A Bayesian Prediction Model for the United States Presidential Election

Steven E. Rigdon
Southern Illinois University Edwardsville

Sheldon H. Jacobson
University of Illinois at Urbana-Champaign

Wendy K. Tam Cho
University of Illinois at Urbana-Champaign

Edward C. Sewell
Southern Illinois University Edwardsville

Christopher J. Rigdon*
Arizona State University

Abstract

It has become a popular pastime for political pundits and scholars alike to predict the winner of the United States presidential election. Although forecasting has now quite a history, we argue that the closeness of recent presidential elections and the wide accessibility of data should change how presidential election forecasting is conducted. We present a Bayesian forecasting model that concentrates on the Electoral College outcome and considers finer details such as third-party candidates and self-proclaimed undecided voters. We incorporate our estimators into a dynamic programming algorithm to determine the probability that a candidate will win an election.

Key Words: Presidential Forecasting Models, Bayesian Statistics, Operations Research.

*Steven E. Rigdon is in the Department of Mathematics and Statistics, Southern Illinois University Edwardsville, Edwardsville, IL 62026-1653. Sheldon H. Jacobson is in the Department of Computer Science, Simulation and Optimization Laboratory, 201 N. Goodwin Avenue (MC–258), University of Illinois, Urbana, IL 61801-2302. Wendy K. Tam Cho is in the Departments of Political Science and Statistics and National Center for Supercomputing Applications, University of Illinois at Urbana-Champaign, Urbana, IL 61801-3696. Edward C. Sewell is in the Department of Mathematics and Statistics, Southern Illinois University Edwardsville, Edwardsville, IL 62026-1653. Christopher J. Rigdon is in the Department of Industrial Engineering, Arizona State University, Tempe, AZ 85287-5906.

We gratefully acknowledge support from the National Science Foundation (Grant IIS 08-27540 SGER).

1 Election Forecasting and Recent Elections

Recent presidential elections indicate that the American voting population is deeply divided. This was plainly obvious in 2000 when Al Gore won more popular votes than George W. Bush, but lost the presidency in the Electoral College tally. In 2004, when Bush secured 50.7% of the popular vote to Kerry's 48.3%, claims of voting irregularities again surfaced. This time the disputes were in Ohio where Bush bested Kerry by 118,775 votes. Although Kerry ultimately decided not to challenge the vote count, if he had somehow mounted a victorious challenge, and Ohio's 20 electoral votes would have been placed in the Democratic camp, Kerry would have won the election with 271 electoral votes—and for the second time in two elections, won while losing the popular vote! When Gore lost, pundits and journalists alike chanted refrains that this phenomenon had not surfaced since 1888 when Benjamin Harrison beat Grover Cleveland in the Electoral College while pulling in 0.8% less of the popular vote. In fact, this oddity is slightly more common, with the more recent manifestation in 1960 when young John F. Kennedy secured the Presidency over Richard Nixon by winning the Electoral College vote, but not the popular vote (Gaines, 2001). These disparities between the winner of the popular vote and the Electoral College are rare, but far from unheard of, and becoming common enough that there have been numerous calls in the last decade for a reform of the Electoral College system (Edwards, 2004; Goux and Hopkins, 2008; Grofman and Feld, 2005; Whitaker and Neale, 2004; Tokaji, 2008; Hiltachk, 2008; Hirsch, 2008; Hansen, 2008; Leib and Mark, 2008; Belenky, 2008)

A reform of the Electoral College system seems unlikely given the intentional difficulty involved in modifying the U.S. Constitution (Bennett, 2006; Rathbun, 2008). While Gore, especially, appeared to be drumming up support for the idea that he was the legitimate winner of the 2000 presidential election, all presidential candidates enter the race knowing full well how votes are counted, and hence, tailor their campaign strategies accordingly. Candidates spend far more of their campaign funds and time in states that are close and pay scant attention to voters in states where the outcome is basically foregone (Althaus, Nardulli and Shaw, 2002; Shaw, 2006; Doherty, 2007). It seems a bit disingenuous after votes are counted to then appeal to voters that a different vote counting scheme conveys the true and legitimate winner. Instead, there is no ambiguity that the division of the popular vote is interesting,

but not decisive. Moreover, it is unclear that the popular vote accurately conveys the preference of the voters better than the Electoral College system. Instead, the weight of the evidence supports the claim that majority will is difficult to measure and fundamentally ambiguous (Rae, 1972; Longley and Dana, 1992; Grofman, Brunell and Campagna, 1997; Gelman, Katz and Tuerlinckx, 2002; Grofman and Feld, 2005).

Also interestingly, presidential vote forecasting models have been heavily focused on the two-party popular vote (Wlezien and Erikson, 2004; Lewis-Beck and Tien, 2004; Abramowitz, 2004; Campbell, 2004; Norpoth, 2004; Holbrook, 2004; Lockerbie, 2004) with only few exceptions (Kaplan and Barnett, 2003). These models have had varying degrees of success by Campbell's 2005 accounting method where less than 2 percentage points is "quite accurate" (Wlezien and Erikson, 2004; Lewis-Beck and Tien, 2004), between 2 and 3 points as "reasonably accurate," closer to the accuracy of the final polls than the pre-campaign polls (Abramowitz, 2004; Campbell, 2004), between 3 and 4 points as "fairly accurate" (Norpoth, 2004), between 4 and 5 points as "inaccurate" (Holbrook, 2004), and in excess of 5 points as "very inaccurate" (Lockerbie, 2004). The Wlezien and Erikson (2004) error was a mere 0.5%, which is undoubtedly an impressive forecast of the popular vote. However, we know, theoretically, through simple math, and empirically, from recent elections, that 0.5% is far and away more than enough votes to swing the election in the Electoral College if they hail from particular geographic locations. It is an artifact of our election system that even the most impressive forecasts of the popular vote, and indeed, the precise popular vote, could point to the loser rather than the victor of a U.S. presidential election.

In the past, a focus on the national two-party popular vote was understandable from a data perspective since it greatly reduces data demands. One needs only a single representative national poll to forecast the popular vote. If one wants to focus instead on the Electoral College, one would need a sample for every state in the union. Today, the proliferation of the Internet has greatly facilitated the transfer of large quantities of data, and made state-by-state polling data readily available. Since the data acquisition obstacle has been dissipated, it seems clear, then, that the ease of data acquisition and the evident dissimilarity between the popular vote and the Electoral College outcome should push presidential forecasting models to shift their attention from the popular vote to electoral votes. In addition, the closeness of recent elections might also imply that forecasting models should include factors

such as third party candidates and the proportion of undecided voters. Both of these entities have the potential to change the course of a close election. Even if elections are not close, a bit of fine tuning, if done well, is always a welcome addition.

In this paper, we propose to move election forecasting in both of these directions with a model focused on the Electoral College that also pays attention to finer, and potentially election-determining, details such as third party candidates and voters who proclaim to be undecided in election polls. Our framework employs a Bayesian estimator that allows one to use both prior and current information for each state to determine each candidate's probability of winning that state in the presidential election. The estimators incorporate both the previous election's results (to capture each state's party tendency) and current polling data (to capture each state's current candidate tendency). We utilize an informative prior based on long term voting trends within a state and different probability distributions. Once we estimate the probability that a candidate will win a state, these values can be used to determine the probability that each candidate will win the election by employing a Monte Carlo simulation or using a dynamic programming algorithm (Kaplan and Barnett, 2003). The idea of using Bayesian estimation rather than frequentist estimation techniques has been explored by others (Kaplan and Barnett, 2003; Jackman and Rivers, 2001), though our formulation differs.

