

A NEW APPROACH TO ESTIMATING THE PROBABILITY OF WINNING THE PRESIDENCY

EDWARD H. KAPLAN

Yale School of Management, Box 208200, New Haven, Connecticut 06520-8200, edward.kaplan@yale.edu

ARNOLD BARNETT

Sloan School of Management, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, abarnett@mit.edu

As the 2000 election so vividly showed, it is Electoral College standings rather than national popular votes that determine who becomes President. But current pre-election polls focus almost exclusively on the popular vote. Here we present a method by which pollsters can achieve both point estimates and margins of error for a presidential candidate's electoral-vote total. We use data from both the 2000 and 1988 elections to illustrate the approach. Moreover, we indicate that the sample sizes needed for reliable inferences are similar to those now used in popular-vote polling.

Received September 2001; accepted June 2002.

Subject classifications: Government: elections, electoral college. Statistics: Bayesian, sampling. Probability: stochastic model applications.

Area of review: Policy Modeling and Public Sector OR.

1. INTRODUCTION

National opinion polls about U.S. presidential races generally focus on candidate standings in the popular vote. But it is the Electoral College where the election is decided and, as the year 2000 reminded us, popular and electoral vote outcomes need not be the same. Thus, the traditional polls are of limited relevance. In this paper, we consider a shift in emphasis in polling to make the Electoral College central to the reported results. We offer evidence that doing so is not intractable and that, indeed, Electoral College polls that use the same sample sizes as national popular vote tallies have a comparable margin of statistical uncertainty. Thus, the opportunity is at hand to link pre-election polls to the system under which the President is actually elected.

The rules of the Electoral College are easy to state. In the College, the number of votes of any state is equal to its number of congressional representatives (two Senators, and one House member from each congressional district). The District of Columbia gets three Electoral College votes. *All* of a state's Electoral College votes go to the candidate who wins the popular vote there (with slight exceptions in Maine and Nebraska). There are 538 electoral votes in total, so a candidate needs 270 such votes to win the presidency. The emphasis in this paper is on converting state-by-state polling results into a probability distribution for a candidate's total number of electoral votes. There have been some prior attempts to simulate such probability distributions (Erikson and Sigman 2000, Jackman and Rivers 2001), but to our knowledge, ours is the first paper to model the electoral college distribution directly.

We start our work in the next section, where we model a candidate's electoral vote distribution given estimates

of the probability that (s)he is ahead for each of the 51 states (including Washington, DC). Then, we discuss how to obtain such state-specific probability estimates, based on "snapshot" polling results from each state and a "gentle" Bayesian prior (§3). We go on to illustrate our approach with several examples that use data from the 2000 campaign (§4). In §5, we consider how setting the goal as estimating electoral vote strength might affect the manner in which polling is conducted. How should a random sample of n voters, for example, be allocated across various states? In exploring this issue with data, we turn to the 1988 presidential election, which comes closer to typifying U.S. presidential contests than other recent elections. In §6, we offer a summary and conclusions.

2. THE PROBABILITY DISTRIBUTION OF ELECTORAL COLLEGE VOTES

In this section we present a model for the probability distribution of the number of Electoral College votes for a given candidate. We first present the model, and then discuss the merits of the assumptions we have made.

We assume that all of the Electoral College votes in a given state are allocated to that candidate who wins the largest number of votes in that state. There are 51 states in the election (corresponding to the 50 actual states plus Washington, DC). We focus on a particular candidate, and let p_i denote the probability that this candidate is genuinely ahead now in the popular vote in state i . There are v_i Electoral College votes at stake in state i . We let the random variable V_i denote the actual number of Electoral College votes won by the candidate in state i , and note that the

probability distribution of V_i is given by

$$\Pr\{V_i = v\} = \begin{cases} p_i & v = v_i, \\ 1 - p_i & v = 0, \\ 0 & \text{all other values of } v, \end{cases} \quad i = 1, 2, \dots, 51. \quad (1)$$

We assume that the random variables V_i are mutually independent, as we discuss shortly.

Now define T_k as the total number of Electoral College votes received by the candidate when considering states 1 through k . Clearly,

$$T_k = \sum_{i=1}^k V_i, \quad k = 1, 2, \dots, 51. \quad (2)$$

The expected value of T_k is given by

$$E(T_k) = \sum_{i=1}^k p_i v_i, \quad k = 1, 2, \dots, 51 \quad (3)$$

while on account of the assumed independence of the random variables V_i , the variance of T_k is equal to

$$\text{Var}(T_k) = \sum_{i=1}^k p_i(1 - p_i)v_i^2, \quad k = 1, 2, \dots, 51. \quad (4)$$

