Bayesian Combination of State Polls and Election Forecasts*

Kari Lock[1] and Andrew Gelman[2]
[1] *Department of Statistics, Harvard University, lock@stat.harvard.edu*
[2] *Department of Statistics and Department of Political Science, Columbia University, gelman@stat.columbia.edu*

31 January 2010

Abstract

A wide range of potentially useful data are available for election forecasting: the results of previous elections, a multitude of pre-election polls, and predictors such as measures of national and statewide economic performance. How accurate are different forecasts? We estimate predictive uncertainty via analysis of data collected from past elections (actual outcomes, pre-election polls, and model estimates). With these estimated uncertainties, we use Bayesian inference to integrate the various sources of data to form posterior distributions for the state and national two-party Democratic vote shares for the 2008 election. Our key idea is to separately forecast the national popular vote shares and the relative positions of the states. More generally, such an approach could be applied to study changes in public opinion and other phenomena with wide national swings and fairly stable spatial distributions relative to the national average.

Keywords: Bayesian updating, election prediction, pre-election polls, shrinkage estimation

1 Introduction

Research tells us that national elections are predictable from fundamentals (e.g., Rosenstone, 1983, Campbell, 1992, Gelman and King, 1993, Erikson and Wlezien, 2008, Hibbs, 2008), but this doesn't stop political scientists, let alone journalists, from obsessively tracking swings in the polls. The next level of sophistication—afforded us by the combination of ubiquitous telephone polling and internet dissemination of results—is to track the trends in state polls, a practice which was led in 2004 by Republican-leaning realclearpolitics.com and in 2008 at the websites election.princeton.edu (maintained by biology professor Sam Wang)

*We thank Aaron Strauss and three anonymous reviewers for helpful comments and the National Science Foundation, Yahoo Research, the Institute of Education Sciences, and the Columbia University Applied Statistics Center for partial support of this work.

and fivethirtyeight.com (maintained by Democrat, and professional baseball statistician, Nate Silver).

Presidential elections are decided in swing states, and so it makes sense to look at state polls. On the other hand, the relative positions of the states are highly predictable from previous elections. So what is to be done? Is there a point of balance between the frenzy of daily or weekly polling on one hand, and the supine acceptance of forecasts on the other? The answer is yes, a Bayesian analysis can do partial pooling between these extremes. We use historical election results by state and campaign-season polls from 2000 and 2004 to estimate the appropriate weighting to use when combining surveys and forecasts in the 2008 campaign.

The year leading up to a presidential election is full of polls and speculation, necessitating a study of the measure of uncertainty surrounding predictions. Given the true proportion who intend to vote for a candidate, one can easily compute the variance in poll results based on the size of the sample. However, here we wish to compute the forecast uncertainty given the poll results of each state at some point before the election. To do this, we need not only the variance of a sample proportion, but an estimate for how much the true proportion varies in the months before the election, and a prior distribution for state-level voting patterns. We base our prior distribution on the 2004 election results and use these to improve our estimates and to serve as a measure of comparison for the predictive strength of pre-election polls.

We use as an example the polls conducted in February, 2008, by SurveyUSA, which sampled nearly 600 voters in each state, asking the questions, "If there were an election for President of the United States today, and the only two names on the ballot were Republican John McCain and Democrat Hillary Clinton, who would you vote for?" and "What if it was John McCain against Democrat Barack Obama?" The poll was conducted over the phone using the voice of a professional announcer, with households randomly selected using random digit dialing (Survey Sampling International, 2008). Each response was classified as one of the two candidates or undecided. For each state the undecided category consisted of 5–14% of those polled, and these people as well as third-party supporters were excluded from our analysis. Likewise, for previous election results, we restrict the population to those who supported either the Democrat or the Republican.

This paper merges prior data (the 2004 election results) and the poll data described above to give posterior distributions for the position of each state relative to the national popular vote. For the national popular vote we use a prior determined by Douglas Hibbs's

"bread and peace model" (Hibbs, 2008), and again merge with our SurveyUSA poll data.

In sections 2 and 3 of this article we ascertain the strength of each source of data in predicting the election. Section 2 contains an analysis of the use of past election results in predicting future election results, ultimately resulting in an estimate for the variance of the 2008 relative state positions given the 2004 election results. Section 3 contains an analysis of the strength of pre-election polls in predicting election results, giving measures both of poll variability and variability due to time before the election. Section 4 brings the sources together with a full Bayesian analysis, fusing prior data with poll data to create posterior distributions.

More generally, an approach such as described here could be applied to study changes in public opinion and other phenomena with wide national swings and fairly stable spatial distributions relative to the national average. For example, Lax and Phillips (2009) compare state-level policies and attitudes on several gay rights questions in the period from 1994 through 2006. The relative rankings of the states on gay rights were fairly stable during a period of great change nationally. In trying to estimate current attitudes within states (or, more generally, within subsets of the population), it makes sense to decompose national and local variation. We illustrate in the present article with forecasts of the 2008 election.