Our paper is organized as follows. Section 2 provides a description of the Bayesian estimators for each candidate's probability of winning each state's Electoral College votes. Section 3 uses these estimators, coupled with polling data that was available, to retrospectively provide a prediction and analysis of the 2004 United States presidential election. Section 4 discusses the results and provides concluding comments and directions for future research.

2 Bayesian Estimators

A Bayesian analysis differs from a frequentist analysis in that it uses observations to update or newly infer some unknown quantity of interest, here, the outcome of the presidential election. The term "Bayesian" arises from the frequent use of Bayes' Theorem, which relates the conditional and marginal probabilities of two events. Two fundamental elements of Bayesian analyses are the prior distribution

and the posterior distribution. The prior distribution encompasses our beliefs prior to observing some data. In our application, the prior distribution is based on historical voting patterns that are used to construct each state's normal vote (Stokes, 1962; Converse, 1966; Nardulli, 2005). We note that our formulated Bayesian estimators could, of course, incorporate any prior, including a noninformative prior or a prior based on a different forecasting model run at the state level (e.g., Wlezien and Erikson (2004)). Constructing a prior based on a popular vote forecasting model is certainly reasonable and would be one way of combining our proposed methods with existing methods. Without other knowledge of the current election, in our application here, our belief about the outcome of the election will be based on historical trends. Consider Table 1, which shows the voting behavior of Alabama and New York, over five election years (1984–2000). Quite clearly, a state's partisan tendency is a useful and informative prior. This "prior" information can be used along with current polling data to formulate the posterior distribution (using Bayes' Theorem) for the probability of each candidate winning a state. The posterior distribution is a transformation of the prior distribution where the transformation embodies an updating of beliefs after observing some data. Since state-by-state polling data are more reliable for predicting the winner of the election than the previous year's election results, the posterior distribution is constructed such that the likelihood function dominates the prior distribution. The posterior distribution for each state is then used to compute the probability that a candidate wins the state's Electoral College votes. Our Bayesian estimators also take into account the impact from voters who declare that they are undecided prior to the election, as well as third party candidates, to obtain the posterior distribution or an estimate of each candidate's probability of winning the election. Finally, these probabilities are used as input into a dynamic programming algorithm to determine the distribution of Electoral College votes for each candidate, and hence, the probability that each candidate wins the election.

2.1 Bayesian Formulation

More formally, define p_i to be the true proportion of voters in a state who intend to vote for candidate i in the election (for simplicity, let $i = 1$ correspond to the Republican candidate, $i = 2$ correspond to the Democratic candidate, $i = 3$ collectively correspond to all third party candidates, and $i = 4$ correspond to no candidate or voters who have declared that they are still undecided). These

Table 1: Voting Behavior of Alabama and New York

Year	**Alabama** Republican Vote	**New York** Democratic Vote
1984	60.5%	45.8%
1988	59.2%	51.5%
1992*	47.7%	49.8%
1996	50.1%	59.5%
2000	56.5%	60.4%

* Ross Perot captured approximately 19% of the popular vote in 1992.

proportions are assumed to be continuous (between zero and one) and sum to one. The joint prior distribution for $\mathbf{p} = (p_1, p_2, p_3, p_4)$ is assumed to be a conjugate prior distribution (i.e., when combined with a multinomial distribution, the same type of posterior distribution is obtained). To satisfy this requirement, assume that $\mathbf{p}$ follows a Dirichlet distribution, $\mathbf{p} \sim \text{DIRICHLET}(b_1, b_2, b_3, b_4)$, which is a multivariate generalization of the beta distribution, and is often used as a prior for the probability of a success in Bernoulli trials. Therefore, the joint probability density function of $\mathbf{p}$ can be written as

$$f(p_1, p_2, p_3, p_4) = c\, p_1^{b_1-1} p_2^{b_2-1} p_3^{b_3-1} p_4^{b_4-1}, \quad p_i \geq 0 \text{ for } i = 1, 2, 3, 4, \ \sum_{i=1}^{4} p_i = 1,$$

where $c = \Gamma\left(\sum_{i=1}^{4} b_i\right) \ / \ \prod_{i=1}^{4} \Gamma(b_i)$.

The probability that a candidate wins a given state can be computed using the marginal probability densities. To obtain these marginals, we sequentially integrated the remaining variables out of the joint Dirichlet probability density function. We now illustrate this process by first rewriting the joint Dirichlet probability density function as

$$f_{1,2,3}(p_1, p_2, p_3) = c\, p_1^{b_1-1} p_2^{b_2-1} p_3^{b_3-1} (1 - p_1 - p_2 - p_3)^{b_4-1}, \quad p_1, p_2, p_3 \geq 0, \ \sum_{i=1}^{3} p_i \leq 1. \tag{1}$$

Integrating over p_3 leads to an expression for the joint probability density of p_1 and p_2,

$$f_{1,2}(p_1,p_2) = \int_0^{1-p_1-p_2} c\, p_1^{b_1-1} p_2^{b_2-1} p_3^{b_3-1} (1-p_1-p_2-p_3)^{b_4-1}\, dp_3 \tag{2}$$

$$= c\, p_1^{b_1-1} p_2^{b_2-1} \int_0^{1-p_1-p_2} p_3^{b_3-1} (1-p_1-p_2-p_3)^{b_4-1}\, dp_3. \tag{3}$$

Using the result

$$\int_0^{1-y} x^{\alpha-1} (1-y-x)^{\beta-1}\, dx = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)} (1-y)^{\alpha+\beta-1},$$

where the gamma function, $\Gamma(x) = \int_0^\infty t^{x-1} e^{-t}\, dt = 2\int_0^\infty u^{2x-1} e^{-u^2}\, du$, leads to the expression

$$f_{1,2}(p_1,p_2) = C_1 p_1^{b_1-1} p_2^{b_2-1} (1-p_1-p_2)^{b_3+b_4-1},\ p_1+p_2 \le 1,\ p_1,p_2 \ge 0, \tag{4}$$

where $C_1 = c \cdot \Gamma(b_3)\Gamma(b_4)\ /\ \Gamma(b_3+b_4)$. Integrating (4) over all possible values of p_2 gives the marginal density of p_1,

$$f_1(p_1) = \int_0^{1-p_1} C_1 p_1^{b_1-1} p_2^{b_2-1} (1-p_1-p_2)^{b_3+b_4-1}\, dp_2 \tag{5}$$

$$= C_1 p_1^{b_1-1} \int_0^{1-p_1} p_2^{b_2-1} (1-p_1-p_2)^{b_3+b_4-1}\, dp_2 \tag{6}$$

$$= C_2\, p_1^{b_1-1} (1-p_1)^{b_2+b_3+b_4-1},\ 0 \le p_1 \le 1, \tag{7}$$

where $C_2 = C_1 \cdot \Gamma(b_2)\Gamma(b_3+b_4)\ /\ \Gamma(b_2+b_3+b_4)$. Therefore, by the form of $f_1(p_1)$, p_1 is distributed as a beta random variable with parameters b_1 and $b_2+b_3+b_4$. Using the identical argument, p_2, p_3, p_4 are also distributed as beta random variables, and hence,

$$p_i \sim BETA\left(b_i, \sum_{k=1}^{4} b_k - b_i\right),\ i = 1,2,3,4. \tag{8}$$