Note that the total number of Electoral College votes won by the candidate is equal to T_{51} . While the mean and variance of T_{51} follow from Equations (3) and (4), we seek the probability distribution of T_{51} , for under the rules of the Electoral College system,

$$\Pr\{\text{candidate wins presidency}\} = \Pr\{T_{51} \geq 270\}. \quad (5)$$

To obtain the probability distribution of T_{51} , we first note that

$$T_{k+1} = V_{k+1} + T_k \quad \text{for } k = 1, 2, \dots, 50. \quad (6)$$

This permits us to argue recursively that for any state k and any number of Electoral College votes t ,

$$\Pr\{T_{k+1} = t\} = (1 - p_{k+1})\Pr\{T_k = t\} + p_{k+1}\Pr\{T_k = t - v_{k+1}\} \quad (7)$$

for $k = 1, 2, \dots, 50$ and $t = 0, 1, \dots, \sum_{i=1}^{k+1} v_i$. Equation (7) is simply the convolution of the distributions for the independent random variables V_{k+1} and T_k . There are only two ways that the candidate could have received exactly t Electoral College votes from the first $k + 1$ states. Either (s)he lost in state $k + 1$ but had already received exactly t votes total in states 1 through k , or (s)he won state $k + 1$ and had already received exactly $t - v_{k+1}$ votes in states 1 through k . Iterating through Equation (7) enables us to efficiently obtain the entire probability distribution for T_{51} , from which we can employ Equation (5) to calculate the probability of winning the presidency.

Some readers might wonder why we concern ourselves with deriving the ‘‘exact’’ probability distribution of Electoral College votes via Equation (7), because from Equation (2) it is clear that we are considering the sum of several (51 to be exact) random variables, so perhaps something close to a normal distribution should be expected to hold in approximation. However, some reflection shows that there is no such guarantee. While the variables V_i are independent by assumption, they are by no means identically distributed. Suppose that having considered all states except California, a normal approximation does work well for the distribution of T_{50} . However, suppose that the candidate in question also has a 50% chance of winning California’s 54 Electoral College votes. The final distribution of Electoral College votes will then be an equally-weighted mixture of two normals spaced 54 votes apart, a result which is decidedly *not* well approximated by a single normal curve. As we will demonstrate later in this paper, the example described above is not simply a theoretical curiosity.

As noted, we assume that the random variables V_i are mutually independent, which is equivalent to assuming that the events that the candidate is leading in various states are mutually independent. This assumption seems reasonable once we know the probabilities p_i that the candidate will win in each state. While it is true that there are some historical correlations in election outcomes across states, trying to exploit them to obtain the joint distribution of the V_i ’s seems far-fetched. Indeed, such correlations need not work against the independence assumption. For example, in the last 10 presidential elections, Massachusetts has voted for the Democratic candidate 9 times while Minnesota has done so 8 times. If outcomes in the two states were statistically independent, one would expect Democratic victories in both of them in $10 \times 0.9 \times 0.8 = 7.2$ times. The actual number of such double-wins was 7. Finally (and importantly), the model above yields results which are eminently reasonable when confronted with the actual results of the 1988 and 2000 presidential elections as we will discuss in §§4 and 5.

3. MODELING THE PROBABILITY OF WINNING A STATE

In this section, we present a Bayesian model for the probability p_i that the candidate wins the popular vote (and hence all v_i Electoral College votes) in state i . We assume for now that there are only two candidates contesting the election (an assumption we will defend shortly). In a two-candidate race, a particular candidate will win if (s)he garners more than half of the votes in the state. The probability that a candidate will win is thus the probability that (s)he has more than half of the votes. A Bayesian approach is natural for this problem, as it leads directly to the updated probability distribution of the fraction of voters that favor a candidate, conditional on observed polling data. Since the procedure to be described will apply to all states, we will drop state-specific subscripts in this section.

We presume that a random sample of the voting population is undertaken in each state to determine which candidate respondents prefer, were the election held at the time of the poll. Denote the sample size of the poll by n , and let X denote the (random) number of respondents in the poll that favor the candidate. As is commonly assumed in electoral polling, we take X to follow the binomial distribution with n trials and success probability Π , the fraction of voters in the state that favors the candidate. However, we treat Π itself as a random variable that, prior to polling, has a probability density $f_{\Pi}(\pi)$. After conducting a poll of size n and observing that the number of respondents that favor the candidate $X = x$, we obtain via Bayes' Rule the posterior density of Π as

$$\begin{aligned} f_{\Pi|X=x,n}(\pi|X=x,n) &= \frac{f_{\Pi}(\pi) \Pr\{X=x|n, \Pi=\pi\}}{\int_{\pi'=0}^1 f_{\Pi}(\pi') \Pr\{X=x|n, \Pi=\pi'\} d\pi'} \\ &\text{for } 0 < \pi < 1. \end{aligned} \quad (8)$$

The probability p that the candidate wins the state is then modeled as

$$\begin{aligned} p &= \Pr\{\Pi > 1/2|X=x,n\} \\ &= \int_{\pi=1/2}^1 f_{\Pi|X=x,n}(\pi|X=x,n) d\pi. \end{aligned} \quad (9)$$

As is common in Bayesian estimation problems with binomially distributed data, we assume a conjugate beta prior density function for Π (Raiffa and Schlaifer 1968):

$$f_{\Pi}(\pi) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \pi^{\alpha-1} (1-\pi)^{\beta-1} \quad \text{for } 0 < \pi < 1, \quad (10)$$

with mean and variance given by

$$E(\Pi) = \frac{\alpha}{\alpha+\beta} \quad (11)$$

and

$$\text{Var}(\Pi) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}, \quad (12)$$

as is well known. With this assumption, after observing x respondents who favor the candidate in a survey of n voters, the posterior distribution of Π is also a beta distribution, but with updated parameters $\alpha+x$ and $\beta+n-x$.

