International Journal of Forecasting 31 (2015) 952–964

Contents lists available at ScienceDirect

International Journal of Forecasting

journal homepage: www.elsevier.com/locate/ijforecast

Combining forecasts for elections: Accurate, relevant, and timely

David Rothschild

Microsoft Research, New York, NY, USA

ARTICLE INFO

Keywords:
Election forecasting
Surveys
Econometric models
Prediction markets
Combining forecasts
Probability forecasting

ABSTRACT

This paper increases the efficiency and understanding of forecasts for Electoral College and senatorial elections by generating forecasts based on voter intention polling, fundamental data, and prediction markets, then combining these forecasts. The paper addresses the most relevant outcome variable, the probability of victory in state-by-state elections, while also solving for the traditional outcomes, and ensuring that the forecasts are easy to update continuously over the course of the main election cycle. In an attempt to maximize both these attributes and the accuracy, I create efficient forecasts for each of these three types of raw data, with innovations in aggregating the data, then correlate the aggregated data with the outcomes. This paper demonstrates that all three data types make significant and meaningful contributions to election forecasting. Various groups of stakeholders, including researchers, election investors, and election workers, can benefit from the efficient combined forecasts defined in this paper. Finally, the forecast is tested on the 2012 elections and excels out-of-sample.

1. Introduction

Polling data has been the most prominent component of election forecasts for decades. From 1936 to about 2000, it was standard in both the academic and popular press to utilize just the raw data, the results of individual voter intention polls, as an implicit forecast of an election. By 2004, poll aggregation became common on the internet. Although aggregated polls provide both stability and accuracy relative to individual poll results, aggregated polls are meant to be a closer approximation, relative to individual poll results, of what an election would look like if it was suddenly held on that day, not an expectation of what will happen on Election Day. By 2008, some websites, run by a mix of academics and non-academics, finally began publishing poll-based forecasts (i.e., forecasts derived from aggregating raw polls then translating the results into a forecast of the election outcome). Furthermore, they shifted the outcome variable to the probability of victory in the Electoral College or senatorial elections, rather than the standard expected vote shares of the national popular vote.

The need to transform raw polling data into a forecast is conclusive in the literature. Campbell (2008) clearly illustrates the anti-incumbency bias, whereby incumbents have lower polling values than the actual election results, and the fading of early leads in polls, whereby election results are tighter than polling numbers. Erikson and Wlezien (2008a) show that translating raw polling data into a forecast makes it more accurate for both the estimated vote share and the probability of victory. Rothschild (2009) improves on the work of Erikson and Wlezien (2008a) by aggregating the daily polls over time, eliminating noisy daily fluctuations, then translating them into a forecast. At the same time, Rothschild (2009) designed his poll-based forecast to be the most accurate forecast using the same general model as Erikson and Wlezien (2008a), leaving open the possibility of creating even more accurate

E-mail address: DavidMR@Microsoft.com.

http://dx.doi.org/10.1016/j.ijforecast.2014.08.006
0169-2070/

transformations through more advanced models of the aggregation and subsequent translation of the polling data into forecasts. The most recent advances in creating forecasts from polls have been in the area of aggregation, including eliminating poll company specific effects and combining the snapshot for any given state with other state and national polls. In this paper, when available, we use both the most transparent and the most efficient method possible, without these further steps, but with Stanford's Simon Jackman's interpretation of these steps, as made available through Pollster.com (Jackman, 2005).

There is a massive body of literature on the modeling of fundamental data, which has found that most models are not useful as forecasts, but rather explain the correlations between different variables and election outcomes. These models use a variety of economic and political indicators, such as past election results, incumbency, presidential approval ratings, economic indicators, ideological indicators, biographical information, policy indices, military situations, and facial features of the candidates. Hummel and Rothschild (2013) provide a substantial list of such models; however, there are several reasons why they are generally not useful for producing forecasts. First, many models are difficult to duplicate, such as that of Armstrong, Green, Jones, and Wright (2010), which utilizes pictures of the candidates. Second, many models incorporate pre-election polls or other late-arriving data; for example, Lock and Gelman (2010) use a model that cannot be resolved until October of the election year. These types of models are designed more to help us obtain an understanding of the correlation between fundamental data and election outcomes, than for forecasting the election during the cycle. Third, most fundamental data models forecast just the presidential national popular vote; examples include those of Abramowitz (2004, 2008). This is a serious issue, not just because it is not the ideal outcome variable, but because it means that there is an extremely limited identification in just one outcome every four years. Fourth, Klarner (2008) pushed the literature forward into the realm of earlier state-by-state forecasts, but still incorporated early polling in the model. In order to compare the value of the different data sources, it is crucial to consider models that use only one data source. Without any polling data, improving on the variable choice and range of data, the model presented by Hummel and Rothschild (2013) has much smaller errors than that of Klarner (2008), and could be put to use by June 15 of the election year. Thus, I utilize the model of Hummel and Rothschild (2013) exclusively as the fundamental model for this paper, because it is the most accurate state-by-state fundamental model for Electoral College and senatorial elections, can be executed early in the cycle, and excludes voter intention polling data. The out-of-sample errors for the model of Hummel and Rothschild (2013) are smaller than the within-sample errors for the most widely circulated state-by-state fundamental models, including Klarner's most recently updated model (Klarner, 2012).[1]

The modern history of the use of prediction markets is not as long as those of the other two data sources. The Iowa Electronic Market launched the modern era of prediction markets in 1988, introducing a winner-takes-all market in 1992. This type of market trades binary options which pay, for example, \$10 if the chosen candidate wins and \$0 otherwise. Thus, an investor who pays \$6 for a 'Democrat to Win' stock, and holds the stock through to Election Day, earns \$4 if the Democrat wins and loses \$6 if the Democrat loses. In that scenario, if there are no transaction or opportunity costs, the investor should be willing to pay up to the price that equals her estimated probability of the Democrat winning the election. The market price is the value at which, if a marginal investor were willing to buy above it, investors would sell the contract and drive the price back down to that market price (and vice-versa if an investor were willing to sell below it); thus, the price is an aggregation of the subjective probability beliefs of all investors.

Both in the last few cycles (Berg, Forsythe, Nelson, & Rietz, 2008; Rothschild, 2009) and in historical elections (Rhode & Strumpf, 2004), scholars have found that prediction market prices can create more accurate forecasts than poll-based forecasts; however, like polling and fundamental data, prediction market prices benefit from a transformation from raw data into a forecast, especially due to the favorite-longshot bias. Berg et al. (2008) conclude that raw prediction market prices are more accurate forecasts of the vote share than raw polling data. However, Erikson and Wlezien (2008a) challenge this finding by comparing raw prediction market prices with properly translated poll-based forecasts; this is confirmed by Rothschild (2009). At the same time, Wolfers and Zitzewitz (2006) highlight the transaction and opportunity costs of investing in prediction markets, Manski (2006) describes how investors in prediction markets behave as if they were risk-loving, and Snowberg and Wolfers (2010) conclude that there are systematic mis-perceptions of probability stemming from prospect theory; when we combine the results of these three papers, we see the favorite-longshot bias for prediction market prices. One hundred days before the election, if an investor believes that the Republican candidate has a 95% chance of winning, there are three reasons for her to bid less than \$0.95 for a contract that pays out \$1.00 if the candidate wins. First, with limited liquidity in the market (i.e., not enough traders and money in the market for all traders to always make their most efficient purchases), she may have to hold the contract until Election Day, thus incurring an opportunity cost. Second, she will incur some transaction costs when she buys and sells the contract, or when it expires. If the opportunity cost is \$0.02 and the transaction cost is \$0.03, then she would not bid more than \$0.90 in order to break even in expectation. Third, investors who behave as if they were risk loving gain