2.2 Calibration of Prior Parameters

The beta distribution is a family of continuous probability distributions characterized by two shape parameters, here, b_1 and $b_2+b_3+b_4$, which must be chosen. Different choices can be incorporated,

which would result in quite varied substantive implications. For example, one possibility is to set these values so that the expected value and the variance, or the first and second moments, for $p_i, i = 1, 2, 3, 4$, closely match observed elections. However, there are an infinite number of combinations of b_1, b_2, b_3, b_4 that result in the same values for the expectations and the variances for b_1, b_2, b_3, b_4. That is, if the expectation and variance of a beta random variable are given by

$$E(p_i) = \frac{b_i}{\sum_{k=1}^{4} b_k}, \tag{9}$$

and

$$\begin{aligned} \mathrm{Var}(p_i) &= \frac{b_i \left(\sum_{k=1}^{4} b_k - b_i \right)}{\left(\sum_{k=1}^{4} b_k + 1 \right) \left(\sum_{k=1}^{4} b_k \right)^2} && (10) \\ &= \frac{\frac{\rho}{(\rho+\tau)} \left(1 - \frac{\rho}{\rho+\tau} \right)}{\rho + \tau + 1}, && (11) \end{aligned}$$

where $\rho = b_i$ and $\tau = \sum_{k=1}^{4} b_k - b_i$, then if $b_1 = b_2 = 4$ and $b_3 = b_4 = 1$ or $b_1 = b_2 = 40$ and $b_3 = b_4 = 10$, the prior means are the same, namely

$$\begin{aligned} E(p_1) &= E(p_2) = \frac{4}{10} = 0.4 \\ E(p_3) &= E(p_4) = \frac{1}{10} = 0.1. \end{aligned}$$

The choice of the shape parameters is essentially arbitrary if we do not issue any constraints or do not employ any substantive guidance. Fortunately, in presidential forecasting, we have a great deal of substantive knowledge that can be integrated. One way to constrain our choices is to choose the b's so that 1) $b_i / \sum_{k=1}^{4} b_k$ equals what p_i is expected to be, prior to observing the polling data, and 2) the spread of the prior distribution for p_i (determined by $\sum_{k=1}^{4} b_k$, with larger values indicating less uncertainty) reflects the perceived uncertainty in p_i.

To illustrate this process, consider the marginal probability density of p_1 in the candidates' home states: Massachusetts and Texas. In 2000, the most recent presidential election before 2004, Bush received 33% of the popular vote in Massachusetts and 59% of the popular vote in Texas. If we let $\sum_{k=1}^{4} b_k = 4$ and set b_1 such that $E(p_1)$ equals the percentage of the popular vote that the Republican candidate won in the 2000 election in the given state (after adjusting for undecided voters—see below), the two marginal densities for p_1 are depicted in Figure 1. Under these choices for the b_i's, Senator Kerry, had a 0.374 probability of winning Texas, while President Bush had a 0.208 probability of winning Massachusetts. These probabilities seem too high given both the historical voting patterns of these states and that these states are the home states of the candidates (Lewis-Beck and Rice, 1983). Another choice would be to let $\sum_{k=1}^{4} b_k = 400$, in which case Senator Kerry has a 0.002 probability of winning Texas, while President Bush has a 10^{-15} probability of winning Massachusetts (see Figure 2). These probabilities seem too low given historical trends. Perhaps a value in between these two choices would be more ideal. Toward this effort, setting $\sum_{k=1}^{4} b_k = 40$ gives Senator Kerry has a 0.177 probability of winning Texas and gives President Bush a 0.010 probability of winning Massachusetts (see Figure 3).[1] Of these values (4, 400, and 40), 40 appears to be the most substantively grounded and so we move forward with $\sum_{k=1}^{4} b_k$ being set to 40.

Our priors for the Republican and Democratic candidates involve the normal vote for all states except Alaska and Hawaii (and also DC). For the combined third party candidates, we chose the mean of the prior to be equal to the combined third party vote in 2000. The prior mean for each of the Republican and Democratic candidates for president was taken to be the normal vote reduced by half of the third part vote for that state.

In order to incorporate the effect from voters who declared that they were undecided, let us assume that the prior mean for undecided voters is 3%. This is an assumption in the purest sense, but it also seems reasonable as a prior given the percentage of undecided voters in our various polls. Of course, our method is not wedded to this value and users are free to incorporate different values as they see fit. We could base this figure on older polls or even an intuition of current trends since the last election.

[1] We did examine other values for $\sum_{k=1}^{4} b_k$, ranging from 10 to 70. Values around 40 yield similar results, while values significantly larger or smaller than 40 may not be appropriate.

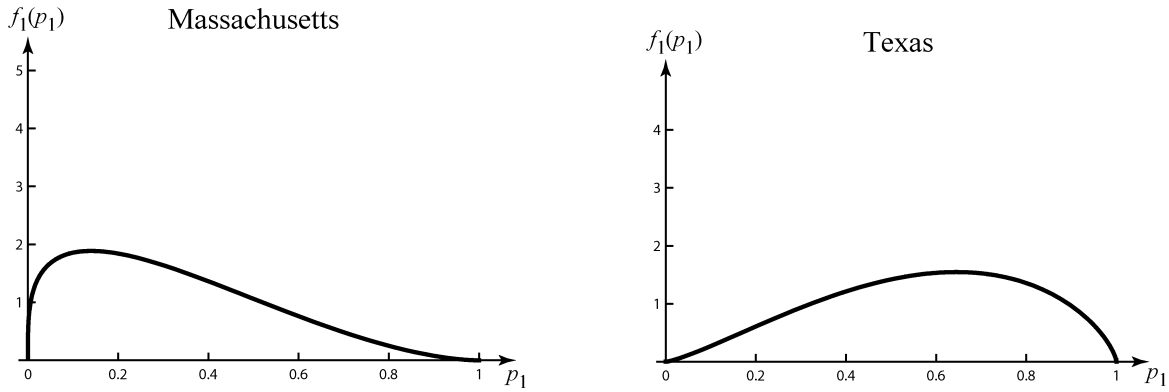

Figure 1: Marginal PDF for p_1 in Massachusetts and Texas when Sum of Prior Parameters is 4.

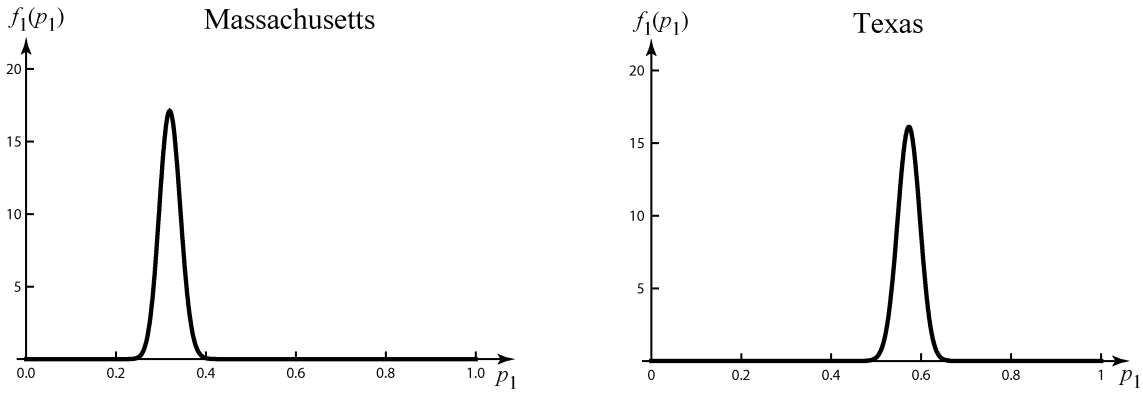

Figure 2: Marginal PDF for p_1 in Massachusetts and Texas when Sum of Prior Parameters is 400.

