Origins of Presidential poll aggregation: A perspective from 2004 to 2012

Samuel S.-H. Wang

Princeton Neuroscience Institute and Department of Molecular Biology, Princeton University, Princeton, NJ 08544.

Address for correspondence:
    Sam Wang
    Princeton Neuroscience Institute, Washington Road
    Princeton University
    Princeton, NJ 08544
    sswang@princeton.edu
    Telephone: 609-258-0388
    Fax: 609-258-1028

Abstract

U.S. political reporting has become extraordinarily rich in polling data. However, this information has not been matched by an improvement in the accuracy of poll-based news stories, which usually examine a single survey at a time without providing an aggregated, more accurate view. In 2004 I reduced polling noise for the Presidential race by developing a meta-analysis that reduced all available state polls into a single snapshot in time, the Electoral Vote estimator. This index, based on polling medians alone, has an accuracy equivalent to less than ±0.5% in national popular-vote margin and outperforms the aggregator FiveThirtyEight and the betting market InTrade. Polling corrections and econometric variables can still improve accuracy, but at the expense of transparency. For purposes of public communication, such improvements should be treated with caution.

1. INTRODUCTION

In 2012, polling aggregation entered the public spotlight as never before. Typically, political horserace commentary in the US is dominated by pundits who are motivated by pressure not to be accurate, but to attract readers and viewers. For example, one day before the 2012 U.S. presidential election, former Reagan speechwriter Peggy Noonan (2012) wrote that "nobody knows anything" about who would win, asserting that Republican candidate Mitt Romney's supporters had the greater passion and enthusiasm. Columnist George Will predicted a Romney electoral landslide (Poor, 2012).

In the end, the aggregators were correct. Pundits largely failed to report the fact that based on public opinion polls with collectively excellent track records, President Obama had an

advantage of 3 to 4 percentage points for nearly the entire campaign season. By ignoring the data, many commentators expressed confidence—and were wrong.

In this article I describe an early approach to the aggregation of Presidential state polls, the meta-analytic method used at the Princeton Election Consortium (PEC; http://election.princeton.edu) since 2004. PEC's approach uses Electoral College mechanisms and can be updated on a daily basis. Its only input is publicly available data and it runs on open-source software, thus offering a high level of transparency. I will describe this method, and give both public and academic perspectives (see also Jones, 2008 for a review).

Outperforming the pundits has been possible at least since 2004, when a number of websites began to aggregate and report polls on a state-by-state basis in Presidential, Senate, and House races. For the Presidency, state polls are of particular interest for three reasons. First, the Presidency is determined via the Electoral College, which is driven by state election outcomes. Second, state polls have the advantage of being, on average, accurate predictors of state election outcomes (**Figure 1a**). National polls can have significant inaccuracies. For example, in 2000 Al Gore won the popular vote by 0.5% over George W. Bush, yet Election-Eve national polls favored Bush by an average of 2.5%, a 3.0% error that got the sign of the outcome wrong. State polls may owe their superior accuracy to the fact that local populations are less complex demographically, and therefore easier to sample, than the nation as a whole. Third and last, state presidential polls are also remarkably abundant: Electoral-vote.com contains 879 polls from 2004, 1189 from 2008, and 924 from 2012.

Early sites – RealClearPolitics in 2002, followed in 2004 by Andrew Tanenbaum's Electoral-vote.com, the Princeton Election Consortium, and several others (Forelle, 2004a) – reported average or median polling margins (i.e. the percentage difference in support between

the two leading candidates) for individual races. An additional step was taken by PEC (then title "Electoral College Meta-Analysis," http://synapse.princeton.edu/~sam/pollcalc.html), which calculated the electoral vote (EV) distribution of all possible outcomes, using polls to provide a simple tracking index, the EV estimator. The calculation, an estimate of the EV outcome for the Kerry v. Bush race, was updated in a low-graphics, hand-coded HTML webpage and a publicly posted MATLAB script. PEC gained a following among natural scientists, political and social scientists, and financial analysts. Over the course of the 2004 campaign, PEC attracted over a million visits, and the median decided-voter calculation on Election Eve captured the exact final outcome (Forelle, 2004b).

In 2008, a full PEC website, unveiled under the banner "A first draft of electoral history," provided results based on decided-voter polling from all 50 states, as well as Senate and House total-seat projections. In the closing week of the campaign, PEC ended up within 1 electoral vote of the final outcome, within 1 seat in the Senate, and exactly correct in the House.

