. Opinion polls report the smallest bias, with a tiny 0.1 percentage points either way being the bias for Democratic (negative) and Republican (positive) polls. It is perhaps to be noted that in the construction of these probabilities (see Section 4.1) we removed biases induced by known Democratic or Republican leaning pollsters, and other measurable biases.[12] Following polls, Intrade reports the smallest bias at 0.4 (positive) percentage points, IEM report biases between 3 and 7 percentage points, while Betfair's bias is 5 percentage points. Thus in terms of 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 3, and naturally we discover that bias corrected forecasts have no bias.

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

[12]Note, also, that the (D) and (R) in tables refer to polls for Democrat or Republican candidates, rather than probabilities from pollsters with known politican leanings.

4.3 Precision

To measure the precision of forecasts, we consider the variance of forecasts from each source, and its reciprocal. The variance is listed on the third row of Table 2, and the precision, on the fourth row. A lower variance is a desirable property, and by construction, a higher value of precision also.

The three prediction markets all have lower variances than opinion polls, and hence higher precision. IEMs show the lowest variance, which cannot be attributed alone to the daily frequency of the data; if we aggregate Betfair and Intrade prices to the daily frequency, they still display greater variance than IEMs.

The bias-corrected Betfair and Intrade probabilities display considerably less precision, and greater variance, likely reflecting the linear bias-correction for what is a heavily non-linear bias, as revealed in Figures 8 and 9.

We can consider also the mean squared error, or Brier score, which is provided in the first row of Table 2, which combines both bias and precision, from (11). We find that Betfair and Intrade perform slightly better than the bias-corrected polls, and considerably better than the (uncorrected) IEM polls, and the bias-corrected Betfair and Intrade predictions provide further improvements in terms of predicting.

Hence the picture appears somewhat mixed. As noted in Section 3, the Hayek hypothesis suggests that decentralised prediction markets ought to provide better forecasts of outcomes than the more centralised opinion polls. Nonetheless, Erikson and Wlezien (2008) proposed that corrected polls can perform comparably well. In line with this, we find that our corrected polls do exhibit less bias, yet they seem generally less precise. In terms of prediction markets, they appear to show small but correctable biases, and are relatively precise. There are differences between the prediction markets nonetheless, with IEM showing greater precision but more bias than the two larger markets, Betfair and Intrade.

5 Conclusions

Election outcomes matter for economic outcomes; as such, determining effective ways to forecast electoral outcomes is important. In this paper we provide an empirical method to transform 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. We correct the polls for a range of known biases, as well as unknown biases in the process of converting vote shares to probability forecasts in order to compare directly to prediction markets, and we find that these corrected forecasts exhibit little bias but unfavourable precision relative to prediction markets.

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.

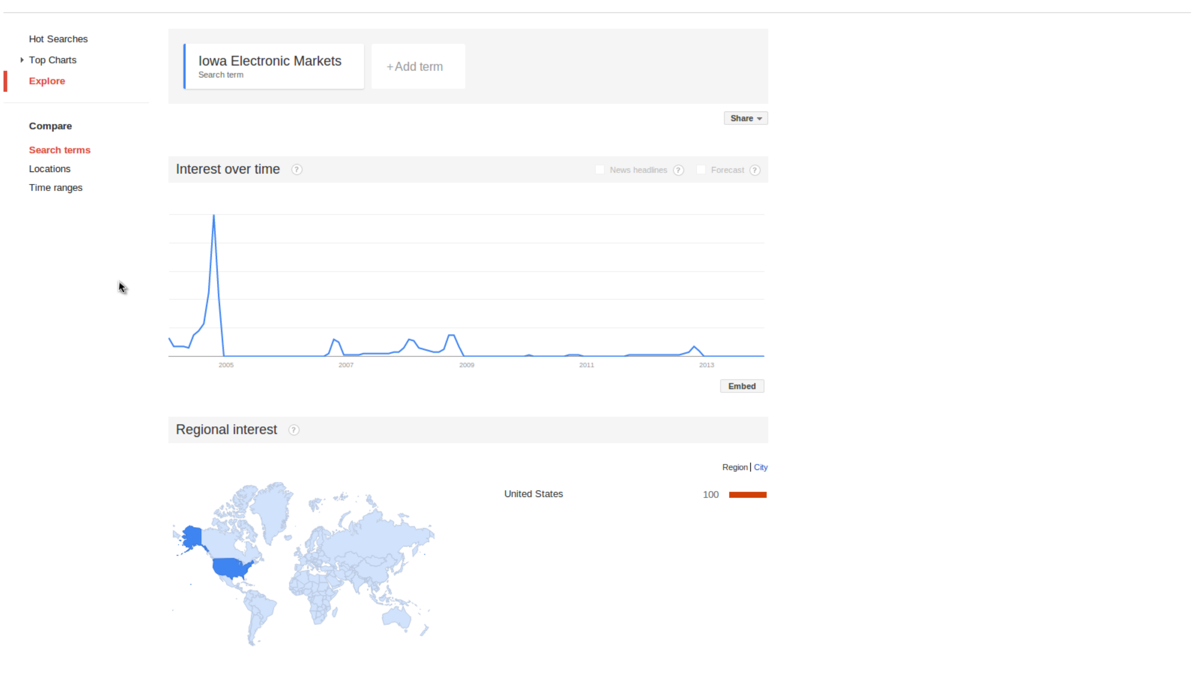

Figure 12: Google Trends information on searches for "Iowa Electronic Markets" over time and geographically (available at goo.gl/NwfGhf).

R.S. Erikson and C. Wlezien. Are Political Markets Really Superior to Polls As Election Predictors? *Public Opinion Quarterly*, 72(2):190–215, Summer 2008.

A. Gelman and G. King. Why are American Presidential Election Campaign Polls so Variable when Votes are so Predictable? *British Journal of Political Science*, 23(4):409–451, October 1993.

A. Graefe. Accuracy of Vote Expectation Surveys in Forecasting Elections. Working paper, Department of Communication Science and Media Research, LMU Munich, 2013.

F.A. Hayek. The Use of Knowledge in Society. *The American Economic Review*, 35(4):510–530, September 1945.

D.F. Hendry and K. Hubrich. Combining disaggregate forecasts or combining disaggregate information to forecast an aggregate. *Journal of Business & Economic Statistics*, 29(2), 2011.

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.

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.

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.

J.A. Mincer and V. Zarnowitz. The evaluation of economic forecasts. In *Economic Forecasts and Expectations: Analysis of Forecasting Behavior and Performance*, pages 1–46. NBER, 1969.

L. Page. Comparing prediction market prices and opinion polls in political elections. *The Journal of Prediction Markets*, 2(1):91–97, May 2008.

P.W. Rhode and K.S. Strumpf. The Long History of Political Betting Markets: An International Perspective. In L. Vaughan Williams and D.S. Siegel, editors, *The Oxford Handbook of the Economics of Gambling*. Oxford University Press, New York, 2013.

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

D. Rothschild and D.M. Pennock. The extent of price misalignment in prediction markets. *Algorithmic Finance*, 3(1-2):3–20, 2014.

D.M. Rothschild and R. Sethi. Trading Strategies and Market Microstructure: Evidence from a Prediction Market. *The Journal of Prediction Markets*, 10(1), 2016.

V.L. Smith. Markets as Economizers of Information: Experimental Examination of the "Hayek Hypothesis". *Economic Inquiry*, 20(2):165–179, April 1982.

L. Vaughan Williams and J.J. Reade. Forecasting Elections. *Journal of Forecasting*, 35(4):308–328, 2016.