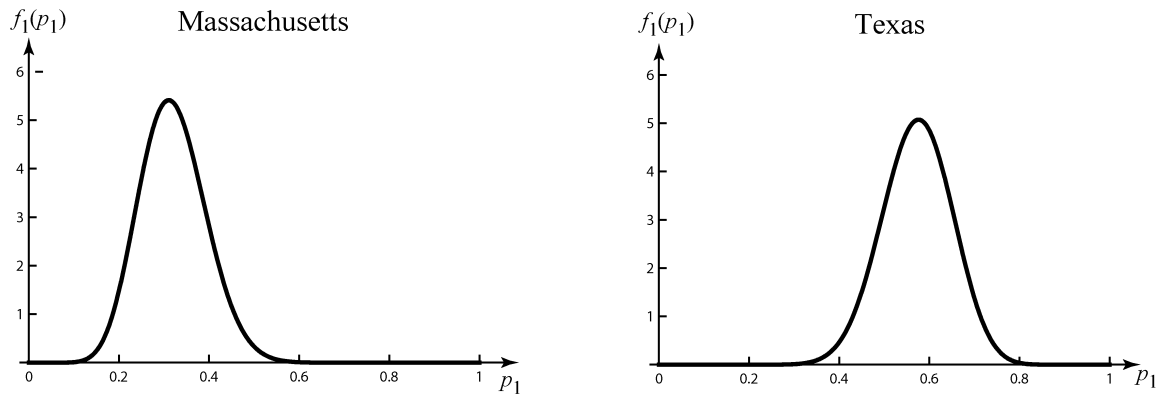

Figure 3: Marginal PDF for p_1 in Massachusetts and Texas when Sum of Prior Parameters is 40.

Thus, the prior parameters for a given state were chosen to be

$$\begin{aligned} b_1 &= 40\,(1-0.03)\left(NV_1 - \frac{1}{2}C_3\right) \\ b_2 &= 40\,(1-0.03)\left(NV_2 - \frac{1}{2}C_3\right) \\ b_3 &= 40\,(1-0.03)\,C_3 \\ b_4 &= 40\,(0.03), \end{aligned}$$

where NV_i is the normal vote for candidate i and C_3 is the proportion of the 2000 vote for third party candidates. Using these parameters, the prior distribution for a given state is

$$\mathbf{p} = (p_1, p_2, p_3, p_4) \sim \text{ DIRICHLET}\left(38.8\left(NV_1 - \frac{1}{2}C_3\right), 38.8\left(NV_2 - \frac{1}{2}C_3\right), 38.8\,C_3, 1.2\right).$$

2.2.1 Illustrations of Priors

The backdrop of the 2000 election may be used to illustrate how this prior works with real data. In 2000, 49% of Floridians voted for Governor Bush, 49% voted for Vice President Gore, and 2% voted

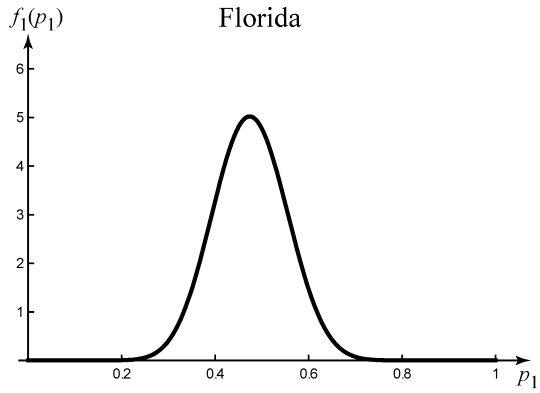

Figure 4: Marginal Density of p_1 for Florida.

for some other candidate. Therefore, the prior distribution for $\mathbf{p}$ in Florida is

$$\mathbf{p} = (p_1, p_2, p_3, p_4) \sim \text{ DIRICHLET}\,(19.012,\ 19.012,\ 0.776,\ 1.2)\,.$$

From Figure 4, we can see that the marginal prior probability density of p_1 and p_2 are identical in this case.

To obtain an expression for the likelihood function, let $\mathbf{X} = (X_1, X_2, X_3, X_4)$ denote the random vector of sample proportions in a state poll for $\mathbf{p}$. Therefore, for a survey with n respondents,

$$\mathbf{X} = (X_1, X_2, X_3, X_4) \sim \text{MULTINOMIAL}\,(n, p_1, p_2, p_3, p_4)\,,$$

where the joint probability density of $\mathbf{X}$ is

$$g\,(\mathbf{X}\,|\,\mathbf{p}) = \frac{n!}{x_1!\,x_2!\,x_3!\,x_4!}\,p_1^{x_1}\,p_2^{x_2}\,p_3^{x_3}\,p_4^{x_4}, \qquad \sum_{k=1}^{4} p_k = 1, x_i \geq 0.$$

As another example, consider a SurveyUSA poll (conducted in October, 2004) in Florida that had 742 respondents. In this poll, 364 expressed an intention to vote for President Bush, 356 favored Senator Kerry, and the final 22 respondents were undecided. Using these polling data, the likelihood function

for Florida voters is

$$g\left(\mathbf{X}\,|\,\mathbf{p}\right) = \frac{742!}{364!\,356!\,0!\,22!}\,p_1^{364}\,p_2^{356}\,p_3^{0}\,p_4^{22}, \qquad \sum_{k=1}^{4} p_k = 1.$$

Bayes' Theorem can be used to obtain the posterior distribution for $\mathbf{p}$. In particular,

$$h(\mathbf{p}\,|\,\mathbf{X}) = C_B\ f(\mathbf{p}) \cdot g(\mathbf{X}\,|\,\mathbf{p}),$$

where

$h\left(\mathbf{p}\,|\,\mathbf{X}\right)$ is the posterior distribution density (for the probability of winning a state)

C_B is a proportionality constant, chosen so that the prior integrates to one

$f\left(\mathbf{p}\right)$ is the prior distribution density (in this case, Dirichlet)

$g\left(\mathbf{X}\,|\,\mathbf{p}\right)$ is the likelihood function density (in this case, multinomial).