The same year, many other aggregators emerged on the scene. The website 3BlueDudes.com documented at least 45 different hobbyists in 2008. One site rapidly emerged as the most popular: FiveThirtyEight. Created by sabermetrician Nate Silver, FiveThirtyEight arose from his efforts at the liberal weblog DailyKos. Silver initially attracted attention for his analysis of the Democratic nomination contest between Hillary Clinton and Barack Obama. In the general election season, Silver provided a continuous feed of news and lively commentary, as well as a prediction of the November outcome based on a mix of econometric assumptions and polling data. FiveThirtyEight was later licensed to the *New York Times* from 2010 to 2012, becoming a major driver of traffic to the *Times* website (Tracy, 2012).

In the academic sector, econometric and polling data have both been used to study Presidential campaigns. Most academic research has focused on time scales of months or longer, usually concentrating on explaining outcomes after the election, or on making a prediction before the general election campaign has begun. Predictions are usually done in the spirit of testing provisional models which then are subject to change (for review see Lewis-Beck and Tien, 2008; Abramowitz, 2008; Jones, 2008; and articles in the current issue of the International Journal of Forecasting).

Polls-only analysis has been done by Gelman and King (1993), who analyzed time trends from national polling data, and since 1996 Erikson and Wlezien (2012) have constructed detailed time series to give post-hoc trajectories of national campaigns. Using Electoral College mechanisms and state polls, Soumbatiants (Soumbatiants, 2003; Soumbatiants et al., 2006) calculated a distribution of probable EV outcomes using Monte Carlo simulation and examined the effects of a hypothetical single-state or nationwide shift in opinion. These scenarios have also been explored from the point of view of a campaign (Strömberg, 2002) or of an individual voter (Gelman et al., 2010). Strömberg (2002) correctly noted the pivotal nature of Florida in the final outcome, and found that campaigns allocated resources in a manner that scaled with the influence of individual states.

In 2012, day-to-day forecasting took three forms. First, the Princeton Election Consortium took a polls-only approach. A second approach was taken by Drew Linzer (http://votamatic.org; Linzer, 2013), who used pre-election variables to establish a prior win probability and updated this in a Bayesian manner using new polling data. The resulting prediction was notably stable for the entire season. Extensive modeling was also done by Simon Jackman and Mark Blumenthal at the Huffington Post (Jackman and Blumenthal, 2013). In the

5

public sphere, FiveThirtyEight combined prior and current information to create a measure that contained mixed elements of both snapshot and prediction in a single measure.

2. DATA

The PEC core calculation is based on publicly available Presidential state polls, which are used to estimate the probability of a Democratic/Republican win on any given date. These are then used to calculate the probability distribution of electoral votes corresponding to all $2^{51} = 2.3$ quadrillion possible state-level combinations.

Data sources and scripts. Polling data came from manual curation (2004), an XML feed from Pollster.com (2008), and a JSON feed from Huffington Post / Pollster (2012). In all cases the data source was selected to include as many polling organizations as possible, with no exclusions. When both likely-voter and registered-voter values were available for the same poll, the likely-voter result was used. For the District of Columbia no polls were available and the win probability was assumed to be 100% for the Democratic candidate. All scripts for data analysis and graphics generation were written in MATLAB and python, posted at http://election.princeton.edu, and deposited at the github software archive. In 2004, updates were done manually. In 2008 and 2012, updates were done automatically from July to Election Day. Update frequency increased with time, with up to six updates per day in October.

3. METHOD

3.1 An exact calculation of the probability distribution

The win probability for any given state *s* at time *t* is termed $p_s(t)$ and assumed to be predicted by the polling margin. Polling margins and analytical results were reported using a

positive number to indicate a Democratic advantage. For any given date, p_s was determined

using the most recent 3 polls, or 1 week's worth of polls, whichever was greater. A poll's date

was defined as the middle date on which it took place and if the oldest two had the same date,

four polls were used. The same pollster could be used more than once for a given state if the

samples contained nonoverlapping respondent populations.

From these inputs, a median margin (M_s) was calculated. The estimated standard error of

the median (σ_s) was calculated as $SD_s = 1.485*(\text{median absolute deviation})/\sqrt{N}$. The Z-score,

M_s/σ_s, was converted to a win probability p_s (**Figure 1b**) using the t-distribution.

The probability distribution of all possible outcomes, *P(EV)* (**Figure 1c**), was calculated

using the coefficients of the polynomial

$$\prod ((1 - p_s) + p_s \, x^{E_s}), \tag{1}$$

where $s=1\dots51$ representing the 50 states and the District of Columbia, E_s is the number of

electoral votes for state s, and x is a dummy variable. In this notation, the coefficient of the x^N

term is the probability of winning a total of N electoral votes, *P(EV=N)*. The fact that in

Nebraska and Maine electoral votes are assigned on a district-by-district basis was not taken into

consideration. The median of P was used as the EV estimator.

For modeling Senate outcomes, the same approach was taken with $E_s=1$ for all races. In

addition, for modeling House outcomes, races were scored as $p=0.5$ for toss-ups as defined by

Pollster.com and set to $p=0$ or $p=1$ otherwise, giving a 68% confidence interval of $\pm\sqrt{N}$ seats for

N toss-up races.