In the applications that follow in the next section, we work with “noninformative” prior distributions that set $\alpha = \beta = 1/2$. As detailed in Box and Tiao (1973), these values uniquely satisfy Jeffreys' Rule (1961), which states that the prior distribution for a single parameter (Π in our application) is (approximately) noninformative if the prior is proportional to the square root of the Fisher information (Jeffreys 1961). Noninformative priors “let the data do the talking” in that the posterior distribution of Π will be dominated by the observed data (see Equations (13) and

(14) below), as seems prudent when tracking voter sentiment over time. Other “weak” priors such as the uniform distribution would lead to virtually identical results in the applications to follow, so there is scant reason to fear that our choice for a prior has unduly influenced the analysis.

Updating the noninformative prior on the basis of (at least) moderately large sample sizes n , the posterior versions of Equations (11) and (12) specialize to

$$E(\Pi|X=x,n) = \frac{x+1/2}{n+1} \approx \frac{x}{n} \quad (13)$$

and

$$\begin{aligned} \text{Var}(\Pi|X=x,n) &= \frac{(x+1/2)(n-x+1/2)}{(n+1)^2(n+2)} \\ &\approx \frac{x}{n} \times \frac{n-x}{n} \times \frac{1}{n} \end{aligned} \quad (14)$$

which are both familiar from standard sampling theory.

Several factors complicate the polls used in determining the p_i 's. There is the obvious point that, in several recent elections, a third-party candidate was a potentially serious force (Anderson in 1980, Perot in 1992, Nader in 2000). In the actual election, such candidates rarely gain a single electoral vote, but the process by which their support “melts away” might affect the split of electoral votes between the two main candidates. (At any given time prior to the election, one could treat the supporters of third-party candidates as having abstained from the real two-candidate race, and could thereby exclude them from the estimation of the present p_i 's.) There are also undecided voters, and the chronic issue of how likely it is that a particular person canvassed will actually vote.

Such difficulties, of course, are already present in popular vote polls. Focusing the polls on electoral vote totals does not make such problems worse and, in some instances, might lessen the difficulties. Suppose, for example, that a third-party candidate's support is concentrated in a few states and that, in every one of these states, his support is far smaller than the spread between the two main candidates. Then, electoral vote calculations should properly show the candidate's irrelevance to the election outcome. Statistics about his national level of popular support might be considerably less transparent.

Before moving to actual numbers, we should mention perhaps the worst-case scenario for our approach, in which one of the two candidates has 49.9% support in every state. Then his Electoral College strength is literally zero. But if we estimated his support from modest-sized state polls with margins of error, it is likely that his various p_i 's would average around 0.5. We would presumably project his mean number of electoral votes at around 269, while the 95% probability interval for his total (covering the 2.5 percentile to the 97.5 percentile the electoral vote distribution) would come nowhere close to the true value of 0.

Such a monotonous pattern of razor-thin margins has no basis in U.S. experience, so great worry about it is excessive. But even if the situation arose, it is not clear that our

method really fails. The aim, after all, is to project who will be the next President. Given inevitable shifts over time in voter preferences, minuscule differences between candidates could often be reversed by Election Day. It might be more plausible to guess that the trailing candidate will ultimately get 269 electoral votes than that he will get none at all.

To put it another way, the p_i 's might ideally reflect both the uncertainty in the candidate's standing at this time and the uncertain relationship between his support level now and his support in the election. In extremely close elections, the fact that our method does, however informally, respond to the second form of uncertainty is arguably more a strength than a weakness.

4. EXAMPLES FROM THE 2000 PRESIDENTIAL CAMPAIGN

4.1. The American Research Group Poll of 30,600 Likely Voters

From September 5 through September 20 of 2000, the American Research Group (ARG) conducted a telephone survey of 600 likely voters in each state plus Washington, DC for a total sample size of 30,600. Those contacted were asked: "If the presidential election were being held today between George W. Bush, the Republican, Al Gore, the Democrat, Ralph Nader, from the Green Party, and Pat Buchanan, from the Reform Party, for whom would you vote—Bush, Gore, Nader, or Buchanan?" The order in which the candidates were named was rotated to remove potential sequencing biases. Using the standard $1/\sqrt{n}$ formula to deduce a 4% margin of error, the ARG reported that Gore led in 15 states (i.e., held more than a 4 percentage point lead) with a total of 204 electoral votes, that Bush led in 17 states with a total of 132 electoral votes, and that the race was "too close to call" in the remaining 19 states with 202 electoral votes.