Therefore,

$$\begin{aligned} h\left(\mathbf{p}\,|\,\mathbf{X}\right) &= C_B\ c\ p_1^{b_1-1}\ p_2^{b_2-1}\ p_3^{b_3-1}\ p_4^{b_4-1}\ \frac{n!}{x_1!\,x_2!\,x_3!\,x_4!}\ p_1^{x_1}\ p_2^{x_2}\ p_3^{x_3}\ p_4^{x_4} \\ &= C\ p_1^{x_1+b_1-1}\ p_2^{x_2+b_2-1}\ p_3^{x_3+b_3-1}\ p_4^{x_4+b_4-1}, \qquad p_i \geq 0, \forall\, i,\ \sum_{i=1}^{4} p_i = 1, \end{aligned}$$

where

$$C = \frac{\Gamma\left(\sum\limits_{k=1}^{4}\left(b_k + x_k\right)\right)}{\prod\limits_{k=1}^{4}\Gamma\left(b_k + x_k\right)}$$

is a constant of proportionality. This result means that the posterior distribution for $\mathbf{p}$ given $\mathbf{X}$ (polling data results) is distributed as DIRICHLET$\left(x_1 + b_1,\ x_2 + b_2,\ x_3 + b_3,\ x_4 + b_4\right)$, which translates to DIRICHLET(383.012, 377.012, 0.776, 23.2) for the Florida data. The marginal densities of p_1 and p_2 are displayed in Figure 5.

2.2.2 Incorporating the Role of Undecided Voters

Our method also includes a means for incorporating the uncertainty presented by voters who remain undecided up until election time. These variable voters are accounted for by different swing scenarios,

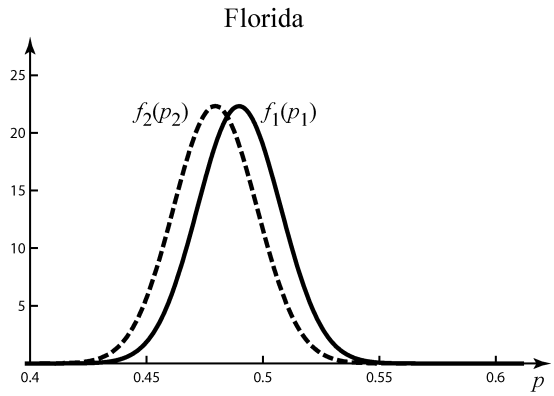

Figure 5: Marginal Density for p_1 (solid line) and p_2 (dotted line).

which we term the "no swing scenario," the "Republican swing scenario," and the "Democratic swing scenario." In a "no swing" scenario, 48% of declared undecided voters vote for the Republican candidate and 48% vote for the Democratic candidate. Alternatively, there may be a "Republican swing" where 52% voted for the Republican candidate and 44% voted for the Democratic candidate or a "Democratic swing" scenario where 44% vote for the Republican candidate and 52% vote for the Democratic candidate. Note that in each of the three scenarios, 4% of the undecided voters are assumed to vote for a third party candidate, which may or may not capture the reality, particularly if undecided voters ultimately abstain from the election. This is an assumption of the model that can be modified by the user if the user feels that it is far flung from any conceivable reality. After perusing exit poll results for our election of interest, we felt comfortable with this choice.

The posterior probability that a particular candidate wins a state can be computed from the joint posterior distribution of the p's. For the "no swing" scenario, we need only the posterior of (p_1, p_2). For example, the probability that the Republican candidate wins a state is

$$\begin{aligned} P\left(p_1 + 0.48\, p_4 > p_2 + 0.48\, p_4 \mid \mathbf{x}\right) &= P\left(p_1 > p_2 \mid \mathbf{x}\right) \\ &= \int_0^{1/2} \int_{p_2}^{1-p_2} f_{12}\left(p_1, p_2 \mid \mathbf{x}\right)\, dp_1\, dp_2. \end{aligned}$$

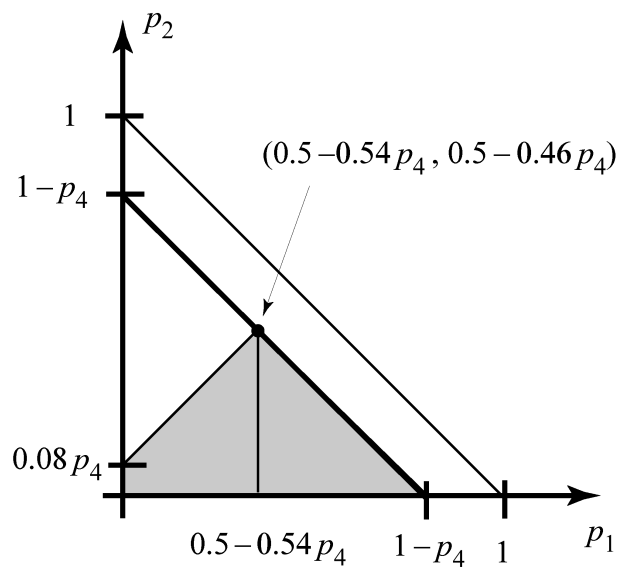

Figure 6: Region of integration in the p_1p_2-plane for the "Republican Swing" scenario.

For the "Republican swing" scenario, the posterior probability must be computed as a triple integral over the appropriate region. Figure 6 shows the region in the p_1p_2-plane for a fixed value of p_4; p_4 then goes from 0 to 1. As this figure suggests, the triple integral must be broken into two parts, giving

$$\begin{aligned} P\left(p_1 + 0.54\,p_4 > p_2 + 0.46\,p_4 \mid \mathbf{x}\right) &= P\left(p_1 + 0.08\,p_4 > p_2 \mid \mathbf{x}\right) \\ &= \int_0^1 \int_0^{0.5-0.54\,p_4} \int_0^{p_1+0.08\,p_4} f_{124}\left(p_1, p_2, p_4 \mid \mathbf{x}\right)\ dp_2\ dp_1\ dp_4 \\ &\quad + \int_0^1 \int_{0.5-0.54\,p_4}^{1-p_4} \int_0^{1-p_4-p_1} f_{124}\left(p_1, p_2, p_4 \mid \mathbf{x}\right)\ dp_2\ dp_1\ dp_4. \end{aligned}$$

The posterior probability under the "Democratic swing" scenario is similarly computed.

Note that this analysis assumes that there is no third-party candidate who has a nonnegligible chance of winning a state. If a viable third-party candidate were present, we would need to make assumptions about the swing of undecided votes among the other candidates. The posterior probability that

$$p_1 + (\text{Republican swing}) * p_4 \quad > \quad p_2 + (\text{Democratic swing}) * p_4$$

and

$$p_1 + (\text{3rd party swing}) * p_4 \quad > \quad p_3 + (\text{3rd party swing}) * p_4$$

would then need to be computed. Although our model formulation can incorporate the impact of a third party candidate, this feature was not needed in 2004 since no third party candidate mounted a viable candidacy in this particular race.

2.2.3 Implementation

In order to obtain the probability that a candidate will win the electoral votes in any given state, we must compute a series of fairly involved formulas. For example, to implement the "no swing" scenario, we must evaluate a double integral. The "Republican swing" and the "Democratic swing" scenarios involve triple integrals. These particular integrals cannot be solved analytically. Instead, they must be approximated numerically. Accordingly, implementation of our model requires a software package that is able to compute numerical approximations for integrals. In addition, because these computations involve both very large numbers (from the Gamma function) and very small numbers (from the fractions raised to large powers), the numerical evaluations are prone to inaccuracies unless a high degree of precision is maintained throughout the computation. We used WinBUGS to obtain the probabilities that each candidate would win a state. The dynamic programming algorithm that uses the state-by-state probabilities to compute the probability distribution for the number of Electoral College votes that a candidate will receive was implemented in Matlab. Implementation of our model, however, can be completed in other software packages as well. For example, Mathematica could be used to evaluate the integrals.