3.2 A polling bias parameter and the Popular Meta-Margin

The snapshot win probability, defined as the probability of one candidate getting 270 or more out of 538 electoral votes, was usually over 99% for one candidate or the other on a given day. A quantity that varied more continuously, and was therefore more informative, was the Popular Meta-Margin (*MM*). *MM* was defined as the amount of opinion swing, spread equally across all polls, that would bring the Median Electoral Vote Estimator to a 269-269 tie. To identify the tie point, *P(EV)* was calculated by varying the offset *x* over a range, i.e. by replacing M_s with M_s+x (**Figure 1d**).

The Meta-Margin allows the analysis of possible biases in polls. For example, if polls understate support for one candidate by 1%, this would reverse the prediction if the Meta-Margin were less than 1% for the other candidate. As a second example, if 1% of voters switch from one candidate to the other, this generates a swing of 2% and can compensate for a Meta-Margin of 2%. In this way, the Popular Meta-Margin is equivalent to the two-candidate difference found in single polls, allows evaluation of a wide variety of polling errors, and provides a mechanism for making predictions.

3.3 Prediction of November outcomes

Prediction for 2012 was done assuming that random drift followed historical patterns for Presidential re-election races. The amount of change between analysis dates between June 1 and Election Day was modeled using a bias parameter *b* applied across all polls, i.e. using margins of M_s+b instead of M_s. Therefore the win probability is the probability that *MM-b>0*.

b was assumed to follow a *t*-distribution setting the number of degrees of freedom equal to three, chosen to allow for the possibility of outlier events such as the 1980 Reagan-Carter race,

during which the standard deviation of the two-candidate margin was ~6% based on national polls (Erikson and Wlezien, 2012). The 2012 distribution of b was estimated using the Meta-Analysis in 2004, a re-election year in which the Meta-margin had a standard deviation (MMSD) of 2.2%. In historical data based on national polls, similar stability can be found in pooled trajectories of re-election races from multiple pollsters (see Figure 2.1 in Erikson and Wlezien, 2012). However, estimating MMSD from national data is difficult because of sampling error. For example, in 2004 Gallup national data showed a standard deviation of 4.9%, and in six re-election races from 1972 to 2004, Gallup data gave standard deviations between 2.9 and 4.9%.

3.4 Voter power

The power of a single voter in state s was calculated by calculating the incremental change in one candidate's election-win probability $\Delta P_s(EV \geq 270)$ arising from a change in M_s of a fraction of a percentage point, and normalized by the number of votes cast in the most recent Presidential election. ΔP_s for each state was normalized to voters in the most-powerful state or to one voter in New Jersey. The latter measure was termed a "jerseyvote."

3.5 Tracking national opinion swings

To track national opinion swings with high time resolution (**Figure 8**), all national polls within a time interval were divided equally into single-day components, then averaged for each day without weighting to generate a time course. After the election, the time course was adjusted by a constant amount to match the actual popular-vote result.

9

4. RESULTS

4.1 Kerry v. Bush 2004: an initial estimate of the bias variable

The first version of the Meta-Analysis, published starting in July 2004, analyzed the closely-fought re-election race of President George W. Bush (R) against his challenger, Senator John Kerry (D). Shortly after the Meta-Analysis was announced on DailyKos.com, it attracted thousands of readers almost immediately, and for good reason. The race was close and suspenseful, and the EV estimator crossed the 270 EV threshold three times during the general election campaign (**Figure 2a**). Meta-analysis was necessary to see this, since the swings were not large in terms of popular support: a 1-point change in the two-candidate margin across all states caused a change of 30 EV in the electoral margin. On Election Eve, the polls-only estimate (i.e. an estimate with bias parameter b=0%) turned out to be exactly correct: Bush 286 EV, Kerry 252 EV. Because the smallest single-state margin was 0.4% (Wisconsin), the uncorrected meta-analysis had an effective accuracy of less than 0.4% in units of popular opinion.

During the campaign, sharp or substantial moves in the EV estimator occurred after the Democratic convention (but not the Republican convention), the Swift Boat Veterans for Truth advertising campaign, and the first Presidential debate. Later debates had little effect, and from October 7th onward the race was static.

Despite the accuracy of the polls-only Meta-Analysis, I personally made an erroneous prediction. In the closing weeks of the campaign, I suggested that undecided voters would vote against the incumbent, a tendency that had been noticed in earlier campaigns. This led me to make an estimate of b=+1.5% toward John Kerry, which led to an incorrect prediction of Kerry 283 EV, Bush 255 EV. The incumbent rule, which was derived from an era in which recent pre-election polls were often not available, was rejected for future analyses. I also concluded that

10

interpreting polling data is susceptible to motivated reasoning and biased assimilation, cognitive biases that occur even among quantitatively sophisticated persons (Kahan et al., 2013). These reasons lead to a strong prescription to set b to zero for tracking purposes.