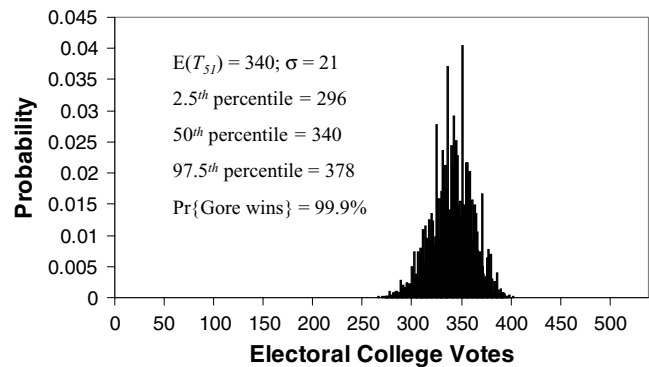
Applying our models to the state-by-state results of the ARG poll, we derived the probability distribution for the number of Electoral College votes Al Gore would receive (see Figure 1). According to this distribution, Gore would have expected 340 Electoral College votes, with a standard deviation of 21. The 2.5 percentile, median, and 97.5 percentile of the distribution were, respectively, 296, 340, and 378. And, according to this distribution, the probability that Gore would have won the presidency had the election been held at the time of the survey equals 99.99%.

This last outcome emphasizes that phrases like "too close to call" and "statistical dead heat" can encourage us to discard highly useful information. If a candidate leads 52-48 in a poll with 600 voters, the chance that he is ahead is not 50% but—even with our neutral prior—about 84%. Thus, the cumulative effect of several "too close to call" results might be an overall pattern that is highly decisive.

4.2. State-by-State Polls: PollingReport.com

To explore the sorts of trends over time in presidential preferences our method might reveal, PollingReport.com

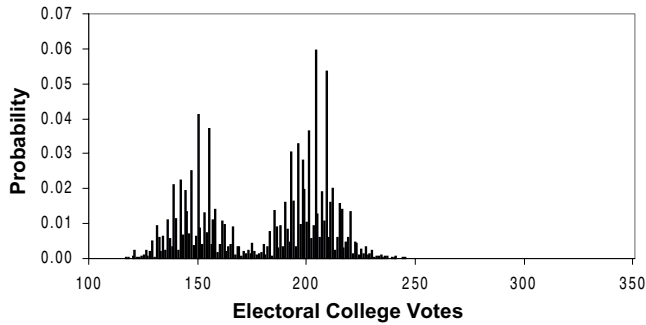
Figure 1. Probability distribution of Gore's Electoral College votes (based on ARG survey of 9/5–9/20, 2000).



granted us access to their database of state-specific voter surveys conducted over the course of the 2000 presidential campaign. This database contains polls conducted by many different polling organizations for newspapers, television stations, and political candidates. Most of these polls were telephone surveys of likely voters, though some interviewed registered voters. We aggregated the 468 different polls we reviewed into seven time periods: January through March, April through June, July, August, September, October 1–15, and October 16 on. Our intent was to create time periods long enough to cover most states, but short enough to enable us to observe trends over time. Also, given the large number of polls conducted as election day neared, we used the data collected from October 16 on to construct forecasts for direct comparison to the actual results of the election.

Unfortunately, polls were not conducted for all states in all time periods, and no polls conducted in Alaska, Kansas, South Dakota, or Washington, DC were present in the database. Our purpose here is to illustrate our methods, so we improvised as follows: For each state, we identified the earliest poll in the database, and then set the results for earlier missing time periods equal to the first-observed results. For example, the first poll conducted in Idaho that appears in the database was conducted in July, so we set the January–March and April–June Idaho results equal to what was observed in July. We filled in missing polls beyond the first available in a similar fashion, only working forwards rather than backwards. So for example, polls conducted in Arkansas were reported in April–June, July, September, and October 16 on. We set the January–March results equal to those observed in April–June, the August results equal to those observed in July, and the October 1–15 results equal to those observed in September. For Alaska, Kansas, South Dakota, and Washington, DC, we simply substituted the results of the September ARG in all time periods. We discarded undecided voters as well as those with a preference for a candidate other than George Bush or Al Gore, and employed noninformative beta priors at the start of each time period. We describe the results below.

Figure 2. Probability distribution of Gore’s Electoral College votes: Jan–Mar, 2000 (based on PollingReport.com).



4.2.1. The Electoral College Distribution as of January–March 2000. As a vivid illustration of why we rely on Equation (7) to compute the probability distribution of the number of Electoral College votes, Figure 2 shows the distribution of Al Gore’s Electoral College votes as estimated from the January–March entries in the PollingReport.com database (after adjusting for missing entries as explained above). The distribution is clearly a mixture of two subdistributions with a separation of 54 votes. California is the key. A January 15 poll and a February 2 poll, both conducted by the Public Policy Institute of California, reported 466 (466) and 456 (466) respondents favoring Bush (Gore) in the respective poll. Taken together, the posterior probability that Gore would receive more than 50% of the vote in California given these data and a noninformative prior equals 0.59, which explains the otherwise odd shape of the Electoral College distribution. Note that according to the distribution in Figure 2, Gore would have had no chance of winning the election as with probability 1, the number of Electoral College votes he would have received would have fallen below the 270 required for victory.