3 Forecasting Presidential Elections

We now illustrate the use of our estimators for forecasting the 2004 United States presidential election. Our state-level surveys for this election were reported by Real Clear Politics, an independent company that gathers information from numerous publishing companies and organizes it for public

consumption on their web site. The polls were gathered from a variety of companies including Zogby, Rasmussen Reports, NBC, the Wall Street Journal, the American Research Group, Fox News, and Survey USA, among others.

For each state, the posterior distribution for the proportion of voters who will vote for a candidate, the prior distribution for $\mathbf{p}$, and the likelihood distribution for $\mathbf{X}$, are all explicitly given in the Section 2. The posterior probabilities, shown in Table 2, can be used to compute the probability that a candidate wins a state. A candidate who wins a state is awarded the number of Electoral College votes associated with that state (with the exception of Maine and Nebraska). To estimate the probability that President Bush wins a state, the posterior probability that $p_1 > p_2$ was computed, which is equivalent to assuming that all third-party candidates have a zero probability of winning a state. In 2004, no third party candidate was in a position to win a state, though third-party candidates were influential in close states.

Once the probabilities of winning each individual state have been computed, we used a dynamic programming algorithm to compute the probability distribution for the Electoral College votes. More formally, in this stage, we number the states (including the District of Columbia) from 1 to 51, and let $p^{(k)}$ be the probability that the candidate wins state k, and v_k be the number of Electoral College votes for state k. Let $P(i,k)$ be the probability that the candidate wins exactly i Electoral College votes in states $1, 2, \ldots, k$. Then, $P(i,k)$ can be computed for $k \geq 2$ via the following recurrence relation:

$$P\left(i,k\right) = \begin{cases} (1-p^{(k)})P(i,k-1) \ + \ p^{(k)}P(i-v_k,k-1) & \text{if } i \geq v_k \\ (1-p^{(k)})P(i,k-1) & \text{if } i < v_k, \end{cases}$$

with boundary conditions (i.e., $k = 1$),

$$P\left(i,1\right) = \begin{cases} 1-p^{(1)} & \text{if } i = 0 \\ p^{(1)} & \text{if } i = v_1 \\ 0 & \text{otherwise.} \end{cases}$$

Table 2: Posterior Probabilities that President Bush Wins Each State in 2004

State	**Dem. Swing**	**No Swing**	**Rep. Swing**	**State**	**Dem. Swing**	**No Swing**	**Rep. Swing**
Alabama	1.00	1.00	1.00	Montana	1.00	1.00	1.00
Alaska	1.00	1.00	1.00	Nebraska	1.00	1.00	1.00
Arizona	1.00	1.00	1.00	Nevada	0.98	0.98	0.99
Arkansas	0.87	0.88	0.89	New Hampshire	0.64	0.66	0.67
California	0.00	0.00	0.00	New Jersey	0.35	0.50	0.64
Colorado	0.81	0.83	0.84	New Mexico	0.22	0.23	0.24
Connecticut	0.00	0.01	0.01	New York	0.00	0.00	0.00
Delaware	0.01	0.03	0.06	North Carolina	0.98	0.98	0.99
D.C.	0.00	0.00	0.00	North Dakota	1.00	1.00	1.00
Florida	0.36	0.37	0.38	Ohio	0.92	0.94	0.96
Georgia	0.99	0.99	0.99	Oklahoma	1.00	1.00	1.00
Hawaii	0.42	0.49	0.57	Oregon	0.05	0.06	0.07
Idaho	1.00	1.00	1.00	Pennsylvania	0.35	0.37	0.40
Illinois	0.00	0.00	0.00	Rhode Island	0.00	0.00	0.00
Indiana	1.00	1.00	1.00	South Carolina	1.00	1.00	1.00
Iowa	0.21	0.22	0.24	South Dakota	1.00	1.00	1.00
Kansas	1.00	1.00	1.00	Tennessee	1.00	1.00	1.00
Kentucky	1.00	1.00	1.00	Texas	1.00	1.00	1.00
Louisiana	1.00	1.00	1.00	Utah	1.00	1.00	1.00
Maine	0.00	0.01	0.01	Vermont	0.00	0.00	0.00
Maryland	0.00	0.00	0.00	Virginia	0.87	0.87	0.88
Massachusetts	0.00	0.00	0.00	Washington	0.14	0.15	0.16
Michigan	0.06	0.07	0.07	West Virginia	0.97	0.97	0.98
Minnesota	0.04	0.05	0.06	Wisconsin	0.05	0.06	0.07
Mississippi	0.99	0.99	1.00	Wyoming	1.00	1.00	1.00
Missouri	0.90	0.91	0.91				

* Normal vote data was unavailable Alaska, Hawaii, and the District of Columbia. We used the 2000 presidential vote outcome as the prior for these states and D.C.

The probability that a candidate wins a total of i electoral votes is given by $P(i, 51)$, and hence, the probability that the candidate wins the presidency is given by the $\sum_{i=270}^{538} P(i, 51)$. The posterior estimates of the probability of a candidate winning state k (described in Section 2) is used in place of $p^{(k)}$ in this algorithm, based on available polling data.

Figures 7, 8, and 9 report empirical probability density functions for the number of Electoral College votes won by the Republican candidate under the "no swing" scenario, "Republican swing" scenario, and "Democratic swing" scenario, respectively. For the "no swing" scenario, the probability that President Bush wins the 2004 election is 0.63, with the expected number of Electoral College votes just over 277. In the final tally, President Bush won 286 Electoral College. The probability that he would have received at least 300 Electoral College votes is 0.16. For the Republican swing scenario, the probability that Bush would win the 2004 election is 0.71, with the expected number of Electoral College votes just over 282. The probability that he would have received at least 300 Electoral College votes is 0.22. For the Democratic swing scenario, the probability that Bush would win the 2004 election is 0.54, with the expected number of Electoral College votes just over 272. In addition, the probability that he would have received at least 300 Electoral College votes is 0.11. These results suggest that the poll data accurately indicated a very tight election.

The output from our model includes three different potential outcomes to capture last minute campaign changes under the assumption that a small percentage of votes are truly swing votes and remain up for grabs even in the last hour of the campaign with the release of the critical information. Surprising last minute news releases are possible and certainly on the strategy radar of the campaigns. If the campaigns cruise into Election Day with the same messages they have touted throughout the election, then the "no swing" scenario is the most likely. If the Republicans are somehow able to mount a concerted and effective effort in the last hours to swing some voters in their direction, then the "Republican swing" distribution would be the most relevant. Similarly, the "Democratic swing" distribution would be the distribution of interest if the Democrats were able to mount an effective final hour campaign. These last minute efforts are commonplace and occur after the polls on which our estimates are based. Accordingly, supplying the three different outcomes gives us a way to account for relevant information that transpires outside of our data window.