2 Past Election Results

The political positions of the states are consistent in the short term from year to year; for example, New York has strongly favored the Democrats in recent decades, Utah has been consistently Republican, and Ohio has been in the middle. We begin our analysis by quantifying the ability to predict a state outcome in a future election using the results of past elections. We do this using the presidential elections of 1976–2004. We chose not to go back beyond 1976 since state results correlate strongly ($.79 \leq r \leq .95$) for adjacent elections after 1972, while the correlation between the 1972 and 1976 elections is only .11.

Figure 1 shows strong correlations in the Democratic share of the vote in each state from one presidential election to the next. But in many cases the proportion for the Democrat is uniformly higher or lower than would have been predicted by the previous election. For example, states had much higher proportions for Clinton in 1992 than for Dukakis in 1988, and much lower proportions for Gore in 2000 than for Clinton in 1996. This does not indicate a change in states' relative partisanship but rather a varying nationwide popularity of the Democratic candidate from election to election. Obama's vote share in a state may differ from Kerry's, but the vote for Kerry in any given state compared to the nationwide vote

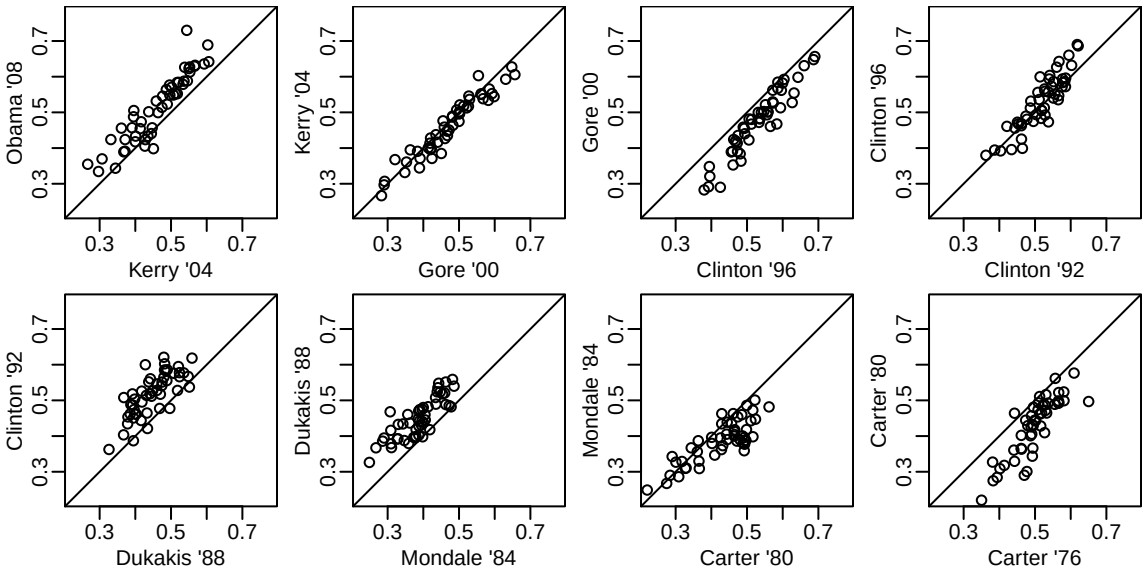

Figure 1: State results from one presidential election to the next, in each case showing the Democratic candidates' share of the two-party vote in each state. The 2008 results are shown here, but this information was not used or available at the time of analysis.

seems to be indicative of Obama's vote in that state as compared to nationwide. For this reason we look at the *relative state positions*, the difference between the proportion voting Democratic in each state and the national proportion voting Democratic.

We tried various models using past elections to predict future elections, but found that not much was gained by using data from elections prior to the most recent election. We imagine that with careful adjustment for economic and political trends, there is useful information from earlier presidential races (as well as data from other elections), but in this paper we keep things simple: In our analysis of 2008 we ignore election data before 2004, and simply consider the proportion of voters in each state choosing John Kerry over George W. Bush in the 2004 election.

After centering around the national vote (Kerry's share of the two-party vote was 48.8% so our prior data become, for each state, the proportion voting for Kerry minus .488), our only adjustment is a home-state correction. We attribute 6% (as determined via analysis of past elections; see Campbell, 1992, and Gelman and King, 1993) of the vote for Bush and Kerry in Texas and Massachusetts, respectively, to a home-state advantage, and we add that same amount in the forecast for McCain in Arizona and Clinton in New York or Obama in Illinois. Further improvement should be possible with careful modeling (or the

sort of careful empricism that political professionals do), but it would not alter our basic point that national and statewide swings can be modeled separately.

To determine the strength of our prior data, we need to know how much these state relative positions vary from election to election. For this, we need data from several elections. Let $d_{s,y}$ be the relative position for state s in year y. We first estimate $\text{var}(d_{s,2008}|d_{s,2004})$ for each state by $\frac{1}{7}\sum_{i=1}^{7}(d_{s,y_{i+1}} - d_{s,y_i})^2$, where $\vec{y} = (1976, \ldots, 2004)$. With only seven data points for each state, however, these estimates could be unreliable. We could get around this problem by assuming a common variance estimate for all states, but rather than forcing either one common estimate or fifty individual estimates, we use shrinkage estimation (also called partial pooling). Exactly how much to pull each estimate to the common mean is determined via a hierarchical model which we fit in R using lmer (Bates, 2005) and is ultimately based upon comparisons of within-state and between-state variability. Before pooling, the estimates of standard deviation for each state range from .012 to .073, with complete pooling the common estimate is .037, and after our partial pooling the estimates range from .029 to .055.