4.2.2. Trends in the Electoral College Distribution and the Probability of Winning the Presidency. Figure 3 reports the probability distributions of Al Gore’s Electoral

Figure 3. Electoral College distributions: Electoral votes for Gore (a: Jan–Mar; b: Apr–Jun; c: July; d: August; e: September; f: Oct 1–Oct 15; g: Oct 16 on).

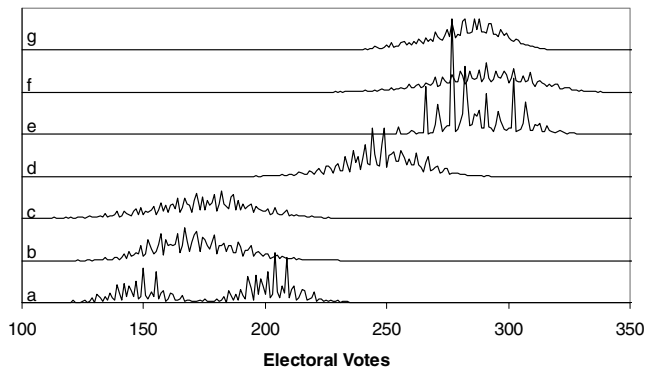
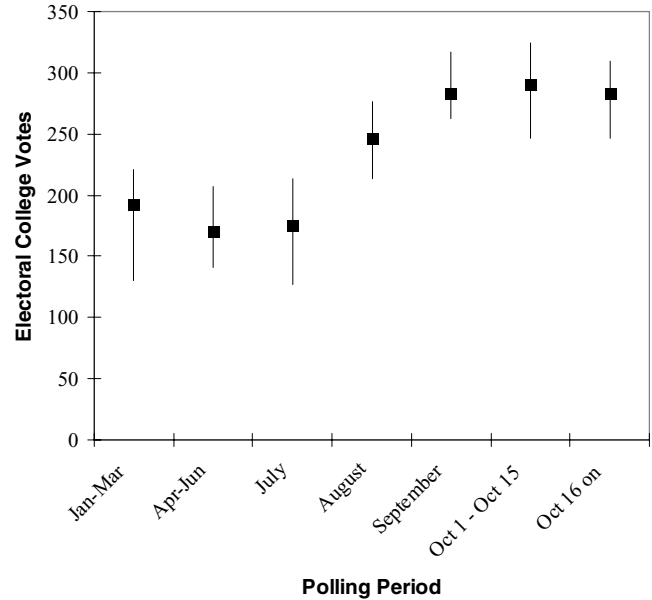


Figure 4. Gore Electoral College votes (95% probability intervals).



College votes for all seven time periods. Together these distributions suggest a growing wave of support for Gore over the course of the campaign. Figure 4 shows the median and 95% probability intervals of these probability distributions, and that the median number of Electoral College votes for Gore jumped from about 180 in the first half of 2000 to about 285 as the election neared.¹ Figure 5 portrays starkly the changes in Gore’s probability of winning the election over time as implied by the PollingReport.com data. Gore’s chance of winning went from literally nothing in the first half of 2000 to an average of about 85% in the months preceding the election.

Figure 6 compares the likelihood of Gore winning over time to the raw overall fraction favoring Gore as evidenced in the PollingReport.com database. Note how very small changes in Gore’s share of the popular vote translate to large differences in Gore’s chance of winning the election. This illustrates an important point: It is precisely in very

Figure 5. Probability that Gore wins the election.

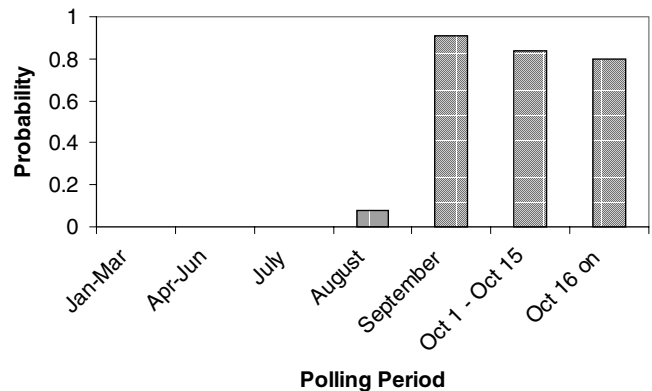
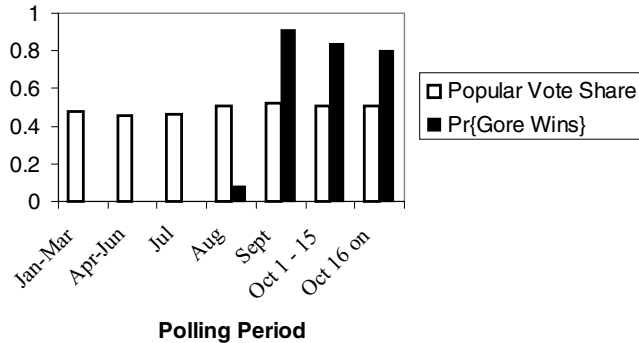