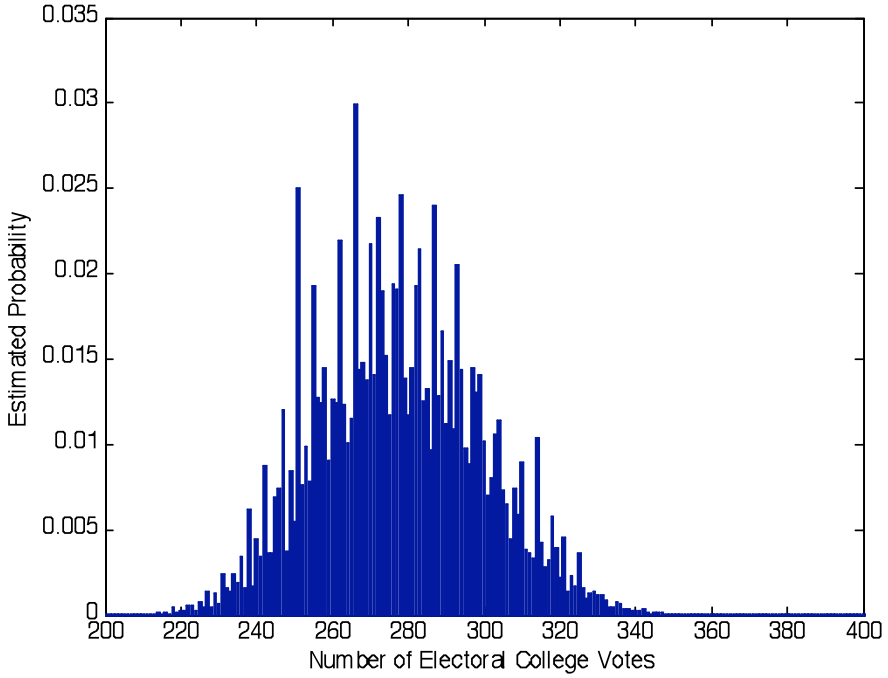

Figure 7: Estimate of Distribution of Electoral Votes in 2004 for President Bush under "No Swing" Scenario.

Unlike non-Bayesian forecasting models, our model supplies a posterior distribution for the resulting Electoral College tally. The distribution allows us to make a point estimate for the Electoral College vote, if such precision is requested, while also allowing one to examine the uncertainty attached to our estimates. Note as well that the final Electoral College tally need not match the results obtained by predicting each state separately and then simply summing up the tally from each state separately. If there are several states where the outcome is close, but leaning to Bush, we would not necessarily expect Bush to win each of these states even though a single point estimate for each state would fall in his favor. In our polls, in 12 of the states, Bush and Kerry were within 5 percentage points of each other. In addition, there were 5 more states where the candidates were 6 points apart. In this election, every state predicted to be in Bush's camp, except New Hampshire, was computed to have at least an 80% probability of voting for Bush. In contrast, there were 208 electoral votes in total from states that had at least an 80% probability of voting for Senator Kerry. That is, accordingly to our analysis, Senator Kerry had to win more of the battleground states than President Bush in order to win the election.

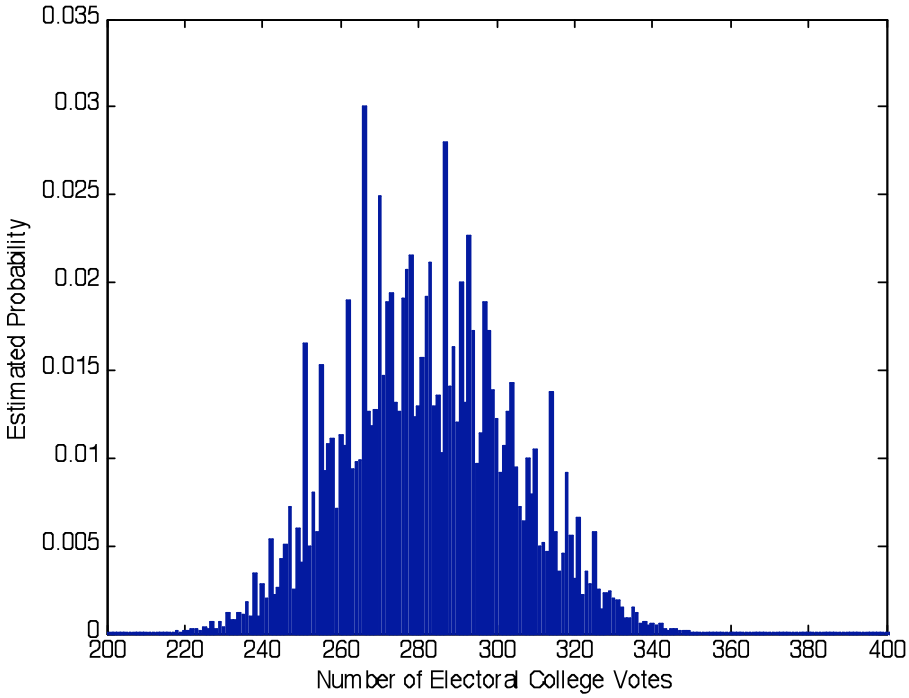

Figure 8: Estimate of Distribution of Electoral Votes in 2004 for President Bush under "Republican Swing" Scenario.

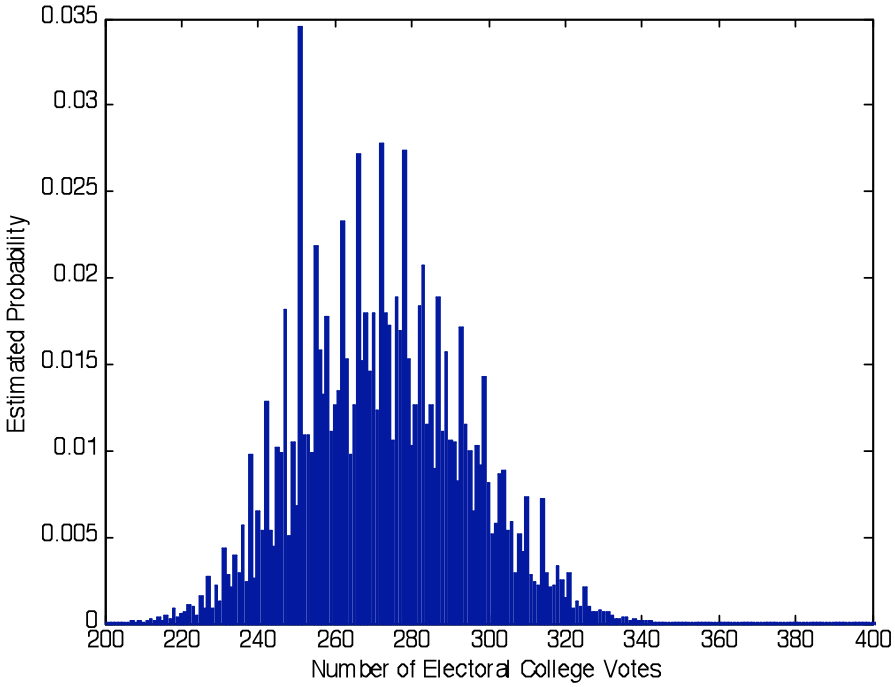

Figure 9: Estimate of Distribution of Electoral Votes in 2004 for President Bush under "Democratic Swing" Scenario.

4 Discussion

In the end, what matters is the Electoral College vote. We argue, then, that election forecasting efforts should be directed at the Electoral College vote rather than the popular vote. Attention to the Electoral College changes forecasting in significant ways. Any prediction involves many predictions, one for every state (and District of Columbia) rather than a single prediction for the national popular vote. More data are involved as well as more analysis. Our method is focused not on single state predictions or even a single national prediction, but on finding an accurate estimate of the *distribution of electoral votes* across the nation. From this distribution, one may obtain a point estimate and measures of uncertainty for the final election outcome.