[1] Klarner (2012) drops the use of voter intention polling data, which were used in early versions of the model; however, his paper was not released until after the initial running of the model for and circulation of this paper, as I wanted to ensure that 2012 would be completely out-of-sample; that change brought Klarner's model closer to that of Hummel and Rothschild (2013), which was already available in a widely circulated working paper at that time. Still, while Hummel and Rothschild (2013) had similar errors for the estimated vote share to those of Klarner (2012) for the Electoral College in 2012, it had significantly smaller errors, nearly a full point on average, for the senatorial elections.

a greater expected utility from buying a longshot than a favorite, all else being equal. Thus, even though the unbiased, risk-neutral market price would be \$0.95, investors do not bid all the way to the biased, risk neutral amount of \$0.90 for a favored candidate, but may move the market to only \$0.85 or less. Rothschild (2009) corrected for this bias with a transformation suggested by Leigh, Wolfers, and Zitzewitz (2007), and proved that corrected prediction market data generate more accurate forecasts than aggregated and corrected polling data.[2] Still, Rothschild (2009) made no attempt to examine or improve upon the transformation suggested by Leigh et al. (2007), which was calibrated on a different type of prediction market data; this paper takes the logical next steps.[3]

There is a rich history of combining data for forecasts, in many domains, including elections, in order to improve the accuracy; however, nothing in the literature has considered state-by-state forecasting with all three types of data. Clemen (1989) provides an early overview across domains. Graefe, Armstrong, Jones, and Cuzán (2014) conclude that there are benefits for accuracy from combining many data types when producing forecasts of the vote share of the national popular vote for president. However, they suggest the simple procedure of using even weights at all points in the cycle, use simplified forecasts from the various data types, and only examine the national popular vote. Erikson and Wlezien (2008b) combine economic indicators and intention polls for forecasting the national popular vote, but do not investigate prediction markets, state-by-state elections, or the probability of victory. Erikson and Wlezien (2012) attempt to demonstrate empirically that raw prediction market prices provide no additional information over polling information, but again look only at a few national popular votes, which is the wrong outcome variable for prediction markets. There is no expectation that the vote share will be identified well by prediction markets, which generally trade on the probability of the election winner (i.e., if a candidate is poised to win big, the candidate should trade near \$1.00 per \$1.00 of payout for the candidate's victory, regardless of whether it is a 5- or 10-point expected victory).

A few recent Bayesian papers, including those of Linzer (2013), Lock and Gelman (2010), and Montgomery, Hollenbach, and Ward (2012), combine fundamental data and voter intention polling, but there are some key differences in the way in which I approach the data. First, Linzer (2013), for example, starts with a fundamental-based forecast for each state and updates the forecasts as the new polls arrive. I translate the separate data types into forecasts independently, rather than jointly, so that I can compare their informational value. This is not a criticism of the work of Linzer (2013) and similar papers, but is essential to the goal of this paper in comparing the different data types separately. Second, I allow the relative values of different data types in my final forecast to fluctuate by day, while Linzer (2013), for example, focuses on the certainty of the data by race. This paper provides a simpler model and highlights some key points about shifting information over time, but again, this is to allow me to optimize over outcomes different to those of Linzer (2013) and similar, and is not a criticism of their work. Third, these papers exclude prediction market data. Prediction market theory suggests that prediction market prices should include the information from voter intention polls and fundamental models already, but the literature lacks any empirical work on the subject. This paper uniquely provides a clean comparison of the three data types in the key setting of state-by-state elections.

This paper defines an efficient election forecast with three attributes geared towards maximizing returns for both election workers and researchers: state-by-state probabilities of victory, regular updating, and accuracy. Historically, election forecasts estimated the vote share rather than the probability of victory for two key reasons: first, the academic literature focuses on incremental improvements over historical forecasts, and the estimated vote share is the historical standard, and second, observers frequently interpret raw polls as naïve estimations of the vote share, making it the simplest rubric. The expected vote share is extremely important for election workers, especially when it is broken down by targetable demographics. However, most stakeholders care about shifts in the expected vote share only insofar as they affect the probability of victory; when judging the impact of a debate on the outcome of the election or deciding which race to invest in, the probability of victory is a more efficient metric. Historically, election forecasts have estimated national rather than state-by-state outcomes, but the US elects its president through the Electoral College, not by the popular vote.[4] Furthermore, any forecast that utilizes national outcomes exclusively has a serious issue with identification, as national elections occur so infrequently; state-by-state elections may be correlated, but they provide extra identification. There is a clear preference for considering the probability of victory at the state-by-state level, as the main popular forecasting sites, from FiveThirtyEight to Princeton Election Consortium, all report the probability of victory in the Electoral College as their main forecast (and barely mention the expected vote share, especially the national popular vote).[5] There are increasing numbers of cases of the probability of victory being utilized in academia, such as Lewis-Beck and Rice (1992), and also state-by-state outcomes, such as Linzer (2013), yet the influential *PS: Political Science*'s special issue in 2012, "Forecasting the 2012 American National Election", had the probability of victory for the national popular vote along with the expected national popular vote (Campbell, 2012), and the Iowa Electronic Market still uses victory in the national popular vote for its main market, not the Electoral College.

[2] Interestingly, Erikson and Wlezien (2008a) recognized the existence of the favorite-longshot bias in prediction markets but did not correct for it.

[3] Leigh et al. (2007) calibrated the transformation on national presidential elections, while Rothschild (2009) utilized state-by-state Electoral College and senatorial elections from recent cycles.

[4] On four separate occasions, the winner of the national popular vote has lost the Electoral College!

[5] FiveThirtyEight accounted for 20% of the New York Times traffic in the lead-up to the 2012 election; see http://money.cnn.com/2012/11/07/news/companies/nate-silver-election/index.html.

Historically, forecasts have typically been updated whenever new information is released, such as a new poll, or right before the event, when all resources have been allocated and major events have already passed; however, forecasts are more valuable further in advance of the event and if they are updated regularly, so that they can exist and be fresh at times when stakeholders can still reallocate investments to more efficient uses.[6] For researchers, the earlier the forecast, the more events they capture and the more time-granular the forecasts, the better, for the sake of studying the effects of small events that occur in the lead up to an election; such forecasts allow researchers to isolate the effects of debates on outcomes, or to study the effects of elections on other events, as Snowberg, Wolfers, and Zitzewitz (2006) did for elections and the economy. The point of accuracy is to achieve the smallest errors, the most efficient calibration, and out-of-sample robustness (i.e., models that explain the future, rather than describing the past). These extra checks, beyond the simple error, help to offset the difficulty in judging forecasts, where there is an uncertain ground truth on all days except the last (i.e., we do not know how the election would have turned out on any day other than the final day).

This paper demonstrates that all three data types, namely polls, fundamentals, and prediction markets, should be part of an efficient election forecast. The academic literature is clear that combining data is generally very effective in increasing the accuracy (Clemen, 1989), yet, overall, three related, but largely non-intersecting, strands of the literature persist for the three data types. The closest work to this paper is that of Linzer (2013), which models state-by-state forecasts for the Electoral College and utilizes polling and fundamental data; however, among other differences, no previous study has also tested prediction market prices for these efficient outcomes. The forecast created in this paper appreciates the fact that the information provided by the three different forecast types shifts over the timeframe under study; 130 days out, the most efficient combined forecast averages the forecasts from all three data types, but the fundamental data's unique information decreases over time, so that the Election Day forecast averages just polling and prediction market data. Various stakeholders, including researchers, election investors, and election workers, can all benefit from the efficient combined forecast of Electoral College or senatorial elections defined in this paper; the forecast is tested and excels out-of-sample during the 2012 elections.

2. Data

This paper calibrates the combined forecast with data from 202 races that span four election cycles. Electoral College races are included for 2004 and 2008, and senatorial elections are included for 2006, 2008, and 2010. In this sample, there are 100 Electoral College races, or 50 for each cycle.[7] There are 33 senatorial elections in 2006, 35 in 2008, and 37 in 2010. I exclude three senatorial elections from the calibration because they are outliers, but include them in any derived forecasts.[8] The 33 senatorial elections and 50 Electoral College races of 2012 serve as an out-of-sample test for the combined forecast.[9] Data from the 2000 Electoral College races and 2004 senatorial elections help calibrate the poll-based forecast.