From the normal approximation, we can expect the difference in 2008 to fall within .06 of the 2004 state difference for the most consistent states and up to .11 away for the least consistent states.

3 Pre-Election Polls

How much can we learn from a February poll of 600 voters in each state? If we ignore that the poll was conducted so early in the year, it appears we can learn quite a lot. Due to sampling variability alone, we would expect the true proportion who would vote Democratic in each state to be within .04 of the sample proportion ($\text{sd} = \sqrt{p(1-p)/n} \approx \sqrt{.5 \cdot .5/600} = .02$). A standard deviation of .02 would make a poll of this size more informative than the 2004 election. Using Monte Carlo techniques, one could simulate many potential "true" proportions for each state, and so many potential popular or electoral college results, as done in Erikson and Sigman (2008). However, this would depict voter preferences *in February*. To get a true measure of variability, we need to consider not only sampling variability and other survey issues, but also uncertainty about opinion shifts between then and Election Day (Strauss, 2007).

We estimate the national-level variance in vote intention during the months before the election using the results of Gallup polls in the presidential election years from 1952 through 2004. The sample size for the Gallup polls averaged 1500 each. Let p_t denote

the true national proportion who intended to vote for the Democratic candidate, t months before the election, $\hat{p}_t$ denote our estimate of p_t from a pre-election poll, and p_0 denote the two-party Democratic vote share in the actual election. Ideally we'd like $\text{var}(\hat{p}_t|p_0)$ as a function of both the poll sample size, n, and the number of months before the election the poll was conducted, t. Decomposing the variance conditionally yields,

$$\begin{aligned}
\text{var}(\hat{p}_t|p_0) &= \text{E}(\text{var}(\hat{p}_t|p_t)|p_0) + \text{var}(\text{E}(\hat{p}_t|p_t)|p_0) \\
&= \text{E}\left(\frac{p_t(1-p_t)}{n}|p_0\right) + \text{var}\left(p_t|p_0\right) \\
&= \frac{\text{E}(p_t|p_0) - \text{E}(p_t^2|p_0)}{n} + \text{var}\left(p_t|p_0\right) \\
&= \frac{p_0(1-p_0)}{n} + \left(\frac{n-1}{n}\right)\text{var}\left(p_t|p_0\right) \\
&\approx \frac{p_0(1-p_0)}{n} + \text{var}\left(p_t|p_0\right). \qquad (1)
\end{aligned}$$

The second term in this expression, $\text{var}\left(p_t|p_0\right)$, represents uncertainty in the underlying true proportion who would vote Democratic t months before the election, and it is not affected by the quality or quantity of polls conducted.

From equation (1), $\text{var}(p_t|p_0) = \text{var}(\hat{p}_t|p_0) - p_0(1-p_0)/n$, and so it can be estimated by empirically calculating $\text{var}(\hat{p}_t|p_0)$ and subtracting off the expected sampling variability.[1] Let $\hat{p}_{t,i}$ and $n_{t,i}$ denote estimated proportion and sample size respectively for the i^{th} poll in a given month, and let N_t be the number of polls we have t months before the election (from Gallup polls 1952 to 2004). We then estimate $\text{var}(p_t|p_0)$ by

$$\widehat{\text{var}}(p_t|p_0) = \frac{\sum_{i=1}^{N_t}\left[(\hat{p}_{t,i} - p_0)^2 - \frac{p_0(1-p_0)}{n_{t,i}}\right]}{N_t}. \qquad (2)$$

The variances estimated in this fashion for each month are displayed in Figure 2(a), along with a line fitted by weighted least squares. (Standard errors are displayed for each point, with larger standard errors in months with less historical polling data available.) The linear trend appears to fit reasonably well, the individual variance estimates are noisy enough that it would be difficult to fit a more elaborate curve. We set the intercept to be 0, assuming the popular vote in November should match that of the election and ignoring issues such as voter turnout.[2] This model gives $\widehat{\text{var}}(p_t|p_0) = .0008t$, with a standard error of .00013

[1] The $p(1-p)/n$ variance estimate is in practice an underestimate of survey error, given clustering, weighting, and other issues that depart from simple random sampling. A more elaborate analysis—using individual respondent data instead of just state averages—could account for these complexities using poststratification.

[2] When we remove the zero-intercept constraint, the estimated intercepts were low and not statistically significantly different from zero.

on the slope, suggesting that the variance in the underlying popular vote increases by .0008 each additional month before the election. Extrapolating to February yields $\widehat{\text{sd}}(p_{\text{feb}}|p_0) = .086$, which is enough higher than forecast uncertainties to imply that February polls contain almost no information about the candidates' *national* vote shares on Election Day.