The bias variable b was still useful to readers who wanted to apply alternative scenarios. If a reader thought turnout efforts would boost his/her candidate by N points, that could be added as $b=N$ and the script recalculated. If he/she thought that one candidate would gain N points at the expense of the other, b could be set equal to $2N$. A map on the PEC website showed the effect of $b=\pm2\%$. For other scenarios, more sophisticated readers could run the MATLAB code.

4.2 Alternative scenarios and the jerseyvote index

As formulated in equation (1), alternative scenarios are explored easily. The most straightforward approach is to directly alter p_s by setting its value to 0 ("what if Romney wins Florida?") or 1 ("what if Obama wins Florida?"). Alternately, the polling margin M_s can be shifted for one or more states.

In 2004 this perturbation approach was introduced using the concept of the "jerseyvote," a fanciful way of expressing the concept of individual voter power. The jerseyvotes calculation was done by shifting all polls to create a near-tied race, adding an additional small change in M_s in a single state, and calculating the resulting change in the win probability. Conceptually, jerseyvotes are related to the Penrose-Banzhaf power index (Banzhaf, 1965). Jerseyvotes express an individual's relative power to influence the final electoral outcome. For example, if a voter in Colorado has ten times as much influence over the national win probability as a voter in New Jersey, the Coloradan's vote is worth 10 jerseyvotes. Sadly for the hosts of PEC, one jerseyvote is not worth very much. PEC advised New Jersey residents to vote early, then amplify their

efforts by tens of thousands by helping Pennsylvania voters get to the polls. In 2008 and 2012, readers were provided with a Voter Influence table (**Table 2**).

4.3 Accuracy in off-year elections, 2006 and 2010

The accuracy of state-level polls was confirmed a second way in the off-year elections of 2006. Using simple polling medians and a compound probability calculation, I estimated the probability of a Democratic takeover of the US Senate at approximately 50%, a higher chance than predicted by pundits or electronic markets. The Democrats (along with two independents) took control of the Senate with a 51-49 majority. A House prediction was note makde.

In the 2010 midterm off-year election, all Senate races were called correctly with the exception of the reelection race of Senator Harry Reid (D-NV) race against Sharron Angle, in which Reid trailed in the last eight pre-election polls, yet won by over five points. This polling error has been ascribed to undersampling of cell-phone-only and Hispanic voters.

In the 2010 House election, Republicans retook control with a 51-seat margin. PEC used district-by-district pre-election polls to predict a 25-seat Republican margin, a substantial underestimate. Most analysts performed similarly, suggesting that in an off-year, district-specific polls may not capture differences in voter intensity between the parties. Congressional generic preference polls on Election Eve showed an average 7-point advantage for Republicans, which would have led to a more accurate prediction.

4.4 Obama v. McCain 2008: Identifying a campaign's turning points

In 2008 the algorithm was kept the same, with the addition of automatic updates to track time trends graphically on a daily basis. This calculation used polling data for all 50 states and the District of Columbia, resulting in the electoral histories shown in **Figure 2**.

The EV estimator and the Meta-Margin showed Senator Barack Obama (D) ahead for almost the entire general election campaign, with an electoral lead of 20 to 200 electoral votes and 1 to 8 percentage points. At times, this lead shifted rapidly (**Figure 2b, Figure 8**). Senator McCain (R) immediately gained a large but transient benefit after the addition of Alaska Governor Sarah Palin as his running mate. Following her riveting convention speech, the Meta-Analysis moved from a large Obama lead to a near-tie. Considering the delays in getting fresh state-level data, it is possible that McCain led Obama at this time. The EV estimator reversed course shortly after Palin's unsuccessful interview with Charlie Gibson on ABC. After that, movement toward Obama accelerated after the collapse of Lehman Brothers, a defining event of that year's economic crash. Movement toward Obama continued after the first Presidential debate, and continued for the rest of October.

By Election Day, the EV estimator had stabilized at 353 EV, with a nominal 68% confidence band of [337,367] Obama EV, and a 95% confidence band of [316,378] EV. These confidence bands included pollster variation (house effects), and so the true uncertainty was likely to be substantially lower. Using a wider time window to minimize the variance in the time series gave a final estimate of 364 EV (**Table 1**), just one electoral vote away from the final outcome, Obama 365 EV, McCain 173 EV. The final Meta-margin, Obama +8.0%, was close to the final national polling median indicating Obama +7.5%. Obama's final margin in the national popular vote was +7.3%.