The polling data are as complete and accurate as possible. I gathered the voter intention poll data from all state-by-state polls listed on PollingReport.com, Pollster.com, and RealClearPolitics.com; I used several sites in order to ensure that I had as many public polls as possible. Pollster.com itself creates a rolling average that incorporates research such as that of Jackman (2005) and Pickup and Johnston (2005) on how to eliminate possible house biases in polls and draw updates from other state and national polls. Neither I, nor Pollster.com, have these aggregated averages historically. To create the historical polling snapshots, I utilized the most advanced, transparent aggregation methods, but I did not go back and correct for house bias historically, because any attempt to do so would suffer from a potential look-ahead bias, and the impact would be minimal. I ensure that this is not an issue by testing my 2012 out-of-sample results with both my method of aggregating polls and the Pollster.com aggregated average (when available); all results are robust to either.

The fundamental model follows the procedure outlined by Hummel and Rothschild (2013). The authors created the models for the Electoral College and senatorial elections separately, using data from the following categories: election results, incumbency, presidential approval ratings, economic indicators, ideological indicators, and biographical information about the candidates. The model is calibrated on Electoral College data from 1972 to 2008 and senatorial data from 1976 to 2010. All of the data for the fundamental model are available publicly on government websites, with exception of presidential approval ratings, which are gathered from Gallup and Pollster.com (for later years).

The prediction market data are all gathered from Intrade in real-time. It sells contracts for all candidates that are worth either $10 if that candidate wins or $0 if that candidate loses. I have a marginal order book for all elections from 2004, 2006, 2008, 2010, and 2012, saved at 10-min intervals throughout the election cycle. For each candidate, the data include: the prices people are willing to buy (bid) and sell, and the last price sold, along with the volume of trade. Intrade is the most liquid market for Electoral College and senatorial elections. Where the data are available, the results are also similar to Betfair, the world's largest prediction market, but I only have Betfair data for the 2010 and 2012 elections.

6 According to the Center for Responsive Politics, $6.3 billion was spent on the 2012 election, with $2.6 billion going to the presidential election alone: http://www.opensecrets.org/bigpicture.

7 I exclude Washington DC, which votes Democratic in all presidential elections with 100% likelihood.

8 There were three unique three-way senate elections in the sample: Connecticut in 2006 and Alaska and Florida in 2010.

9 This paper was circulated publicly prior to the 2012 election in order to ensure that the forecasts could be considered ex-ante.

In order to approximate a random draw of the demand for information, I use one observation per forecast type for each day. For the polling data, I date polls by the last day they are in the field; if anything, this is biased in favor of the relevancy of polls, as polls are frequently released at least one day after they leave the field. For the fundamental model used in this paper, the value is the same for the entire cycle. The model shows that there is little or no added information from including late-breaking fundamental data. For the prediction market data, I use the average of the bid and ask at noon on the noted day.

I analyze all data for the time period between 0 and 130 days before the election; 130 days generally falls in late June of the election year and is when the fundamental model is realized. The standard start of the campaign season in the United States is Labor Day, which falls roughly 60–65 days before the election, so this allows me to review a sizable length of time in the periods both before and after this.

Not all of the three forecast types occur on all days. First, there is a liquidity issue for prediction markets early in the cycle, where the lack of meaningful trading can lead to some imprecise pricing, although the markets all exist. Second, some senatorial elections never have major polls over the entire election cycle, while many do not start having polls until midway through the timeframe here. Third, I have a slight data issue where my program did not record any prediction market data for 10 random days in 2008.

There are three additional dimensions of the data to consider beyond the data type: the election type, the number of days before the election, and the certainty of the raw data. First, where the identification exists, I create different forecast models for the Electoral College and senatorial elections. Second, some parameters vary depending on the number of days until the election. Finally, some parameters vary depending on the certainty of the raw data. This can be important, because, for example, a prediction market price of near 50% may be correlated with the expected outcomes differently from a price near 95%, or a poll that gives a 2-point lead may behave differently from a poll with a 10-point lead.

The first step in this process is to create the most efficient forecasts possible for estimating the vote share and probability of victory for all three forecast types. I could go through the process of aggregating all information simultaneously into one combined forecasts, but there are two major advantages of first creating three separate forecasts and then combining them. First, it puts all of the data at the same scale, thus allowing us to understand how the information mixes in the combined forecast. Second, not all data are available at all times, so it is beneficial for forecasting to have both the separate and combined versions.

This paper emphasizes the probability of victory over the estimated vote share, but the estimated vote share is included so as to allow for historical comparisons.

3. Estimation strategy/results in creating separate forecasts

The fundamental model of Hummel and Rothschild (2013) utilizes an OLS regression for predicting the expected vote share and a probit regression for predicting the probability of victory. The models are capable of creating forecasts more than 130 days before the election for both types of elections. I encourage readers to review the study of Hummel and Rothschild (2013) if they wish to recreate the fundamental model, but, for reference, the Electoral College model has the following coefficients for Democratic vote share: 0.41 on (presidential approval − 42) ∗ incumbency, −1.88 on two or more terms ∗ incumbency, 0.72 on state vote four years ago − national vote, 0.12 on state vote eight years ago − national vote, 0.21 on change in state income from 9th to 13th quarter of the term, −0.02 on the sum of ACU rankings for senators − the average sum of ACU senators, 0.08 on the change in %Dems in lower house of state legislature, 4.83 on home state if less than 10 million in population, −3.04 on home state from last cycle if less than 10 million in population, and 47.37 as a constant. The senatorial model has the following coefficients for Democratic vote share: 0.13 on (presidential approval − 50) ∗ presidential incumbency, 12.82 on incumbency, −2.80 on midterm ∗ presidential incumbency, 0.36 on last presidential vote − national vote, 0.11 on state vote six years ago − national vote, 0.21 on change in state income from 9th to 13th quarter of the term ∗ presidential incumbency, 0.09 on Reps ACU rating − 74, 0.08 on change in %Dems in lower house of state legislature, between 3.87 and 10.39 on previous job of senatorial candidates, and 46.55 as a constant.[10]

The voter intention polling and prediction market models follow the procedures created by Rothschild (2009), but incorporate some major advances beyond that paper. I update the debiasing techniques (i.e., the translation of the aggregated raw data into a forecast) used for both forecast types. The polling debiasing technique varied by election type and days before the election for Rothschild (2009); this paper also allows variation by the size of the lead in the two-party vote share. For prediction markets, that article had a one-size-fits-all debiasing, but here I examine the value of allowing it to fluctuate along each of three dimensions: election type, days before the election, and the certainty of the raw pricing (i.e., how close the price is to \$0.50 rather than \$0.00 or \$1.00 for contracts that pay out \$1.00).

The first step in creating a voter intention poll-based forecast is to create a snapshot of the estimated two-party vote share of the two candidates (i.e., the support for candidate one divided by the sum of the support for candidates one and two) if the election were held that day. This aggregation helps to smooth the random fluctuation that occurs in raw daily polls. The method creates a linear regression of all polls up to that day, and the snapshot is the trend of that regression.[11] The data are de-meaned around zero, so that the two-party vote share runs from −0.5 to 0.5, where the leading candidate is above zero and the losing candidate is

[10] I exclude all coefficients that do not apply to the elections in 2012 and forward. I have included the coefficients for the probit, to create the probability of victory, in the Appendix.

[11] The linear trend is the simplest and most transparent method of creating a consistent poll average on any given day, especially in races with limited numbers of polls. The robustness of this method will be shown in the next section.