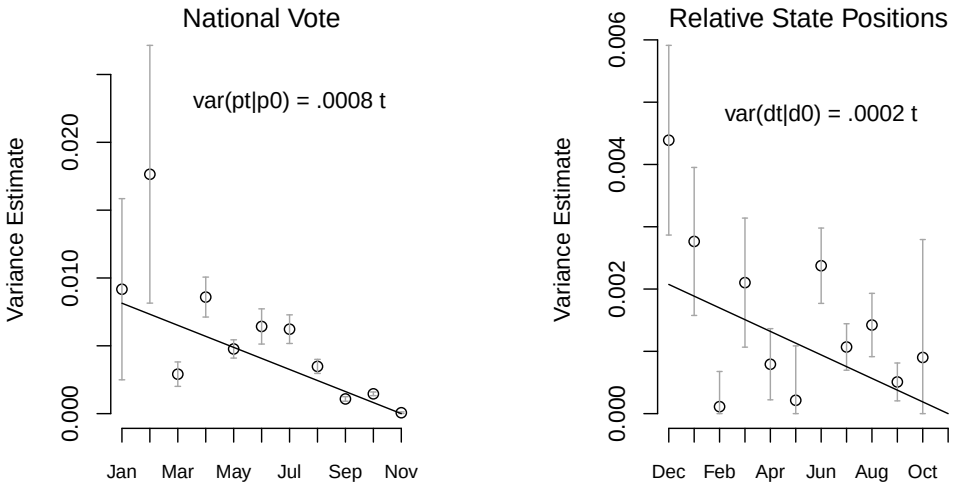

Figure 2: (a) Estimated variances of the popular vote in each month given the popular vote in the election. (b) Estimated variances of the relative position of each state in each month, given the relative position of the state in the election. Error bars represent ± one standard error.

We now repeat the above calculations, this time to estimate the variance of the relative positions of the states during the months before the election. We do this using the National Annenberg Election Survey, a large rolling cross-section poll conducted in 2000 and 2004 by the Annenberg Public Policy Center at the University of Pennsylvania. Again restricting our analysis only to those who say they would vote for the Democrat or the Republican, we have 43,373 people polled in 2000 and 52,825 in 2004.

Now we want $\text{var}(\hat{d}_{s,t}|d_0)$ as a function of n and t, where $d_{s,t}$ is the relative position, t months before the election, of state s. We follow the same logic as with the popular vote, except now instead of averaging over multiple years worth of pre-election polling data, with only two years to work with we have to average over the states. We average over all states, assuming a common variance across states. We tried computing seperate estimates for small and large states, or for Democratic, Republican, and battleground states, but the differences in estimated variances between these different sorts of states were small and not statistically significant. Due to the sample sizes in many states we chose a common estimate rather than noisier alternatives. For each state in each month, sample sizes range from 0

to 844, but with 42% having less than 30 people polled. Sample sizes this small lead to unreliable estimates, so we tweak (2) slightly and take a weighted average, weighting by sample size. We thus estimate $\text{var}(d_{s,t}|d_0)$ by

$$\widehat{\text{var}}(d_{s,t}|p_{s,0}) = \frac{\sum_{y\in\{2000,2004\}}\sum_{s=1}^{50} n_{s,y,t}\left[(\hat{d}_{s,y,t} - d_{s,y,0})^2 - \frac{p_{s,y,0}(1-p_{s,y,0})}{n_{s,y,t}}\right]}{\sum_{y\in\{2000,2004\}}\sum_{s=1}^{50} n_{s,y,t}}. \tag{3}$$

This isn't quite as straightforward as the calculation for (2), since we don't observe the national opinion at time t so can't actually observe $\hat{d}_{s,t}$ (we only have $\hat{p}_{s,t}$). To get around this, we estimate the national popular vote each month before the elections of 2000 and 2004 using both the Annenberg state polls and Gallup poll data. In practice, the abundance of large national polls should give a good estimate of the national opinion at any point in time. We use these estimates to calculate each $\hat{d}_{s,t}$, which then allows us to compute (3) for each month. The estimated variances are shown in Figure 2(b). A weighted linear regression on these data points, again with intercept 0, gives the equation $\widehat{\text{var}}(d_{s,t}|d_{s,0}) = .0002t$, with a slope standard error of .00005. This estimates $\widehat{\text{sd}}(d_{s,feb}|d_{s,0}) = .041$, about half the standard deviation of the national mean.

4 Posterior Distributions

With the variance estimates derived in sections 2 and 3, we are all set to go forth with the full Bayesian analysis. We first look only at the relative positions of the states, and momentarily ignore the national popular vote. Our poll and prior distributions can be represented as follows:

$$\text{Poll:}\quad \hat{d}_{s,t}|d_{s,0} \sim N\left(d_{s,0},\ \frac{p_{s,0}(1-p_{s,0})}{n_{s,t}} + \text{var}(d_{s,t}|d_{s,0})\right) \tag{4}$$

$$\text{Prior:}\quad d_{s,0}|d_{s,2004} \sim N\left(d_{s,2004}, \text{var}(d_{s,0}|d_{s,2004})\right). \tag{5}$$

Here $d_{s,0}$ is equivalent to the notation $d_{s,2008}$ used in section 2; both refer to the relative position of state s at the time of the 2008 election.

Model (4) gives the distribution of a state poll conducted t months before the election (relative to the national opinion at that time), given that state's ultimate relative position at the time of the election. The poll variance has a component based on the poll sample size and a component based on time before the election. In section 3 we estimated the variance due to time before the election to be $\text{var}(d_{s,t}|d_{s,0}) \approx .0002t$. This estimate was calculated using the Annenberg state polls from the 2000 and 2004 elections. Normality is justified by the large sample size of each poll.