Downticket, polls showed comparable overall levels of accuracy (**Table 3**). In the Senate, the median outcome was 58-59 Democratic+Independent seats, with the Minnesota race (D-Franken v. R-Coleman) too close to call. The final outcome was 59 Democratic+Independent seats. In the House, taking all polls available at Pollster.com and assigning each winner to the leader, Democrats were predicted to win 257 ± 3 seats (68% confidence interval, 254-260 seats) assuming binomial random outcomes for close races. The final outcome was 257 Democratic seats.

4.5 Covariation between states adds modest uncertainty

State polls are partially interdependent samples because they are conducted by a smaller group of polling organizations. This raises the possibility that any systematic error would be shared by multiple states. One upper bound to the cumulative electoral effect of systematic error is the nominal 95% confidence band (gray bands in **Figure 2**). To test whether covariation was a likely contributor to the overall error, *b* was set to a range of Δ=[-1, +1]% or [-2, +2]% and the resulting EV probability distributions averaged over all values of Δ. The results for an August 2008 dataset are shown in **Figure 3**.

All three cases showed the same median (298 EV) and mode (305 EV). With no covariation, the 68% confidence interval was [280, 312] EV, a width of 32 EV. With ±1% covariation, the confidence interval widened by 3 EV to [279, 314]. With ±2% covariation, the interval widened by 12 EV to [275, 319] EV. Thus even perfect covariation leads only to modest changes in the overall shape of the outcomes distribution.

4.6 Obama v. Romney 2012

Re-election races are generally thought to be a referendum on the incumbent. President Obama came into the general election campaign with a united Democratic party and a number of accomplishments, including the rescue of the auto industry and the passage of the Affordable Care Act. However, the economy was still weak and the opposition party was polarized and combative. Most econometric models gave the President a slight to moderate advantage for re-election.

Viewed as a whole from June 1 through Election Day (**Figure 2c**), the electoral history fluctuated around an equilibrium of Obama 312 ± 16 EV (mean $\pm$ SD), and a Meta-Margin of $3.0 \pm 1.2\%$. The distributions were not long-tailed (kurtosis $=2.7$ for EV, 2.5 for the Meta-Margin, 3 for a normal distribution). Thus the race varied over about half the range of the 2004 election and was notably stable.

Four major events appeared to precede local maxima and minima in the electoral history: the addition of Rep. Paul Ryan (R) as Mitt Romney's running mate (helping Romney), the Democratic National Convention (helping Obama), the first Presidential debate (helping Romney), and the vice-presidential debate (helping Obama). In each case, the tight temporal association suggests that the event was the cause (**Figure 4**). Unlike a mixed polling/econometric approach, a polls-only approach is able to resolve notable campaign events to within a few days.

4.7 A prediction with no econometric assumptions

Starting in 2012, PEC began to provide a prediction. This was a true prediction, yet did not rely on econometric conditions. Prediction was done using the same tool used to calculate the Meta-Margin and the effects of covariation. The prediction was constructed on the assumption

that long-term movement in candidate preference moved uniformly in all states by an amount *b*, with *b* following a symmetric distribution with μ=MM and σ=2.2%. The parameter σ was estimated from movement of the Meta-margin in the 2004 and 2008 races. Since the actual σ was 1.2% in 2012, this parameter was in retrospect set conservatively.

An assumption of uniform shift is equivalent to assuming perfect covariation, which means that the the amount of predicted variation is an upper bound. The November prediction was plotted in the style of a hurricane strike zone, with the one-sigma band based on the parameter *b* (68% confidence interval) plotted in red, and a 95% confidence interval that included both long-term movement and pollster variations plotted in yellow (**Figure 2c**). This random-drift prediction approach gave an Obama win probability of 90% in July.

To determine how quickly the shift *b* developed, I calculated the average change in the Meta-margin for varying amounts of time from all dates in the 2008 general campaign season (**Figure 5**). This quantity increased with a half-rise time of 20 days. Its time course was similar to a square root function, consistent with a random walk. Therefore, for short-term predictions as the election drew near, I modeled movement in 2012 using σ=2.2*$\sqrt{(D/20)}$, where D was the number of days to the election. Under these assumptions, the Obama win probability increased to a maximum of 99.2% on Election Eve.

National polls could be added as a Bayesian prior to inform an estimate of the national popular vote (**Figure 6**). On the day before the election, the national poll median (Obama +0.0%) was assumed to predict the Meta-Margin as a *t*-distribution with σ=2.5%, a weak prior because of the substantial potential for systematic error. When combined with a state-polls-based prediction of Obama +2.9±1.5%, the predicted popular-vote margin was Obama +2.4%, with a

win probability >99.9%.The final two-party popular-vote margin was Obama +4.0%. Thus, state polls alone outperformed national polls in predicting the national popular vote.