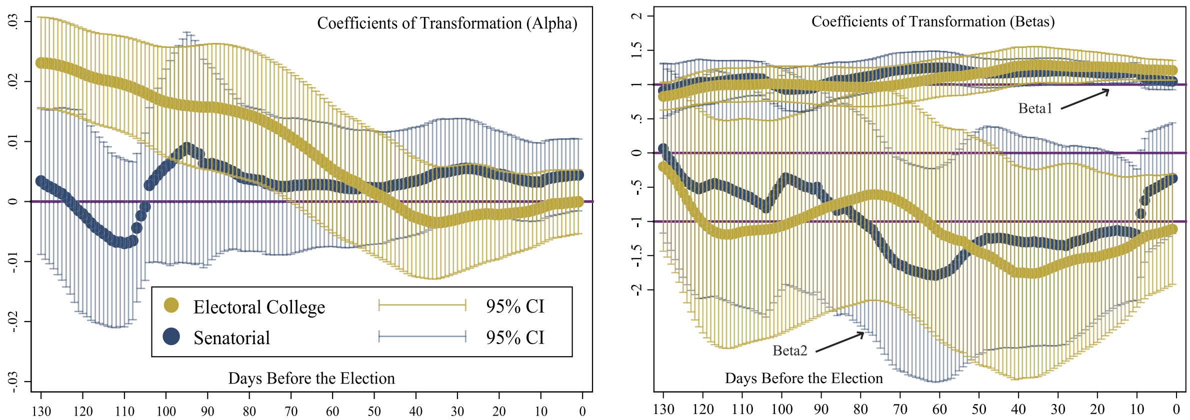

Fig. 1. Alpha (left) and beta (right) for Electoral College and senatorial elections, as derived in Eq. (1). Each point plots the value of the coefficient on a given day before the elections. An alpha greater than 0 represents an anti-incumbency bias; the incumbent gains points in expectation above their snapshot. A beta1 less than 1 represents a reversion to the mean and a beta2 less than 0 means that the reversion to the mean increases as the snapshot widens. The standard errors are clustered by race (i.e., state and year).

below zero. All vote shares are in terms of the state's incumbent (i.e., the winner of the state's electoral votes four years before and the state's senatorial election six years before).

The second step in creating a poll-based forecast is to create an estimated vote share for Election Day. To create the estimated vote share, I regress the final vote share on the poll for each day before the election in previous election years:

$$V_r = \alpha + \beta_1 S_r + \beta_2 S_r|S_r| + e_r, \qquad (1)$$

where r is a given race (i.e., state and year), and S is the snapshot. I use the absolute value of S_r, rather than the simpler square term, so that the β_2 term moves the estimate further from parity (if greater than 0) or closer to parity (if less than 0), rather than up or down in absolute terms for the incumbent. This β_2 term explains how the transformation varies as the snapshot indicates a wider distance between the candidates.[12] The regressions are calibrated using elections from 2000, 2004, and 2008 for the Electoral College, and 2004, 2006, 2008, and 2010 for the senatorial elections.[13] I calibrate the parameters separately for each election type and for each day before the election. Thus, I recover a unique α, β_1, and β_2 for each day before the election (T) and election type (q). The daily estimated vote shares are created as follows: $\hat{V}_r = \alpha_{T,q} + \beta_{1,T,q} S_r + \beta_{2,T,q} S_r|S_r|$, where the alpha term corrects for the anti-incumbency bias and the beta terms correct for reversion to the mean.

While the Electoral College has a meaningful and statistically significant anti-incumbency bias, the senatorial elections do not; furthermore, the results add a different perspective to the theory behind the anti-incumbency bias. The left side of Fig. 1 shows that a presidential candidate whose party won the state in the previous election can expect to gain one to two percentage points of the two-party vote share from his/her opponent, depending on the number of days before the election. This bias decreases towards zero a week or two after Labor Day. I therefore tested coding the Electoral College in terms of the national incumbent party and the state-by-state winner from the previous election. For the model, the data is in terms of the state-by-state incumbent, because it has lower forecast errors, which correlates with a more meaningful alpha term, relative to using the national incumbency. This is the first time anyone has addressed the anti-incumbency bias at the state level for the Electoral College or senatorial elections.[14]

The right side of Fig. 1 shows that the reversion to mean increases as the snapshot indicates a wider distance between the candidates; this is new to the literature as well. It is apparent from the large negative coefficients for the snapshot squared, while, if the plain snapshot's coefficient drifts away from one, it is actually above one. This novel result suggests that I can characterize reversion to the mean more by a 10-point lead in the polls preceding a narrow victory, than by a 2-point lead preceding a toss-up.[15]

The third step in creating a poll-based forecast is to create a probability of victory, which is the probability that the two-party vote share is greater than 50%. The method assumes that the actual vote share on Election Day is drawn from a normal distribution centered around the estimated vote share. For a given estimated vote share, the more accurate the estimation is, the tighter the distribution of true outcomes, and the higher the percentage of probable outcomes where the favored candidate has the greater number of votes. I determine the optimal sigma ($\sigma_{T,q}$) for each day and election type by running a probit of the binary

12 I use ±7 days of data for all of the parameters, in order to obtain consistency, relative to the daily random variation of Erikson and Wlezien (2008a).

13 The data are collected from: PollingReport.com, Pollster.com, and RealClearPolitics.com. I fill in missing data using the method of Wlezien and Erikson (2002), for historical data only, with the linear interpolation from the poll before and after any missing day.

14 I also tested both Electoral College and senatorial elections in terms of party affiliation, but that provided even less accurate forecasts.

15 While there are sizable standard errors on this coefficient, it is an extremely consistent, meaningful, and, for both types of elections, statistically significant finding for long periods of time.

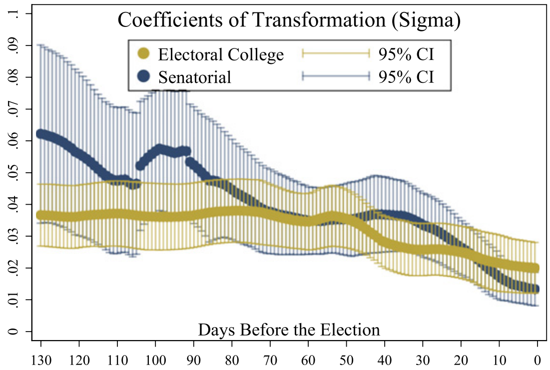

Fig. 2. Sigma for Electoral College and senatorial elections, derived in Eq. (2). Each point plots the value of the coefficient on a given day before the elections. The standard errors are clustered by race (i.e., state and year).

election outcomes on the expected vote share derived using the coefficients from Eq. (1):

$$P_r = \Phi(\widehat{V_{T,q}}/\sigma_{T,q}), \tag{2}$$

where P_r is the probability of victory for a given race and $\widehat{V_{T,q}}$ is the estimated vote share we derived using the coefficients from Eq. (1). Unsurprisingly, sigma gets smaller as the number of days before the election decreases, as shown in Fig. 2. It is impressive to see how small sigma, representing the standard deviation, becomes as Election Day approaches. For most of the early days of the cycle, the sigma coefficient is larger in the senatorial than in the Electoral College, because the estimated vote shares are less accurate, due to less polling and less accurate polling.[16]

The raw prediction market data translate into outcomes over a few steps. First, I take the average of the bid and ask for the stock that pays out if the Democrat wins on Election Day. If the bid–ask spread is greater than five points, I take the last sale price.[17] This is the raw prediction market price. Second, I correct for the favorite-longshot bias as per Rothschild (2009), using the transformation suggested by Leigh et al. (2007): $P_r = \Phi(1.64\Phi^{-1}(price))$.[18]

My datasets for examining prediction market data only include 2004 and 2008 for the Electoral College and 2006, 2008, and 2010 for senatorial elections. This limited dataset is used for producing the combined forecast because of the relatively limited availability of prediction market data, compared to polling and fundamental data. As a first step in determining the most efficient prediction market model, I duplicated the procedure of Leigh et al. (2007); that is, I took the probit $P_r = \Phi(\beta\Phi^{-1}(price))$ and regressed it over all of the prediction market data in my datasets and recovered $\beta = 1.67$; this is remarkably close to the 1.64 that Leigh et al. (2007) determined over a totally different dataset.

For both fundamental data and polling data, I create separate models for the different election types, but I do not have the necessary identification to enable prediction market prices to be split by election type. I do not want to over-estimate the coefficients, and I do not feel comfortable with just two election cycles for the Electoral College.