The prior gives a distribution for the state relative positions in the 2008 election given each state's relative position in the 2004 election. The prior variance, $\text{var}(d_{s,0}|d_{s,2004})$, is estimated in section 2 using the results of past elections. Estimated variances range from $.029^2$ to $.056^2$, differing by state. Normality is justified by the general lack of outliers in state election returns (an assumption that didn't quite work in Hawaii in 2008).

Combining these distributions will provide our quantity of interest, a posterior distribution for the true state relative positions at the time of the election, given poll data and the 2004 election results. With the normal-normal mixture model, we weight by information, the reciprocal of variance. Our posterior takes the form:

$$d_{s,0}|\hat{d}_{s,t}, d_{s,2004} \sim N\left(\frac{\frac{1}{\text{var}(\hat{d}_{s,t}|d_{s,0})}\hat{d}_{s,r} + \frac{1}{\text{var}(d_{s,0}|d_{s,2004})}d_{s,2004}}{\frac{1}{\text{var}(\hat{d}_{s,t}|d_{s,0})} + \frac{1}{\text{var}(d_{s,0}|d_{s,2004})}}, \quad \frac{1}{\frac{1}{\text{var}(\hat{d}_{s,t}|d_{s,0})} + \frac{1}{\text{var}(d_{s,0}|d_{s,2004})}}\right). \quad (6)$$

We illustrate with the February Survey USA state polls described in section 1. We first calculate $\hat{d}_{s,feb}$ for each state. We don't know the popular vote in February so can't compute these exactly, but can get a pretty close estimate given that we have a sample size exceeding 500 in each state. In section 3, we estimated $\text{var}(d_{s,feb}|d_{s,0}) \approx .041^2$, so from (4) we get the poll distribution

$$\hat{d}_{s,feb}|d_{s,0} \sim \text{N}\left(d_{s,0}, \frac{p_{s,0}(1-p_{s,0})}{n_{s,feb}} + .041^2\right). \quad (7)$$

The sample sizes range from 500 to 600, leading to standard deviations ranging from .045 to .047. (Our model assumes the state poll gives an unbiased estimate of the true opinion at that date. The analysis becomes more difficult if, for example, pollsters are performing their own Bayesian adjustments and shrinking down outliers before reporting their survey numbers.)

For most states, the poll standard deviation (.045 to .047) is higher than the prior standard deviation (.029 to .056). This means that most posteriors will place more weight on the estimates based on the 2004 election results than on the Feburary poll estimates. For a typical state, (6) simplifies to something like

$$d_{s,0}|\hat{d}_{s,feb}, d_{s,2004} \sim \text{N}(.4\hat{d}_{s,feb} + .6d_{s,2004}, .03^2), \quad (8)$$

with the weight on the poll estimate ranging from .29 to .59 and the standard deviations ranging from .025 to .036. States with higher prior variances place more weight on the polls, and have higher posterior standard deviations. Figure 3 shows the posterior predictions for the relative positions of the states for both Clinton and Obama. (The poll was conducted

before the Democratic candidate was chosen, and our prior applies to any Democratic candidate.) In retrospect (and, perhaps, even before the election) the estimates are not perfect—for example, should Texas really have been viewed as close to a toss-up state for Obama?—and such discrepancies should motivate model improvement. (From a Bayesian perspective, if you produce an estimate using correct procedures but it still looks "wrong," that means you have additional information that has not yet been included in the model as prior or data.)

We now move on to creating a posterior for the national popular vote. We construct our prior based on the estimate and predictive standard deviation from Hibbs (2008), who predicts the national two-party Democratic vote share based only on two factors: weighted-average growth of per capita real personal disposable income over the previous term (with the weighting estimated based on past election results), and cumulative US military fatalities owing to unprovoked hostile deployments of American armed forces in foreign conflicts. To determine the variance in the success of this model we look at its predictions for the previous 14 elections (1952 to 2004). The sample standard deviation of (predicted − actual) is .021 (quite accurate for only two predictors and no polling information!). Shortly before the election, Hibbs predicted that Obama would get 53.75% of the two-party vote.

Thus for the national popular vote, we have

$$\text{Poll:} \quad \hat{p}_t|p_0 \sim \text{N}\left(p_0,\ \frac{p_0(1-p_0)}{n_t} + \text{var}(p_t|p_0)\right) \tag{9}$$

$$\text{Prior:} \quad p_0 \sim \text{N}\left(.5375,\ .021^2\right) \tag{10}$$

$$\text{Posterior:} \quad p_0|\hat{p}_t \sim \text{N}\left(\frac{\frac{1}{\text{var}(\hat{p}_t|p_0)}\hat{p}_t + \frac{1}{.021^2}.5375}{\frac{1}{\text{var}(\hat{p}_t|p_0)} + \frac{1}{.021^2}},\ \ \frac{1}{\frac{1}{\text{var}(\hat{p}_t|p_0)} + \frac{1}{.021^2}}\right). \tag{11}$$

With our February poll data, we get the estimated popular vote by weighting the sample poll proportion voting Democratic in each state by the number of voters in that state in the 2004 election. This gave a national estimate of 51.44% for Obama. From section 3, $\text{var}(\hat{p}_{\text{feb}}|p_0) = p_0(1-p_0)/n + .086^2 \approx (.51 \cdot .49)/27000 + .086^2$, giving a standard deviation of 8.6 percentage points. This variance may not be entirely accurate since the variance was estimated in section 3 using polls of a nationwide sample rather than a sample within each state, but we didn't have sufficient state level data from enough past elections to provide a better estimate. This estimate (.086) is much larger than the standard deviation associated with our prior (.021), so the posterior will be strongly weighted towards Hibbs's estimate.