Figure 6. Gore’s popular vote share versus probability that Gore wins the election.



tight races where the idiosyncracies of the Electoral College system for choosing the President matter the most. Thus, though it is historically true that in most elections, the popular vote and Electoral College winners have been the same, it is in close elections where differences are likely to occur, and where our proposed approach might therefore yield the most valuable information. Close elections occur with some frequency: 1960, 1968, 1976, 1980, and 2000 were all tight races.

4.2.3. The Actual Election: An “Out of Sample” Experience. The polls from October 16 on provide a nearly complete data set (only Connecticut, North Dakota, Utah, and Wyoming are missing in addition to the four states identified earlier) sufficiently close to the actual election to make a comparison between the surveys and the actual election meaningful. The comparison is even more meaningful if we exclude the state of Florida and the associated ballot confusion in Palm Beach County. Excluding Florida, our model suggests that Gore could have expected 262.5 Electoral College votes with a standard deviation of 12. Gore won the rights to 267 electoral votes (though he actually received only 266, as one elector from Washington, DC refused to vote as a protest against DC’s lack of representation in the United States Congress), well within the chance bounds of our model. Comparing our individual state forecasts with actual state results, we find that the projected winner actually won in all but four states (Delaware, Missouri, Oregon, and Florida). In these states, the trailing candidate was assigned a chance of winning between 14% and 37%. If the trailing candidate is estimated to have probability q_i of winning state i , then we would expect to be “surprised” $\sum_{i=1}^{51} q_i$ times; for these data, $\sum_{i=1}^{51} q_i = 3.7$. Thus, in some sense, the reversals we saw are more consistent with our model than an absence of reversals would have been.

5. ALLOCATING A FIXED SAMPLE SIZE

Until now, we have applied our model in opportunistic fashion using whatever polls we were able to locate for analysis. Suppose instead that one wished to adopt our procedures prospectively for use in future presidential elections. An important design question to answer is, how should a

sample of fixed total size be allocated across the states? Thinking of the entire Electoral College probability distribution, we propose an approach in this section that focuses on minimizing the variability of that distribution. More precisely, we seek to minimize the *prior* expectation of the *posterior* variance of the Electoral College distribution.

Given that the distribution of a candidate’s electoral votes need not be normal, minimizing the variance is not equivalent to minimizing variability under other criteria (e.g., achieving the narrowest 95% probability interval for the candidate’s standing). But focusing on the variance is a reasonable approach that preserves tractability.

We continue to focus only on the case of two candidates, and let n_i denote the number of persons sampled in state i who have a preference for either the Republican or the Democrat. Suppose that having sampled n_i persons with such preferences in state i , we discover that x_i favor the candidate in question. From Equation (9), recall that given this result, the probability that the candidate wins state i is, in an obvious notation, modeled as

$$\begin{aligned} p_i(x_i, n_i) &= \Pr\{\Pi_i > 1/2 | X_i = x_i, n_i\} \\ &= \int_{\pi=1/2}^1 f_{\Pi_i | X_i=x_i, n_i}(\pi | X_i = x_i, n_i) d\pi. \end{aligned} \quad (15)$$

The posterior variance of the number of Electoral College votes in state i given that x_i respondents favor the candidate is then given by

$$\sigma_i^2(x_i, n_i) = p_i(x_i, n_i) \times (1 - p_i(x_i, n_i)) \times v_i^2. \quad (16)$$

Now, recall that the distribution of the number surveyed in state i that favor the candidate, X_i , is distributed binomially with parameters n_i and Π_i where Π_i itself has a prior density $f_{\Pi_i}(\pi)$. The marginal prior distribution of X_i given the sample size n_i is then the mixture of binomials given by

$$\Pr\{X_i = x_i | n_i\} = \int_0^1 \Pr\{X_i = x_i | n_i, \Pi_i = \pi\} f_{\Pi_i}(\pi) d\pi. \quad (17)$$

The *prior* expectation of the *posterior* variance of the number of Electoral College votes for the candidate in state i is thus given by

$$\bar{\sigma}_i^2(n_i) = \sum_{x_i=0}^{n_i} \sigma_i^2(x_i, n_i) \times \Pr\{X_i = x_i | n_i\}. \quad (18)$$