However, I do want to examine two other dimensions: the number of days before the election and the certainty of the price. The chart on the left side of Fig. 3 shows how that same coefficient, β, shifts with the number of days before the election (dbe) within the equation:

$$P_r = \Phi(\beta\Phi^{-1}(price)). \tag{3}$$

There is no smoothing between dbe in this chart; each point is a separate regression. The coefficient is amazingly close to 1.64 the entire time, until the last few days. Second, I checked to see how the coefficient would change as the price moves away from 50. I therefore ran Eq. (3) again for every price between 50 and 100, where I inverted prices below 50 to be above 50 (i.e., 30 became 70, or all prices are in terms of the most likely candidate).[19] The chart on the right side of Fig. 3 shows that the efficient coefficient is relatively stable around 1.64, regardless of the extremity of the price. The only bump is near the middle, where the data are much more sparse and the parameter has much less of an impact when it is applied (i.e., even an infinite coefficient does nothing at 50). Thus, in the absence of any compelling evidence to change it, I keep 1.64 as the same coefficient for debiasing all prediction market data.

4. Estimation strategy/results in combining forecasts

In comparing the three forecast types, we are limited to the overlapping elections of the data, namely 2004 and 2008 for Electoral College elections and 2006, 2008, and 2010 for senatorial elections; this is the dataset that I have for prediction markets, the narrowest dataset.

Liquidity is an issue through Election Day. From 130 days before the election, I have data for all three forecast types for all Electoral College elections. The fundamental forecast provides a forecast for all elections at all times. The prediction markets can be very thin early in the cycle, but they always provide a forecast; these early forecasts can provide relatively large errors in relatively easy to predict elections if there has been little action in the market. Polls are completely absent in some senatorial elections. Out of a possible 102 senatorial elections, polling data range from just 41 elections 130 days before the election (dbe), to 74 at 100 dbe, to 86 at 1 dbe. Early in the cycle they are absent in a somewhat random selection of elections, while they are likely to be absent in easy-to-predict elections late in the cycle.

The three forecast types differ in their accuracy as the number of days before the election decreases. Fig. 4 shows

[16] If there were a sizable correlated shock in a given cycle, then the standard deviation would be too small and the probabilities too confident. However, this method is standard, has worked historically, and is relatively robust to the sample size, and, in addition, I do not have enough cycles to derive these coefficients confidently out-of-sample.

[17] This procedure is adapted from Snowberg et al. (2006).

[18] This transformation was suggested (and estimated) by Leigh et al. (2007) prior to Rothschild (2009), using data from Presidential predication markets from 1880 to 2004.

[19] I use ±5 points of data in order to gain consistency.

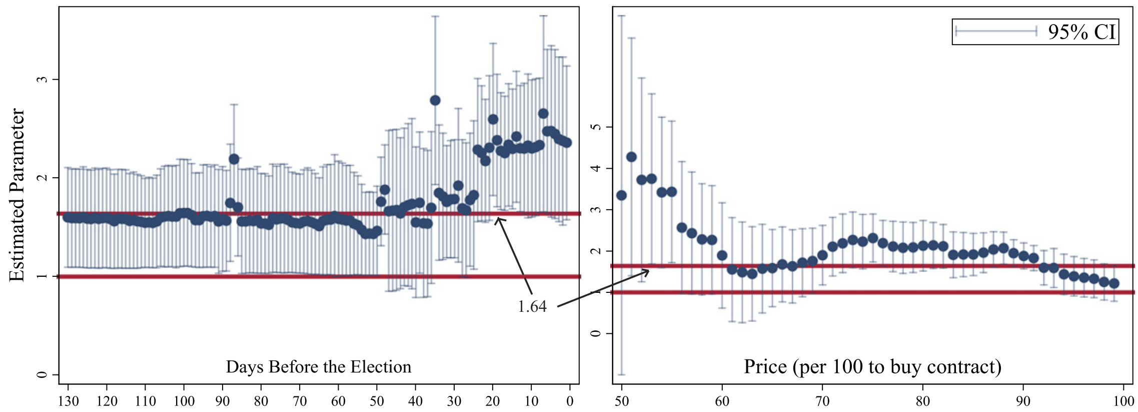

Fig. 3. Sigma for Electoral College and senatorial elections in Eq. (3). Each point in the left panel plots the value of the coefficient on a given day before the elections. Each point in the right panel plots the value of the coefficient at a given price from 50 to 100. The standard errors are clustered by race (i.e., state and year).

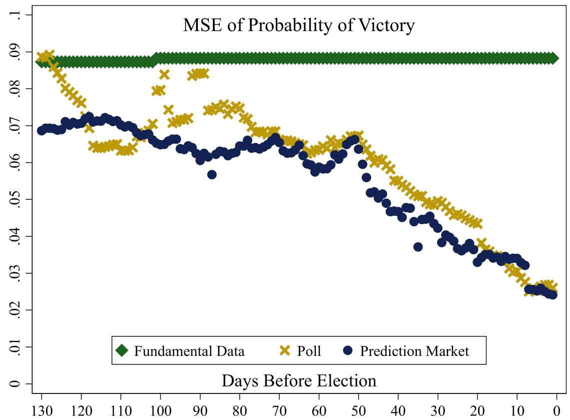

Fig. 4. Accuracy of probability of victory estimates for Electoral College and senatorial elections using fundamental data, voter intention polls, and prediction market-based forecasts. There are 202 observations per dbe for the combined and prediction market-based forecasts. If poll data do not exist, prediction market data take their place in the combined forecast. Poll-based forecasts vary from 141 to 186 observations per dbe.

the mean square error of the probability of victory, relative to the final outcome (i.e., 1 for a win and 0 for a loss), for both the Electoral College and senatorial elections. No one knows the true probability of victory 50 or 100 days before an election, but the mean square error relative to the final outcome is a key rubric for determining the accuracy of forecasts over time. The accuracy of the fundamental model never moves, because it does not shift during the election.[20] The other two forecast types are very similar, with prediction markets having slightly smaller errors, with pockets of time where the difference is statistically significant. Only as the campaigns move into the main timeframe, after Labor Day, do the errors on the poll and prediction market-based forecasts drop well below the fundamental model-based forecasts. The chart is extremely similar if I use only elections where all three data types have forecasts; prediction market-based forecasts have some periods of time with statistically significantly smaller errors than poll-based forecasts early in the cycle, and very similar errors after Labor Day. Consistent with the literature, raw prediction market prices regularly have higher errors than properly debiased prediction market-based forecasts, and not debiasing polls leaves raw polls with substantially larger errors than properly debiased poll-based forecasts.

When creating the combined forecast, I can combine the individual forecasts in many different ways, but I am even more concerned with ensuring that I avoid over-estimating the coefficients here than for the individual forecasts. Again, with only two presidential cycles, I avoid any attempt to separate the parameters by election type. The accuracy of the forecast should shift with time, being highly correlated with variations in the quantity of polls available and the liquidity in prediction markets. In the interests of simplicity, I do not allow the aggregation to vary for different forecast types within a given day. Thus, I focus exclusively on the number of days before the election, which Fig. 4 shows should be a major factor.

I combine the forecasts of the probability of victory very directly with a probit of the inverse normal of their probabilities:

$$P_r = \Phi(\Sigma\beta + \Sigma\gamma\,\text{dbe})(\Phi^{-1}(P_{r,F}) + \Phi^{-1}(P_{r,Poll}) + \Phi^{-1}(P_{r,PM})), \tag{4}$$

where I allow the parameters to shift linearly according to the dbe.[21]

This produces a very clean result; when everything is added together, the coefficients vary from an approximately equal weighting of each forecast on 130 dbe, to averaging just the poll and prediction market-based forecasts on Election Day. By the definition of the model, the weight on the fundamental data moves in a linear manner. The coefficients derived are given in Table 1.