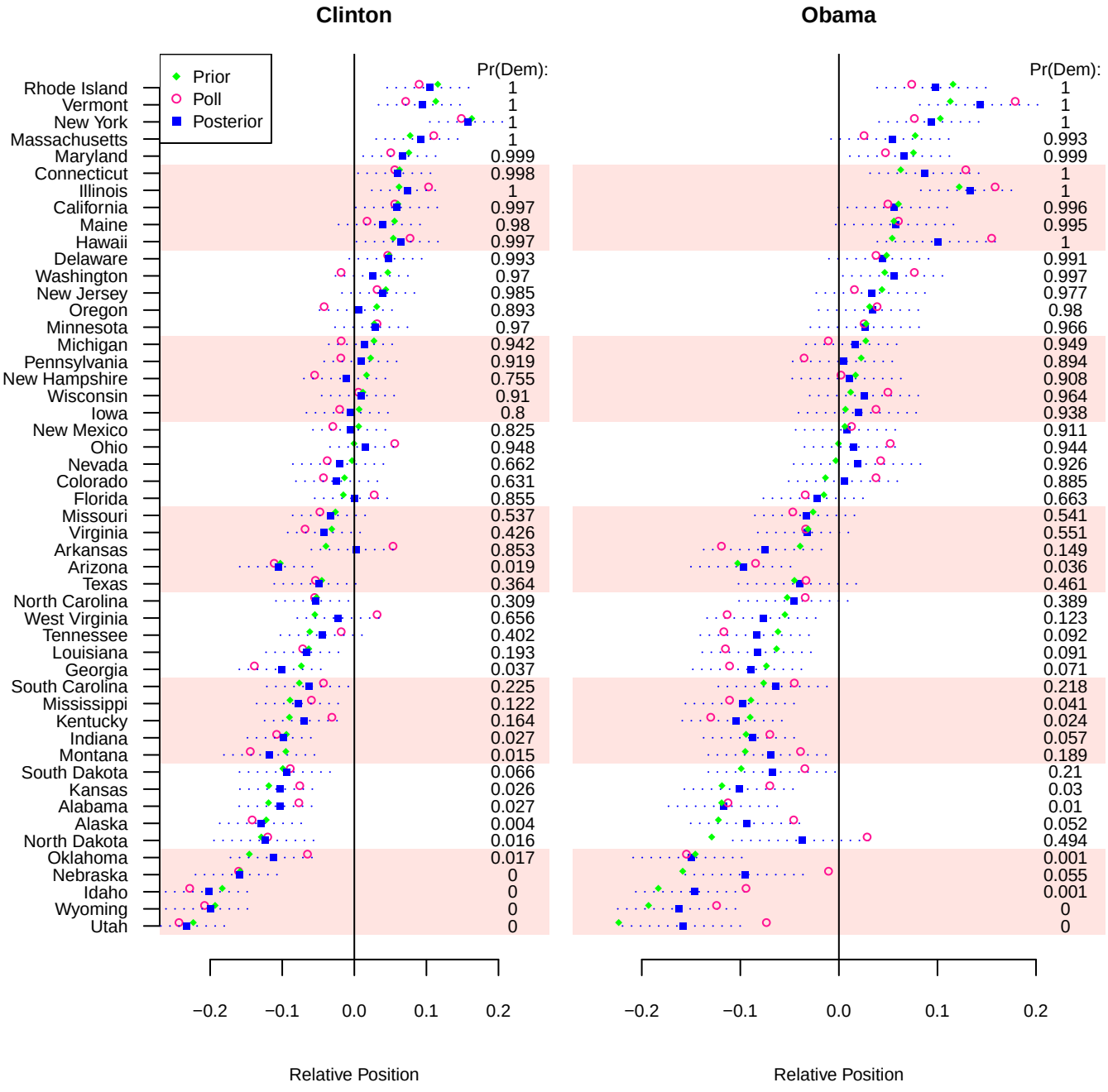

Figure 3: 95% posterior intervals for the relative position of each state, alongside prior and poll point estimates. The left column gives the probability of each state going Democratic (which incorporates the posterior for the national popular vote). States are ordered by 2004 Democratic vote share.