4.8 Presidential coattails in the 2012 Senate race

Senate polls were analyzed using the same probabilistic algorithm as the EV tracker. Movement in this index was driven largely by seven close races: Missouri (D-i-McCaskill v. R-Akin), Indiana (D-Donnelly vs. R-Mourdock), Massachusetts (D-Warren v. R-i-Brown), Montana (D-i-Tester vs. R-Rehberg), North Dakota (D-Heitkamp v. R-Berg), Virginia (D-Kaine v. R-Allen), Wisconsin (D-Baldwin v. R-Thompson). PEC polling medians called the winner correctly in all seven races (**Table 3**).

Over time, the Senate seat-number tracking index (**Figure 7**) moved up and down in parallel with the Presidential race. From mid-September to Election Day, the probability of retained Democratic control stayed in the 80-99% range. A sharp dip in the Democratic/Independent seat count occurred in mid-August after the Ryan vice-presidential nomination, a steady and large increase occurred starting at the time of the Democratic convention, and a small decrease occurred after the first Presidential debate. Similar to the Presidential EV tracker, the Republican convention led to little change in the Senate seat count, and if anything, a slight movement toward Democrats.

These results indicate that Presidential and Senate preferences moved in tandem with one another, consistent with a coattail effect, i.e. similar party preference at different levels of the ticket. The first Presidential debate had a relatively weaker effect on Senate races than on the Presidential race, suggesting that the two levels are not always coupled equally.

17

5. DISCUSSION

In this work I have described a simple method for generating snapshots and predictions based on state polls alone. The Meta-analysis combines these polls to give a single snapshot with a temporal resolution of days, and accuracy equivalent to less than half a percentage point of difference in national support between the two candidates. Taken together, these qualities make the Meta-analysis a sensitive indicator of the ups and downs of a national campaign – in short, a precise electoral thermometer.

A post-election analysis (Muehlhauser and Branwen, 2012) has reviewed PEC's polls-only peformance and found it to be significantly superior to other aggregators and the betting site InTrade (**Table 3**). This is made possible by the fact that pollsters show a wisdom of crowds in which their net bias is near zero. Enough state polls have been available to track presidential races since 2000, when Ryan Lizza at *The New Republic* compiled state polls. On the day before the election, that compilation indicated that the outcome would hinge on Florida, as ultimately occurred. In 2004-2012, state poll Meta-Margin has come within an average of 1.6% of the national popular vote, making no sign errors (**Table 1**). National margins in 2000-2012 have done worse, getting the sign of the popular-vote margin correct in only two years (2004 and 2008) and deviating by an average of 2.1% from outcomes.

House-effect corrections of individual pollsters, as done by aggregators such as FiveThirtyEight, appear to be unnecessary for making an accurate prediction. To date, such corrections have not yielded much benefit in electoral-vote estimation (**Table 3**). Corrections for "house effects" are, however, useful for statistical error analysis. In 2004, 2008, and 2012, the nominal confidence interval of the EV estimator was wider than the event-related swings in each

18

race. Accurate estimation of confidence intervals would require removing the contribution of house effects in individual polls before they are entered into the EV estimator.

The results here demonstrate that a model composed of uncorrected polls and random drift over time is enough to make an accurate prediction. The question arises, then: what contribution can econometric data make to true prediction? It has been demonstrated (Abramowitz, 2008; Linzer, 2013) that econometric variables have predictive value before a general election campaign, when polls are scarce. After the season begins, opinion polls provide a direct measurement of opinion, at which point the question becomes one of estimating how opinion will evolve over time. A true prediction properly done should not change much over time, as seen in the work of Linzer (2013). A snapshot to track the current state of the race does the converse. Adding random drift to the snapshot lacks an explanatory component, but it has the advantage of generating a reliable forecast.

One way to incorporate econometric modeling while retaining the news power of the snapshot is to estimate the direction of drift going forward in time from the snapshot. For example, it would be possible to quantify how 2nd-quarter unemployment and July-to-November poll *movement* were related, and with what distribution. In this manner, polling data at any moment in time could be used as a starting point for future projections.

Although national polls are inferior at presidential race prediction, they have the advantage of high time resolution due to their frequency. In contrast, the state-poll snapshot takes at least one week to equilibrate after a major campaign event. In the future, it should be possible to use national polling data to estimate day-to-day shifts in opinion (**Figure 8**) and apply this as a correction to the EV estimator using the bias variable *b*, thereby achieving both accuracy and temporal sensitivity.

6. CONCLUSION

What is the future of poll aggregation? In addition to news value, poll aggregation has other applications. One is election integrity. In cases where substantial pre-election polling is available, fraud is made more difficult by the presence of concrete opinion data. A second application is resource allocation (Strömberg, 2002), both by candidate campaigns and by activist organizations. A third potential application is a reduction in media chatter concerning individual polls.