[20] The small bump around day 100 is the addition of the West Virginia senate race in 2010, which did not exist until Senator Robert Byrd's death after the cycle had already begun.

[21] Allowing the coefficients to vary daily was an over-estimation. By regressing across time, I introduce correlated error, so that I cluster by race. Since individual races are not always independent, this should be taken as a lower bound of error, but that should not affect the point estimates of the coefficients.

Table 1
The coefficients for combining the three forecast types to give a combined probability of victory, Eq. (4): $P_r = \Phi(\Sigma\beta + \Sigma\gamma\,\text{dbe})(\Phi^{-1}(P_{r,F}) + \Phi^{-1}(P_{r,Poll}) + \Phi^{-1}(P_{r,PM}))$. The standard errors are clustered by race (i.e., state and year).

Variable	Coefficients for the probability of victory
Fundamental data	0.229 (0.125)
Polling	0.645* (0.247)
Prediction market	0.726* (0.210)
Fundamental data ∗ dbe	0.003* (0.001)
Polling ∗ dbe	−0.001 (0.003)
Prediction market ∗ dbe	−0.003 (0.003)

* Denotes significance at the 5% level.

To keep everything simple and to avoid over-estimating the model, I simplify a few things when I use this model to create out-of-sample real-time forecasts. First, I drop insignificant variables, which leaves just β for polls and prediction markets, and γ for fundamental data. Second, I recalibrate the coefficients so that they sum to 1 at any given dbe; I do this because I have no reason to believe that they should go beyond 1 and I want to be conservative until I have more election cycles. Finally, as they are already very close, I round the coefficients to exactly 0.333 each at 130 days before the election and allow them to move linearly until they are 0.5 each for the polling and prediction market-based forecasts, and 0 for fundamental-based forecasts, on Election Day.

So far, the results have been within-sample, so, despite the simplifications of the coefficients, it is almost by definition that the combined forecast is more accurate than the individual forecasts.[22] As Fig. 5 shows, it is not surprising that combining the three forecasts to give the combined forecast provides the greatest potential benefit early in the cycle, where the fundamental data are involved heavily and there is wide variation in the accuracy of the forecasts. Late in the cycle, the poll and prediction market forecasts converge, and thus, by definition, they converge with a combined forecast which is mainly a combination of the two forecasts. However, the combined forecast generally performs better than any of the individual forecasts for almost any given dbe. The combined forecast is equally dominant if I only use elections where we have all three data types.

Researchers should judge forecasts on their calibration, as well as on the size of the errors. Calibration is unique to the probability of victory; it measures how often an event occurs relative to the forecasted probability. For example, if a properly calibrated forecast declares that 100 events have about a 75% probability of occurring, then seventy-five of the events should occur. The goal of calibration is to see how well the forecast discriminates certainty; this metric actually rewards a well calibrated forecast of a toss-up, while the error would be large regardless of the outcome. Fig. 6 shows the relationship between the probability of the leading candidate winning (so that all forecasts are between 50% and 100%) and the percentage of elections won

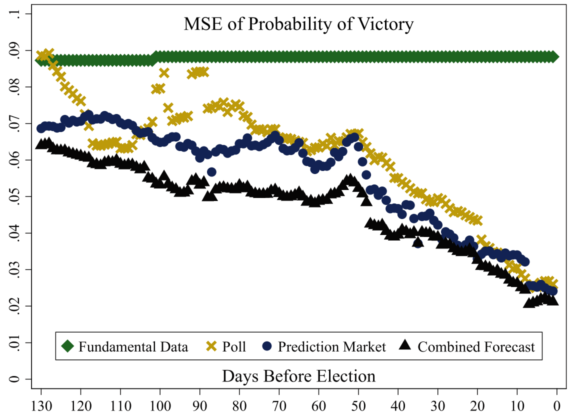

Fig. 5. Accuracy of the probability of victory estimates for Electoral College and senatorial elections by fundamental data, voter intention polls, and prediction markets-based forecasts, along with the combined forecasts. There are 202 observations per dbe for the combined and prediction market-based forecasts. If poll data do not exist, prediction market data take their place in the combined forecast. Poll-based forecasts vary from 141 to 186 observations per dbe.

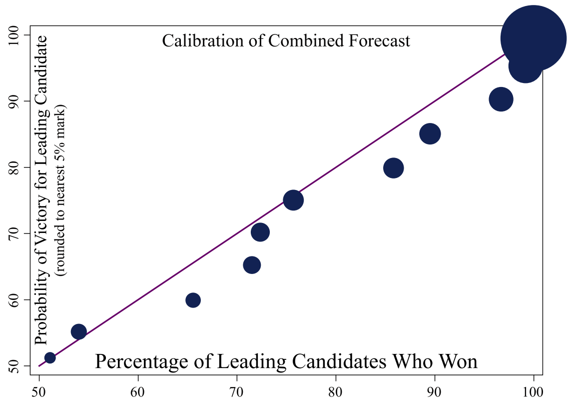

Fig. 6. Calibration of the probability of victory estimates for Electoral College and senatorial elections by the combined forecast. There are 202 observations per dbe for the combined forecasts, and this chart is aggregated over the final 130 dbe. If poll data do not exist, prediction market data take their place. Poll-based forecasts vary from 141 to 186 observations per dbe. The sizes of the circles are correlated with the numbers of observations in each bucket. The bucket sizes range from 315 near 50% to 16,303 near 100%. The number increases monotonically.

by the leading candidate. Every probability is rounded to the nearest 5% mark and the sizes of the circles are correlated with the numbers of observations in each bucket. A well-calibrated forecast is close to the 45° line shown in purple; the combined forecast is generally close to the 45° line as you move from the lowest probabilities, which occur with low frequency, to the highest probabilities, which occur with high frequency.

I do not want this paper to dwell too much on a single year, as the correlations between the outcomes decrease the explanatory power of even 83 different outcomes somewhat.[23] However, the combined forecast does perform well for predicting the 2012 election out-of-sample; Fig. 7 shows the errors every 4 h for the last 130 days before

22 I say 'almost by definition' because I simplified the parameters from the regression results, which will increase the mean square error of the forecast.

23 We drop DC's Electoral College election.

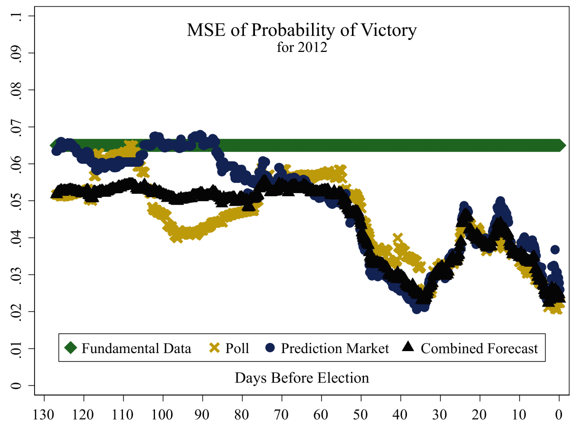

Fig. 7. Accuracy of the probability of victory estimates for Electoral College and senatorial elections by fundamental data, voter intention polls, and prediction market-based forecasts, along with the combined forecasts for 2012. There are 83 observations per dbe for all forecasts, at all dbe.