We have presented a methodology for predicting the outcome of the United States presidential election that uses a Bayesian estimation approach that incorporates polling data. Our model includes the effect of third party candidates and declared undecided voters as input to a dynamic programming algorithm to build the Electoral College vote probability distribution for each candidate. There are other ways to proceed as well. For instance, we could have used the individual state predictions to generate the Electoral College distribution via Monte Carlo simulations. This is an alternative method that would also yield an Electoral College distribution. However, our approach using the dynamic programming algorithm is advantageous since it computes the probability mass function for the Electoral College votes exactly. An infinite number of Monte Carlo simulations would yield the same distribution, while a smaller number of simulations would include more variability. Plainly, there are different ways to proceed. Our point it not that we have identified the only method of obtaining an Electoral College distribution or that all methods need to be Bayesian in nature, but that the Electoral College distribution is the entity of interest in forecasting models. We have demonstrated a simple means of identifying this distribution that is substantively consistent and allows us to incorporate finer details that may affect the election outcome, especially in the types of close elections that have typified presidential elections recently. We illustrated the predictive capability of our particular estimation procedure on actual elections. Our methods are flexible and also applicable to different scenarios such as House and Senate races as long as state-by-state polling data are available for these races.

Given the recent trends in presidential elections, the methodology presented provides a rigorous approach for transforming state level polling data into presidential election forecasts. By incorporating third party candidates and undecided voters (using swing voter effects), the emotionally heated peaks and valleys of tightly contested elections are transformed into a rational probabilistic representation of the likely winner of the race. Surely, as polling data changes, so too do the predictions, and our methodology can be used at any point in the campaign cycle to compare and capture these changes and allocates appropriate weight to them so as to provide reasonable predictions even in the most volatile elections.

References

Abramowitz, Alan I. 2004. "When Good Forecasts Go Bad: The Time-for-Change Model and the 2004 Presidential Election." *PS: Political Science and Politics* 37:745–746.

Althaus, Scott L., Peter F. Nardulli and Daron R. Shaw. 2002. "Candidate Appearances in Presidential Elections." *Political Communication* 19:49–72.

Belenky, Alexander S. 2008. "The Good, the Bad, and the Ugly: Three Proposals to Introduce the Nationwide Popular Vote in U.S. Presidential Elections." *Michigan Law Review First Impressions* 106:110–116.

Bennett, Robert W. 2006. *Taming the Electoral College*. Stanford, CA: Stanford University Press.

Campbell, James E. 2004. "Forecasting the Presidential Vote in 2004: Placing Preference Polls in Context." *PS: Political Science and Politics* 37:763–767.

Campbell, James E. 2005. "Introduction - Assessments of the 2004 Presidential Vote Forecasts." *PS: Political Science and Politics* 38(1):23.

Converse, Philip. 1966. The Concept of a Normal Vote. In *Elections and the Political Order*, ed. A. Campbell, P. Converse, W.E. Miller and D.E. Stokes. Wiley.

Doherty, Brendan J. 2007. "Elections: The Politics of the Permanent Campaign: Presidential Travel and the Electoral College, 1977–2004." *Presidential Studies Quarterly* 37(4):749–773.

Edwards, George C. 2004. *Why the Electoral College is Bad for America*. New Haven, CT: Yale University Press.

Gaines, Brian J. 2001. "Popular Myths About Popular Vote-Electoral College Splits." *PS: Political Science and Politics* 34(1):70–75.

Gelman, Andrew, Jonathan N. Katz and Francis Tuerlinckx. 2002. "The Mathematics and Statistic of Voting Power." *Statistical Science* 17(4):420–435.

Goux, Darshan J. and David A. Hopkins. 2008. "Empirical Implications of Electoral College Reform." *American Politics Research* 36(6):824–843.

Grofman, Bernard and Scott L. Feld. 2005. "Thinking about the Political Impacts of the Electoral College." *Public Choice* 123:1–18.

Grofman, Bernard, T.L. Brunell and J. Campagna. 1997. "Distinguishing the Difference between Swing Ratio and Bias: the U.S. Electoral College." *Electoral Studies* 16:471–487.

Hansen, John Mark. 2008. "Equal Voice by Half Measures." *Michigan Law Review First Impressions* 106:100–105.

Hiltachk, Thomas W. 2008. "Reforming the Electoral College One State at a Time." *Michigan Law Review First Impressions* 106:90–95.

Hirsch, Sam. 2008. "Awarding Presidential Electors by Congressional District: Wrong for California, Wrong for the Nation." *Michigan Law Review First Impressions* 106:95–100.

Holbrook, Thomas M. 2004. "Good News for Bush? Economic News, Personal Finances, and the 2004 Presidential Election." *PS: Political Science and Politics* 37:759–761.

Jackman, Simon and Douglas Rivers. 2001. "State-level Election Forecasting During Election 2000 via Dynamic Bayesian Hierarchical Modeling." Working Paper, Department of Political Science, Stanford University.

Kaplan, E.H. and A. Barnett. 2003. "A New Approach to Estimating the Probability of Winning the Presidency." *Operations Research* 51(1):32–40.

Leib, Ethan J. and Eli J. Mark. 2008. "Democratic Principle and Electoral College Reform." *Michigan Law Review First Impressions* 106:105–110.

Lewis-Beck, Michael S. and Charles Tien. 2004. "Jobs and the Job of President: A Forecast for 2004." *PS: Political Science and Politics* 37:753–758.

Lewis-Beck, Michael S. and Tom W. Rice. 1983. "Localism in Presidential Elections: The Home State Advantage." *American Journal of Political Science* 27(3):548–556.

Lockerbie, Brad. 2004. "A Look to the Future: Forecasting the 2004 Presidential Election." *PS: Political Science and Politics* 37:741–743.

Longley, L.D. and J.D. Dana. 1992. "The Biases of the Electoral College in the 1990s." *Polity* 25:123–145.

Nardulli, Peter F. 2005. *Popular Efficacy in the Democratic Era: A Re-examination of Electoral Accountability in the U.S., 1828–2000.* Princeton, NJ: Princeton University Press.

Norpoth, Helmut. 2004. "From Primary to General Election: A Forecast of the Presidential Vote." *PS: Political Science and Politics* 37:737–740.

Rae, Douglas W. 1972. *Political Consequences of Electoral Laws.* New Haven: Yale University Press.

Rathbun, Daniel P. 2008. "Ideological Endowment: The Staying Power of the Electoral College and the Weaknesses of the National Popular Vote Interstate Compact." *Michigan Law Review First Impressions* 106:117–122.

Shaw, Daron R. 2006. *The Race to 270.* Chicago: University of Chicago Press.

Stokes, Donald E. 1962. "Party Loyalty and the Likelihood of Deviating Elections." *Journal of Politics* 24(4):689–702.

Tokaji, Daniel P. 2008. "An Unsafe Harbor: Recounts, Contests, and the Electoral College." *Michigan Law Review First Impressions* 106:84–88.

Whitaker, L. Paige and Thomas H. Neale. 2004. The Electoral College: An Overview and Analysis of Reform Proposals. Technical report Congressional Research Service.

Wlezien, Christopher and Robert S. Erikson. 2004. "The Fundamentals, the Polls, and the Presidential Vote." *PS: Political Science and Politics* 37:747–751.