Substituting these numbers into (9)–(11) yields,

$$\text{Poll:} \quad \hat{p}_{\text{feb}}|p_0 \sim \text{N}\left(p_0, .086^2\right) \tag{12}$$

$$\text{Prior:} \quad p_0 \sim \text{N}\left(.5375,\ .021^2\right) \tag{13}$$

$$\text{Posterior:} \quad p_0|\hat{p}_{\text{feb}} \sim \text{N}\left(.06\hat{p}_{\text{feb}} + .94\hat{p}_{\text{hibbs}}, \frac{1}{\frac{1}{.086^2} + \frac{1}{.021^2}}\right) \sim N\left(.536, .020^2\right). \tag{14}$$

While the weight on our February poll data is relatively low (.06 for the popular vote and about .4 for the state relative positions), if the same polls had been conducted in October the weight on the poll estimates would shift to .35 for the popular vote, and around .9 for the state relative positions. The time the poll is conducted is key for determining the appropriate weights to place on the prior and the poll, and so for creating the posterior distributions.

Now that we have posterior distributions for both the national popular vote and each state's position relative to this, we can simply add them together to get posterior distributions for the proportion voting Democratic in each state. To create a posterior distribution for Obama's electoral college vote share, we simulate 100,000 elections, each time randomly drawing first a national popular vote from (14), and then simulating each state outcome by adding a draw from (8) to the simulated popular vote. The simulated electoral vote outcomes are shown in Figure 4(a) and have a posterior mean of 353 and standard deviation of 28. Of the 100,000 simulated elections, Obama won 99,870.

5 Discussion

5.1 Retrospective evaluation of our forecast

Our predictions were based on the SurveyUSA February poll data (for both the relative state positions and the popular vote estimate), the 2004 election results (for the relative state positions), and Hibbs' October estimate of the popular vote. Our analysis and the first draft of this paper up to this point were completed before November, 2008, and we added the present paragraph just after the election, allowing us to compare our posterior estimates with the actual election results. The actual two-party popular vote for Obama was 53.7%, very close to our posterior predictive mean of 53.6%. (Given our standard error, we do not claim any special magic in our method; we just happened to get lucky that it was so close.) At the national level, our forecast is barely distinguishable from that of Hibbs or, for that matter, mamy other political science forecasts based on "the fundamentals" (see Wlezien and Erikson, 2007). Where we go further is by using state-level information to get

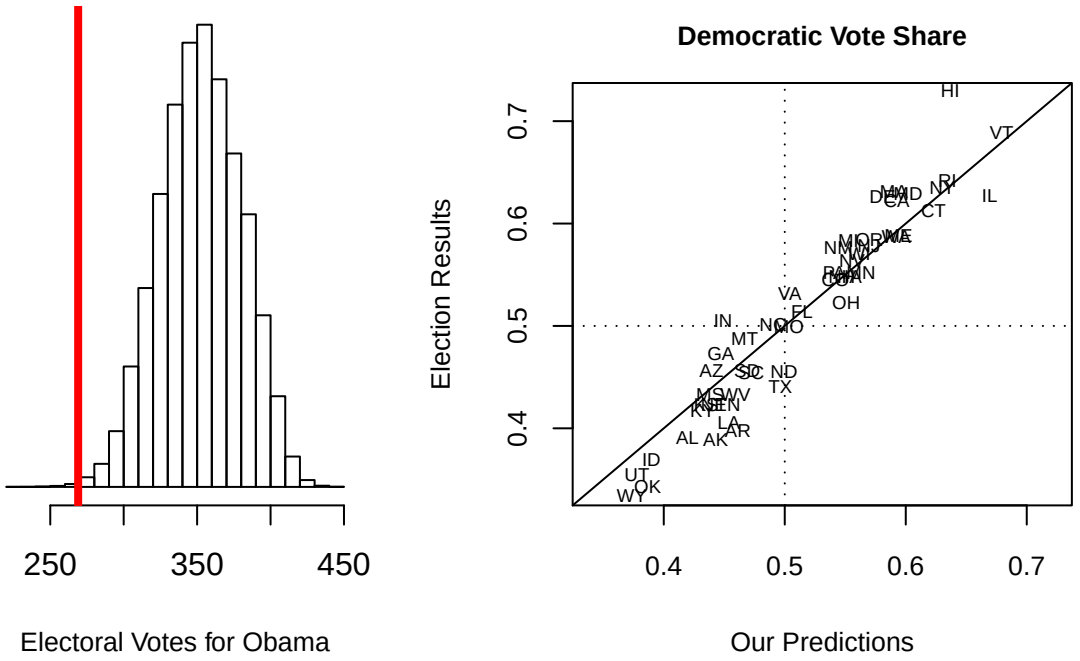

Figure 4: (a) Posterior distribution for Obama's electoral college vote share. Anything ≥ 270 indicates an Obama victory. (b) Actual election results, plotted against our prediction of the Democratic share of the two-party vote in each state.

a state-level forecast. The current state of the art in political journalism is poll aggregation, which is fine for tracking current opionion but doesn't make the best use of the information for the purpose of state-to-state forecasting.

Figure 4(b) shows the actual Democratic vote share for each state as compared to our predictions. We came quite close for most states, but we tended to overestimate Obama's popularity in Rebublican states and underestimate in Democratic states (a problem that also was present in pre-election poll aggregations; see Figure A.15 of Gelman et al., 2009). The correlation between our predicted values and actual values is .96, and the root mean square error (RMSE) of our estimates is $\sqrt{(1/50)\sum_{s=1}^{50}(p_{s,predicted} - p_{s,actual})^2} = .031$. The RMSE for fivethirtyeight.com's estimates, which use polls leading up the election, is .025. It is not surprising that you get closer to the truth using pre-election polls right before the election, but it is remarkable that we can do so well without using any polling data collected beyond February.