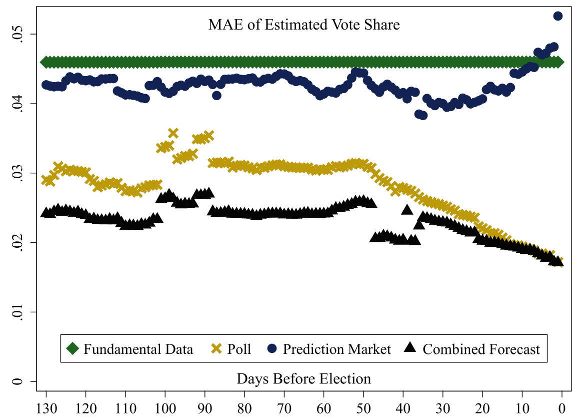

Fig. 8. Accuracy of the estimated vote shares for Electoral College and senatorial elections by fundamental data, voter intention polls, and prediction market-based forecasts, along with the combined forecast. There are 202 observations per dbe for the combined and prediction market-based forecasts. If poll data do not exist, fundamental data take their place in the combined forecast. Poll-based forecasts vary from 141 to 186 observations per dbe.

the election in 2012. Again, I created the model for the combined forecast prior to the 2012 election; a publicly available website published the combined forecast during the leadup to the election and updated it every few minutes.[24] Unlike the within-sample years, it was not dominant at every point in the cycle, but it was the most consistent forecast. For a period of about 30 days early in the cycle when poll-based forecasts had lower errors than prediction market-based forecasts, the combined forecast was either below or near the poll-based forecast. Over the period from near the end of the summer until the last month of the campaign, a span of about 45 days when the prediction market-based forecast had lower errors than polls, the combined forecast again held closely to the lowest errors. At any given moment from 130 days before the 2012 election to Election Day, the combined forecast is likely to have a lower error than either the poll-based or prediction market-based forecasts.[25]

5. Estimation strategy/results in creating separate forecasts: expected vote share

Both fundamental data (in Hummel & Rothschild, 2013) and polling (within the last section in Eq. (1)) have already been translated into estimated vote shares. Prediction market data translate into estimated vote shares by regressing the inverse of the price on the vote share and predicting the value:

$$V_r = \beta\Phi^{-1}(price). \qquad (5)$$

There should be no meaningful distinction between the probabilities of a big certain win and a small certain win (i.e., if Candidate A is estimated to receive 51% of the vote in her election and Candidate B is estimated to receive 75% of the vote in a different election, but both are certain to win their respective elections, both of their prediction market prices will approach \$1.00 for \$1.00 contracts, providing no identification for their estimated vote shares).

The combined forecast of the expected vote share is determined in the same manner as the probability of victory. I use the following regression:

$$V_r = (\Sigma\beta + \Sigma\gamma\,\text{dbe})(\widehat{V_{r,F}} + \widehat{V_{r,Poll}} + \widehat{V_{r,PM}}), \qquad (6)$$

where I allow the coefficients to shift linearly based on the dbe. All coefficients, except for the β for prediction markets, are highly significant. The coefficients derived are given in Table 2. The poll-based forecast is weighted the most for the entire length of the campaign, starting with just a little more weight than the fundamental-based forecast 130 days before the election and increasing to over 80% of the weight by Election Day. The absolute error of the resulting combined forecast is compared with those of the three individual forecasts in Fig. 8. Unsurprisingly, the forecast converges with the poll forecast as Election Day approaches.[26]

6. Conclusion

This paper combines three forecasts based on polling data, fundamental data, and prediction market data. This combined forecast accounts for shifting levels of information by allowing the parameters to adjust with the relative values of the separate data types over time. The paper

24 The combined forecast had well over 1,000,000 pageviews between October 1, 2012 and Election Day 2012.

25 The combined forecast and the prediction market-based forecast both use the average of the raw prices from Betfair and Intrade, when both markets were available. The combined forecast and the poll-based forecast are tested for both the poll aggregation method described in this paper and the more complicated aggregation method utilized by Simon Jackman for Pollster.com, where both are then translated into a forecast using the same method; while the method used in this paper has smaller errors on average, the differences between the two aggregation methods are not significant to the findings.

26 This result validates and expands the findings of Erikson and Wlezien (2012), who determined that prediction markets do not provide information for estimating the national popular vote. Of course, a central thesis of this paper is that that is not the correct outcome variable for relevant election forecasts.

Table 2
The coefficients for combining the three forecast types into a combined estimated vote share, Eq. (6): $V_r = (\Sigma\beta + \Sigma\gamma \text{ dbe})(\widehat{V_{r,F}} + \widehat{V_{r,Poll}} + \widehat{V_{r,PM}})$. The standard errors are clustered by race (i.e., state and year).

Variable	Coefficients for the estimated vote share
Fundamental data	0.173* (0.036)
Polling	0.886* (0.072)
Prediction market	0.001 (0.051)
Fundamental data ∗ dbe	0.001* (0.000)
Polling ∗ dbe	−0.004* (0.001)
Prediction market ∗ dbe	0.002* (0.001)

* Denotes significance at the 5% level.

shows that all three data types are meaningful and significant in an efficient forecast of state-by-state elections; to create efficient forecasts, it is crucial that researchers be agnostic toward differences in data type. The benefits of combining are highest earlier in the cycle, as polls and prediction markets converge towards Election Day, and thus, the combined forecast becomes very close to the two forecasts.

A secondary contribution of the paper is to provide new insights into the translation of raw polling and prediction market data into forecasts. For example, the anti-incumbency bias for polls does not extend to senatorial elections. This is probably due to the longer time period between elections that allows for more change, and the massive jumps between parties that can occur with periodic retirements. In another example, the favorite-longshot bias for prediction markets does not vary significantly with time. In theory, there are more concerns about liquidity the longer the contract is held, as the opportunity costs increase and the liquidity in the market is lower, which should increase the favorite-longshot bias with the number of days before the election. However, this does not seem to have any impact within a 130-day window.

Some researchers may be surprised that fundamental data provide anything which is not found in either polls or prediction markets, but in actual fact fundamental data provide two things: stability and liquidity. First, fundamental forecasts provide stability, when prediction markets or polls may be chasing short-term fluctuations in sentiment. This was certainly the case in the 2012 presidential election, when Mitt Romney's bump after the first debate ultimately dissipated. Second, neither polls nor prediction markets are that well formed 130 days before the election. Only after Labor Day are the polls and prediction markets both fully liquid.[27]

Some researchers may be surprised that prediction markets add any information beyond polls, but there are three types of data to consider. First, prediction markets can incorporate the impacts of major events that polls take several days to acquire (e.g., the release of a secret video). Second, prediction markets can incorporate the impacts of events that have not yet occurred, but that users know will impact the electorate (e.g., the likely impact of a debate). Finally, prediction markets can aggregate idiosyncratic information about the election from self-selected users with high levels of information that poll respondents are not able to incorporate into their poll responses, or do not have access to (e.g., major differences in get-out-the-vote efforts).

The value of the regularly updating nature of the forecast, which can only be provided by prediction market data, was seen in 2012. Forecasts based only on fundamental and/or polling data would not update around the conventions or debates. With only the Labor Day weekend between the Republican and Democratic debates, it was difficult for most polling companies to field polls that showed the impacts of the conventions separately. Similarly, while observers could speculate on the way in which the polls would move in the days following Romney's triumph over Obama in the first debate, the combined forecast moved during the debate and in its immediate aftermath.

Finally, some researchers may be surprised that polls add any information over and above prediction markets. In theory, prediction markets should already include all of the latest polling information. It is possible that some of the prediction market data suffer from enough liquidity issues at various points in the cycle that aggregation with other data types is beneficial in correcting those issues. Furthermore, prediction markets are prone to small price shifts from users whose incentive is not to maximize their return in the market, but to hedge other investments or to influence the campaign (Rothschild & Sethi, 2013). This paper does not challenge the efficient market principle in general, but shows empirically that polls are necessary to fill in holes in the information which powers a forecast that runs continuously.

The combined forecast is a practical forecast; a publicly available website published the forecasts from this model over the election cycle. Publishing the forecast during the campaign not only ensured that 2012 was cleanly out-of-sample, but also forced me to consider data issues that can be ignored if the forecast is run ex-post. For example, some states do not hold their senatorial primaries until the second week in September. Since senatorial polling is candidate-to-candidate, this makes using polling data early in the cycle difficult, as I had to weigh forecasts based on the likelihood of victory in the primary, but is no problem after the election, when a forecaster can determine with certainty which matchup occurred.

More refined combinations of data are certainly possible, but I am wary of over-estimating my models. There are not enough elections to enable us to identify overly intricate models, nor is there enough independence within elections. As it is, my standard errors could be slightly higher, due to correlations between outcomes. That is why I round my coefficients for the combined forecast and try to be as conservative as possible with these results. While the question of standard errors is not very germane to my generalized results, it should serve as a warning to future research that may attempt to estimate their coefficients too tightly.