Our proposal for allocating a sample of n two-party respondents is to solve the following knapsack problem:

$$\min_{n_1, n_2, \dots, n_{51}} \sum_{i=1}^{51} \bar{\sigma}_i^2(n_i) \quad (19)$$

subject to the constraints

$$\sum_{i=1}^{51} n_i = n \quad (20)$$

and

$$n_i \geq 0 \text{ and integer for } i = 1, 2, \dots, 51. \quad (21)$$

Solving this sample allocation problem is aided greatly by the observation that the functions $\bar{\sigma}_i^2(n_i)$ are decreasing and convex. This means that a Marginal Allocation (or *greedy*) Algorithm will provide the optimal solution. In such a scheme, each sample is allocated to that state with the largest *marginal* reduction in uncertainty. More formally, define

$$\Delta_i(n_i) = \bar{\sigma}_i^2(n_i) - \bar{\sigma}_i^2(n_i + 1) \quad (22)$$

and initially set $n_i = 0$ for $i = 1, 2, \dots, 51$. Let m be a counter that will run from 1 through n . Then the algorithm runs as follows:

Marginal Allocation Algorithm

For: $m = 1$ to n

Define: $i^* = \arg \max_{1 \leq i \leq 51} \Delta_i(n_i)$

Set: $n_{i^*} \leftarrow n_{i^*} + 1$

Next: m

Ties can be broken arbitrarily in defining i^* in the algorithm above. We next present some examples illustrating the use of this algorithm.

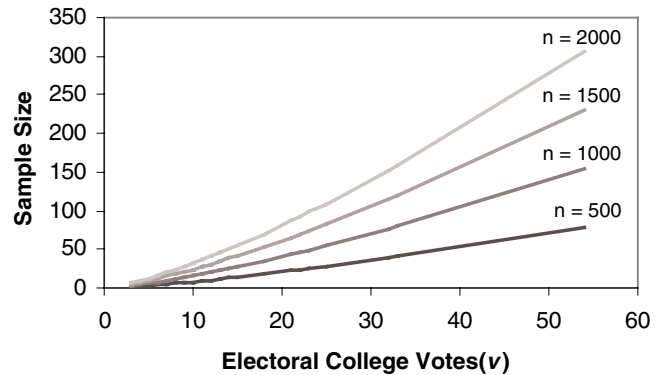
5.1. The American Research Group Poll Revisited

Recall the 30,600 person state-by-state ARG poll. Of the 30,600 persons sampled, 26,076 expressed a preference for George Bush or Al Gore. Assuming noninformative priors, had these 26,076 observations been allocated in accord with the Marginal Allocation Algorithm, the *prior* expectation of the *posterior* variance in the Electoral College distribution would equal 80.66, which implies a pseudo-standard deviation of 8.98, or about 1.7% of the electoral-vote total. Suppose that the ARG faced a factor-of-ten budget slash and wished to optimally allocate 2,000 Bush/Gore samples across the states. Again, assuming noninformative priors, the Marginal Allocation Algorithm achieves a variance of 197.5, with an associated standard deviation of 14.05, 2.6% of the number of electoral votes.

5.2. Optimal Sample Size as a Function of the Number of Electoral College Votes: Noninformative Priors

Continuing with the use of noninformative priors, Figure 7 reports how the optimal sample allocations determined by the Marginal Allocation Algorithm vary with the number of Electoral College votes at stake for total sample sizes of 500, 1,000, 1,500, and 2,000. The optimal sample sizes increase in convex fashion with the number of Electoral College votes. This is not surprising when one realizes that

Figure 7. Optimal state-specific sample sizes.



the prior expectation of the posterior variance of the number of electoral votes in state i , $\bar{\sigma}_i^2(n_i)$, is proportional to v_i^2 . In trying to estimate the distribution of Electoral College votes, getting California right is clearly more important than North or South Dakota!

5.3. Optimal Sample Size with “Last Minute” Informative Priors

To illustrate how one can proceed with a prior that incorporates the latest polling data, imagine the following scenario: A political consultant, aware of the survey results from a family of polls taken prior to the election, decides to use the results of these polls to form “last minute” informative priors for allocating a sample with a target total of 500 Bush/Gore respondents. We assume that the consultant has access to prior polls with sample size m_i in state i . Having observed a fraction g_i for Gore in the prior sample, she updates her noninformative prior for Π_i , as described in §3. As a numerical example, we consider the proportions favoring Gore reported in the last wave of PollingReport.com data, but we reduce the prior sample size m_i to 10% of the total number of Bush/Gore respondents in the data to better reflect the information an individual consultant might have at her disposal (because many of the polls in the data were private at the time).

The results are shown in Table 1. Not surprisingly, in most states the priors are sufficiently strong to obviate the need for further sampling. Sampling would only progress in eight states under this scenario. Of the 500 samples sought, 279 would have been allocated to Florida, a state with both a tight race and a large number (25) of Electoral College votes at stake. The situation in California was more certain, but with 54 Electoral College votes, it would still be prudent to collect an additional 77 samples there. Table 1 also shows that it might not always be worth it to take the results of the Marginal Allocation Algorithm literally—is it really that important to obtain two additional samples from Arizona, three from New Mexico, and 6 from Colorado? Including these samples leads to an expected posterior variance of 793.4, while excluding them raises this to 794.4, a trivial change.