An open question for the future is whether poll aggregation will continue to perform well in the future. The answer depends in large part on the availability of accurate polling data. Economic tension exists between polling organizations which release individual data points as a means of calling attention to themselves, news organizations for which a poll is cheaper to run than a reporter is to pay for generating a story, and aggregators who obtain a far more accurate result by collecting many polls. Although one possible outcome is that fewer polls will be available, even if they were halved in number, the meta-analysis would be minimally affected. Conversely, journalism might benefit from the weeding-out of low-information news stores about single polls. Ideally, this would clear the way for more substantive coverage of political races.

REFERENCES

Abramowitz, A.I. (2008) Forecasting the 2008 presidential election with the time-for-change model. *PS-Political Science & Politics*. 41, 691-695.

Banzhaf, J.F. (1965) Weighted voting doesn't work: A mathematical analysis. *Rutgers Law Review* 19, 317-343.

Campbell, J.E. (2004) Forecasting the presidential vote in 2004: placing preference polls in context. *PS: Political Science and Politics*, 37, 733-735.

Erikson, R.S. & Wlezien, C. (2012) The timeline of Presidential elections: how campaigns do (and do not) matter. University of Chicago Press, Chicago.

Forelle, C. (2004a) For math whizzes, the election means a quadrillion options. *Wall Street Journal*, October 26, 2004, page A1.

Forelle, C. (2004b) Winner at picking electoral vote. *Wall Street Journal*, November 4, 2004, page D9.

Gelman, A. and King, G. (1993) Why are American Presidential-election campaign polls so variable when votes are so predictable? *British Journal Of Political Science*, 23, 409-451.

Gelman, A., Silver, N., & Edlin, A. (2010) What is the probability your vote will make a difference? *Economic Inquiry*, 50, 321-326.

Jackman, S. and Blumenthal, M. (2013). Using model-based poll averaging to evaluate the 2012 polls and pollsters. In *AAPOR 68th Annual Conference*.

Jones, R.E. Jr. (2008) The state of presidential election forecasting: the 2004 experience. *International Journal of Forecasting*, 24, 310-321.

Kahan, D.M., Peters, E., Dawson, E.C, and Slovic, P. (2013) Motivated numeracy and enlightened self-government. http://dx.doi.org/10.2139/ssrn.2319992.

Lewis-Beck, M.S. and Tien, C. (2008) Forecasting presidential elections: when to change the model. *International Journal of Forecasting*, 24, 227-236.

Linzer, D.A. (2013) Dynamic Bayesian forecasting of Presidential elections in the states. *Journal of the American Statistical Association*, 108, 124-134.

Muehlhauser, L. and Branwen, G. (2012) Was Nate Silver the most accurate 2012 election pundit? *Center for Applied Rationality*, November 9, 2012. http://rationality.org/2012/11/09/was-nate-silver-the-most-accurate-2012-election-pundit/

Noonan, P. (2012) Monday morning, *Wall Street Journal* online, http://blogs.wsj.com/peggynoonan/2012/11/05/monday-morning/, November 5, 2012.

Poor, J. (2012) George Will predicts 321-217 Romney landslide. *Daily Caller*, Novvember 4, 2012, http://dailycaller.com/2012/11/04/george-will-predicts-321-217-romney-landslide/

Soumbatiants, S.R. (2003) Forecasting the probability of winning the U.S. presidential election. Doctoral thesis, University of South Carolina.

Soumbatiants, S., Chappell, H., & Johnson, E. (2006) Using state polls to forecast U.S. presidential election outcomes. *Public Choice*, 123, 207-223.

Tracy, M. (2012) Nate Silver is a one-man traffic machine for the Times. *The New Republic*, November 6, 2012.

Table 1. Comparison of polling meta-analysis with election outcomes, 2004-2012. The win probability was calculated assuming symmetric drift (t-distribution, 3 degrees of freedom) with σ=2.2% between July 1 and Election Day. The Meta-margin standard deviation was calculated from June 1 to Election Day. National polls were calculated as the median of all polls conducted from November 1 to Election Day.

| Year | PEC forecast / snapshot | | | National polls | Outcome | |
	July 1 win Democratic win probability	November 1 Democratic EV estimate	November 1 Meta-margin MM (SD)	poll median	Democratic EV	Popular vote outcome (two-party)
2000				Bush +2.5%	266 EV	Gore +0.5%
2004	38%	252 EV	Bush +0.7% (1.2%)	Bush +2.0%	252 EV	Bush +3.0%
2008	90%	364 EV	Obama +8.0% (2.2%)	Obama+7.5%	365 EV	Obama+7.3%
2012	90%	315 EV	Obama +2.6% (1.2%)	Tie (+0.0%)	332 EV	Obama+4.0%

Table 2. The power of an individual voter. As an example calculation, a listing of voter power as calculated on Election Eve, November 5, 2012.