The model in this paper is designed to be easy to duplicate in real time for future elections. Future testing will

27 Fundamental forecasts are still interesting in October for explaining the correlations between the world and the election (e.g., people want to know the expected impact of new economic indicators). However, they should not be confused with updated and accurate forecasts.

consider the benefits of allowing the coefficients for combining the forecasts to vary by race, based on metrics of certainty. This would require additional data that I do not have historically; then, having collected that data, I would have to weigh any possible accuracy benefits against the offsetting costs of data collection, running the model in real-time, and possible over-estimation. The central question would be whether or not there is enough identification to enable the accurate modeling of shifting weights both by day, as is currently done, and by certainty of the data types within day.

This paper takes steps to clarify the objectives of creating forecasts in order to provide relevant stakeholders with the most efficient forecasts. This includes defining the relevant outcome variable as the probability of victory at a state-by-state level, ensuring that the forecasts update continuously during the course of the main election cycle, and testing the calibration, along with traditional measures of error.

I have framed this paper with the aim of making the most relevant, timely, and accurate forecasts, but this ignores a key fourth component: cost efficiency. Fundamental forecasts are extremely costly, in that researchers need to derive new models for every outcome. Standard polling is becoming increasingly expensive as response rates plummet. Polling does exist in senatorial races, but is extremely sparse. Prediction markets can expand with no marginal cost, but may have trouble in getting liquidity for some events. Overall, prediction markets are the most cost efficient, but different situations will cause this relationship to vary. As we move into a digital age, with more data and new ways to contact people, we should be mindful that more cost efficient forecasts will help us to answer more questions in more domains. This will add efficiency to our decisions and allow researchers to answer new questions.

Acknowledgments

I thank the editors and reviewers for their extremely helpful criticisms, and Sunshine Hillygus and Jolie Martin for their insightful comments.

Appendix. Coefficients for the fundamental model

The coefficients for the expected vote share from Hummel and Rothschild (2013) are in the main text; here, I give the coefficients for the probability of victory which should be used as input for a probit. The Electoral College model has the following coefficients for a Democratic likelihood of victory: 0.15 on (presidential approval − 42) $*$ incumbency, −0.65 on two or more terms $*$ incumbency, 0.22 on state vote four years ago − national vote, 0.07 on state vote eight years ago − national vote, 0.15 on change in state income from 9th to 13th quarter of term, −0.004 on sum of ACU rankings for senators − the average sum of ACU senators, 0.05 on change in %Dems in lower house of state legislature, 1.71 on home state if less than 10 million in population, and −0.94 as a constant. The senatorial model has the following coefficients for a Democratic likelihood of victory: 0.02 on (presidential approval − 50) $*$ presidential incumbency, 1.46 on incumbency, −0.50 on midterm $*$ presidential incumbency, 0.06 on last presidential vote − national vote, 0.02 on state vote six years ago − national vote, 0.05 on change in state income from 9th to 13th quarter of term $*$ presidential incumbency, between 0.66 and 1.13 on previous job of senatorial candidates, and no constant. There are fewer variables in the probability of victory versus the estimated vote share, because some variables are significant for the size of victory, but not the binary outcome.

References

Abramowitz, A. I. (2004). When good forecasts go bad: the time-for-change model and the 2004 presidential election. *PS: Political Science and Politics*, *37*(4), 745–746.
Abramowitz, A. I. (2008). Forecasting the 2008 presidential election with the time-for-change model. *PS: Political Science and Politics*, *41*(4), 691–695.
Armstrong, J. S., Green, K. C., Jones, R. J., & Wright, M. J. (2010). Predicting elections from politicians' faces. *International Journal of Public Opinion Research*, *22*(4), 511–522.
Berg, J., Forsythe, R., Nelson, F., & Rietz, T. (2008). Results from a dozen years of election futures markets research. In *Handbook of experimental economics results*: *Vol. 1* (pp. 742–751).
Campbell, J. E. (2008). *The American campaign: US presidential campaigns and the national vote*: *Vol. 6*. Texas A&M University Press.
Campbell, J. E. (2012). Forecasting the 2012 American national elections. *PS: Political Science and Politics*, *45*(4), 610–613.
Clemen, R. T. (1989). Combining forecasts: a review and annotated bibliography. *International Journal of Forecasting*, *5*(4), 559–583.
Erikson, R. S., & Wlezien, C. (2008a). Are political markets really superior to polls as election predictors? *Public Opinion Quarterly*, *72*(2), 190–215.
Erikson, R. S., & Wlezien, C. (2008b). Leading economic indicators, the polls, and the presidential vote. *PS: Political Science and Politics*, *41*(4), 703–707.
Erikson, R. S., & Wlezien, C. (2012). Markets vs. polls as election predictors: an historical assessment. *Electoral Studies*, *31*(3), 532–539.
Graefe, A., Armstrong, J. S., Jones, R. J., Jr., & Cuzán, A. G. (2014). Combining forecasts: an application to elections. *International Journal of Forecasting*, *30*, 43–54.
Hummel, P., & Rothschild, D. (2013). Fundamental models for forecasting elections. ResearchDMR.com/HummelRothschild_FundamentalModel.
Jackman, S. (2005). Pooling the polls over an election campaign. *Australian Journal of Political Science*, *40*(4), 499–517.
Klarner, C. (2008). Forecasting the 2008 US house, Senate and presidential elections at the district and state level. *PS: Political Science and Politics*, *41*(4), 723–728.
Klarner, C. E. (2012). State-level forecasts of the 2012 federal and gubernatorial elections. *PS: Political Science and Politics*, *45*(4), 655–662.
Leigh, A., Wolfers, J., & Zitzewitz, E. (2007). Is there a favorite-longshot bias in election markets? *Preliminary version presented at the 2007 UC riverside conference on prediction markets, Riverside, CA*.
Lewis-Beck, M. S., & Rice, T. W. (1992). *Forecasting elections*. CQ Press.
Linzer, D. A. (2013). Dynamic Bayesian forecasting of presidential elections in the states. *Journal of the American Statistical Association*, *108*(501), 124–134.
Lock, K., & Gelman, A. (2010). Bayesian combination of state polls and election forecasts. *Political Analysis*, *18*(3), 337–348.
Manski, C. F. (2006). Interpreting the predictions of prediction markets. *Economics Letters*, *91*(3), 425–429.
Montgomery, J. M., Hollenbach, F. M., & Ward, M. D. (2012). Improving predictions using ensemble Bayesian model averaging. *Political Analysis*, *20*(3), 271–291.
Pickup, M., & Johnston, R. (2005). Measurement error and house bias in 2004 presidential campaign polls. In *Annual meeting of the American Political Science Association, Washington, DC, September* (pp. 2–5).
Rhode, P. W., & Strumpf, K. S. (2004). Historical presidential betting markets. *The Journal of Economic Perspectives*, *18*(2), 127–141.
Rothschild, D. (2009). Forecasting elections comparing prediction markets, polls, and their biases. *Public Opinion Quarterly*, *73*(5), 895–916.

Rothschild, D. M., & Sethi, R. (2013). *Trading strategies and market microstructure: evidence from a prediction market*. Available at SSRN 2322420.
Snowberg, E., & Wolfers, J. (2010). *Explaining the favorite-longshot bias: is it risk-love or misperceptions?* Tech.rep. National Bureau of Economic Research.
Snowberg, E., Wolfers, J., & Zitzewitz, E. (2006). *Party influence in congress and the economy*. Tech.rep. National Bureau of Economic Research.
Wlezien, C., & Erikson, R. S. (2002). The timeline of presidential election campaigns. *The Journal of Politics*, *64*(4), 969–993.
Wolfers, J., & Zitzewitz, E. (2006). *Interpreting prediction market prices as probabilities*. Tech.rep. National Bureau of Economic Research.