While the accuracy of our predictions is important, we also care about the calibration of our variance estimates, as every prediction needs an accompanying degree of uncertainty. The RMSE for our estimated state relative positions is .031, while our posterior standard

deviations range from .025 to .036, helping to improve the credibility of our variance estimates. The true position of each state falls within our 95% posterior intervals for 49 of the 50 states (we underestimated Hawaii), giving 98% coverage. For the relative state positions, we have 94% coverage, missing Hawaii, Arkansas, and Indiana. (Some of this has to be attributable to luck—the state estimates are correlated, and a large national swing could easily introduce a higher state-by-state error rate.)

5.2 The fundamental contradiction of up-to-the-minute poll aggregation

Polls can be aggregated to get a snapshot (or moving average) of public opinion, at the state or national level, but, as Wang (2008) has pointed out, such a snapshot is not the same as a forecast. For one example, presidential horse-race polls predictably jump during the parties' summer nominating conventions, but only a naive reader of the news would think that such jumps represent real increases in the probability of a candidate winning.

Tracking public opinion is a worthy goal in its own right, but if you are trying to forecast the presidental election, our message from this paper is that frequent polling provides very little information. Thus, as poll aggregation sites such as the Princeton Election Consortium, RealClearPolitics, FiveThirtyEight become more and more sophisticated at election forecasting, they will ultimately provide less and less in the way of relevant updates for their news-hungry consumers. This is not a bad thing—as with baseball statistics, the leading political statistics websites have already been moving from raw numbers and simple summaries toward more analytical modeling—and we hope the present article will do its part to shift political reporting toward informatino for the general voter and analysis for the political junkies, rather than horse-race summaries for both.

5.3 Conclusions

This paper has the goal of determining the strength of past elections and of pre-election polls in predicting a future election, and combining these sources to forecast the election. We found that to predict the current election, using the results of the most recent election is a good predictor of the way each state votes compared to the nation, but not necessarily of the national vote.

Hence, past election data are best used with a current estimate of the popular vote (such as can be obtained from polls or from forecasts that use economic and other information). Thus, our key contribution here is to separate the national forecast (on which much effort has been expended by many researchers) from the relative positions of the states (for which

past elections and current polls can be combined in order to make inferences). Pre-election polls, not surprisingly, are more reliable as they get closer to the election. Our advance with this analysis is quantification of this trend. Further work could be done (following Rosenstone, 1984, Campbell, 1992, and many others) in incoporating additional state-level economic and political information, while working within our framework that separates the national swing from relative movement among states. And we believe these ideas would be helpful in studying state-level public opinion and, more generally, any phenomenon that admits separate aggregate and relative forecasts.

References

[1] Annenberg Public Policy Center (2008). `www.annenbergpublicpolicycenter.org/AreaDetails.aspx?myId=1`, June.

[2] Bates, D. (2005). "Fitting linear models in R using the lme4 package," *R News*, 5(1): 27–30. `cran.r-project.org/doc/Rnews/Rnews_2005-1.pdf`

[3] Campbell, J. E. (1992). "Forecasting the Presidential Vote in the States," *American Journal of Political Science*, 36: 386–407.

[4] Erikson, R. S., and Sigman, K. (2008). "Guest Pollster: The Survey USA 50 State Poll and the Electoral College," Pollster.com, March. `www.pollster.com/blogs/guest_pollster_the_surveyusa_5.php`

[5] 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.

[6] Gelman, A., Park, D., Shor, B., and Cortina, J. (2009). *Red State, Blue State, Rich State, Poor State: Why Americans Vote the Way They Do*, second edition. Princeton University Press.

[7] Hibbs, D. A. (2008). "Implications of the "Bread and Peace" Model for the 2008 US Presidential Election," *Public Choice*, September.

[8] Lax, J. R., and Phillips, J. H. (2009). "Gay Rights in the States: Public Opinion and Policy Responsiveness," *American Political Science Review*, 103.

[9] Rosenstone, S. J. (1983). *Forecasting Presidential Elections*, New Haven, Conn.: Yale University Press.

[10] Silver, N. (2008). `www.fivethirtyeight.com/`, August.

[11] Strauss, A. (2007). "Florida or Ohio? Forecasting Presidential State Outcomes Using Reverse Random Walks," Princeton University Political Methodology Seminar.

[12] Survey Sampling International (2008). `www.surveysampling.com`, June.

[13] Wang, S. (2008). Princeton Election Consortium FAQ. `http://election.princeton.edu/faq/`

[14] Wlezien, C., and Erikson, R. S. (2004). "The Fundamentals, the Polls, and the Presidential Vote," *Political Science and Politics*, 37: 747–751.

[15] Wlezien, C., and Erikson, R. S. (2007). "The Horse Race: What Polls Reveal as the Election Campaign Unfolds," *International Journal of Public Opinion Research*, 19: 74–88.