State	Median polling margin	Power
NH	Obama +2%	100.0
IA	Obama +2%	82.2
PA	Obama +3%	77.8
OH	Obama +3%	74.0
NV	Obama +5%	71.9
VA	Obama +2%	71.0
CO	Obama +2%	63.7
WI	Obama +4.5%	44.7
NM	Obama +6%	30.1
FL	Tied	26.6
MI	Obama +5.5%	21.6
OR	Obama +6%	19.1
NC	Romney +2%	5.2
MN	Obama +7.5%	3.2
LA	Romney +13%	0.9
NJ	Obama +12%	0.0009149

Table 3. Performance comparisons in 2008 and 2012. Presidential predictions and results are listed for Barack Obama. *Brier scores come from Table 5.2 of Muehlhauser and Branwen (2012), and are defined so that lower numbers indicate better performance. The 2012 Senate close races are listed in section 4.8.

	FiveThirtyEight	Linzer (Votamatic)	InTrade	Polls alone (PEC)	Actual outcome
2008					
Presidential EV	348.5 EV	-	364 EV	353/364 EV	365 EV
Popular vote	52.3%	-	-	53.0%	52.9%
Senate	58-59 D	-	-	58-59 D	59 D
House	-	-	-	257 D	257 D
2012					
Presidential EV	313 EV	332 EV	303 EV	312 EV	332 EV
*Brier score, Pres.win	0.0083	0.0001	0.1170	0.0000	0.0000
*Brier score, state win	0.009	0.004	0.028	0.008	0.000
Senate close races	5/7	-	5/7	7/7	7/7
*Brier score (30 races)	0.045	-	0.049	0.012	0.000
*Brier score, combined Presidential/Senate	**0.023**	**-**	**0.037**	**0.009**	**0.000**

FIGURE LEGENDS

Figure 1. Foundations of the Presidential meta-analysis. (a) State-by-state election margins as a function of final pre-election polls in the 2004 Kerry v. Bush race. (b) Pre-election win probabilities and actual outcomes in the 2012 Obama v. Romney race. (c) A snapshot of the exact distribution of all 2^{51}=2.3 quadrillion outcomes calculated from win probabilities in (b). The electoral vote estimator is defined as the median of the distribution. (d) Electoral Effect of uniform shift in state polls by a constant swing. The gray band indicates nominal 95% confidence interval including uncorrected pollster-to-pollster variation.

Figure 2. Time series of the meta-analytic electoral vote predictor, 2004-2012. The EV estimator for the most recent available state polls plotted as a function of time for (a) 2004, (b) 2008, and (c) 2012. The arrows indicate notable campaign events. Upward-pointing arrows indicate events likely to benefit the Democratic candidate, downwarw-pointing arrows the Republican candidate. DNC, Democratic National Convention. RNC, Republican National Convention. SBVT, Swift Boat Veterans for Truth ad campaign. HRC, Hillary Rodham Clinton. The gray band indicates nominal 95% confidence interval including uncorrected pollster-to-pollster variation.

Figure 3. Effects of covariation among state polls. The effect on (a) the uncorrected snapshot electoral vote estimator of adding a bias of (b) -1 to +1% or (c) -2 to +2% to state polls. The center of the distribution does not change but its width increases modestly.

Figure 4. Turning-point events in Presidential campaigns. An expanded view of significant campaign-moving events in 2008 and 2012, followed by subsequent events reported to have worked in the opposite direction. (a) Sarah Palin (R) vice-presidential nomination accceptance speech at the Republican convention, followed by her interview with Charlie Gibson on ABC, John McCain (R) appearance on The View, and the Lehman Brothers collapse. (b) The announcement of Paul Ryan (R) addition as vice-presidential nominee, followed by Rep. Todd Akin (R) comment on "legitimate rape." (c) The first Obama-Romney presidential debate in 2012, followed by the Biden-Ryan vice-presidential debate and the second Presidential debate.

Figure 5. A random-drift Bayesian prediction model for Presidential campaigns. (a) Average change in Meta-margin over the 2012 campaign season. (b) Application of drift in (a) to make a prediction. The red zone indicates the one-sigma range. The yellow indicates the union of the two-sigma range and the 95% nominal confidence interval.

Figure 6. Using state and national polls to predict the popular vote. National polls and the state-poll-based meta-analysis are combined to make a prediction of the national populat vote. The state-polls-only estimate performed better than the combined estimate.

Figure 7. Coattail effects in the U.S. Senate elections, 2012. Polling snapshot of Senate outcomes as a function of time, based the most recent available Senate data.

Figure 8. Increased time resolution from day-by-day averaging of national polls. National polling margins. Each available poll at Huffington Post/Pollster.com was distributed over the dates it was conducted and the average calculated. The time series was shifted so that the last day matched the actual popular vote outcome on Election Day. "Sandy" indicates Hurricane Sandy.