Table 1. Optimal sample sizes (informed priors, $n = 500$).

State	Optimal Sample Size
Arizona	2
California	77
Colorado	6
Florida	279
Georgia	25
New Mexico	3
North Carolina	85
Pennsylvania	23

5.4. An Experiment with the 1988 Election

We now turn to the 1988 presidential election to illustrate a complete “start to finish” application of our approach. That year, George Bush Sr. defeated Michael Dukakis, gaining 426 of the 538 electoral votes and 54% of the popular vote. The 1988 election was the most “normal” of recent contests: In 1984 and 1996, there were landslides, while in 1992 Ross Perot launched the strongest third-party candidacy in recent memory. The premise of the experiment is that a pollster can canvass n voters across the nation on the eve of the election, and that polling in individual states will reflect without bias how the state will vote on Election Day. Thus, for example, Dukakis got 43% of the vote in New Jersey; we therefore assume that any person polled in that state would have a 0.43 chance of supporting Dukakis and a 0.57 chance of supporting Bush.

We allowed the total number of voters polled to vary, from an average of 10 per state (510 voters in all) to 150 per state (7,650 in total). We also allowed the allocation of samples across the states to vary according to four different schemes:

(1) *Homogeneous*—The same sample size in all 51 states.

(2) *Simple Marginal Allocation*—Samples are allocated across states according to their electoral vote totals, so as to minimize posterior variance (as described in §5).

(3) *Homogeneous/Marginal Hybrid*—First, half the sampling is divided equally among the 51 states. Then, the initial results are used to achieve revised distributions of state vote splits. These distributions are used in the Marginal Allocation Algorithm to set the numbers sampled in each state among the remaining $n/2$ people polled.

(4) *Marginal/Marginal Hybrid*—Here the first $n/2$ samples are divided among states according to the Marginal Allocation Algorithm. Then, the results generate a revised distribution of state vote splits, which is then used in a second run of the Marginal Allocation Algorithm for the remaining $n/2$ samples.

In all we considered six values of n and four sample-allocation rules, yielding a total of 24 results. The basic procedure in each instance was:

(i) Take a sample of size m in each state, and assume that the observed number for Dukakis would be binomially distributed with m trials and success probability q , where

$q =$ Dukakis’ actual share of the vote on Election Day in 1988. (As noted, m might partially depend on outcomes in the first half of the sampling.)

(ii) Use the observed sampling result and the Bayesian prior as in §3 to estimate the probability that Dukakis would carry each state (i.e., its p_i).

(iii) Go back to the original recursion of Equation (7) in §2 to find the probability distribution for Dukakis’ total number of electoral votes.

Table 2 summarizes the results of the experiment. In each instance, we present the 2.5 percentile, the median, and the 97.5 percentile of the estimated number of Dukakis electoral votes (recall that the 95% probability interval for the total is the range from the 2.5 percentile to the 97.5 percentile), along with the probability of a Dukakis win. Dukakis actually won 112 electoral votes.

The results are generally encouraging, though not without some disappointments. In all 24 scenarios, Dukakis was accurately assigned a small chance of winning the presidency. The probability was about 5% when 2,040 people were polled (mean of 40 per state), and fell to about 1 in 1,000 when an average of 150 people were polled per state. But there was a systematic tendency to overestimate Dukakis’ electoral-vote total. Although his actual showing fell within the 95% probability interval once n reached 2,040, he was below the median even for $n = 7,650$.

We suggested the reason for this latter tendency earlier, where we noted that our method might overestimate the chance a candidate will win a state when he is only modestly behind there. The Bayesian prior starts by assigning $p_i = 0.5$, and only gradually moves from that assessment as polling data come in. If, for example, only 20 voters are selected in a state in which the true Dukakis vote share is 0.44, then even an “accurate” vote split of 9 for Dukakis and 11 for Bush would only push p_i down to 0.33. The problem would ultimately disappear once sample sizes become huge; as the calculation suggests, however, huge might mean considerably larger than typical national polls.

Of course, large state-by-state polls routinely take place for various reasons, as we saw in 2000. Thus, if the Electoral College poll could sensibly “piggy back” on individual-state surveys already undertaken, then achieving very high accuracy need not entail inordinate extra expense.

Perhaps counterintuitively, the homogenous/marginal allocation method fared better in the experiment than the marginal/marginal method. The danger to the latter approach might be its predisposition to favor the larger states in the first half of canvassing, even though uncertainty about outcomes might be greater in smaller states. Thus, samples wasted on “belaboring the obvious” in large states might preclude enough sampling in close small states to provide reliable information there.

6. CONCLUSION

National voter surveys estimate the likely popular vote, but such polls do not directly estimate the probability of

Table 2. Results from the 1988 Dukakis Electoral College distribution experiment (2.5 percentile, Median, 97.5 percentile, and $\Pr\{\text{Dukakis Wins}\}$).

Sample Size	Homogeneous	Marginal Allocation	Homogeneous/Marginal Hybrid	Marginal/Marginal Hybrid
	2.5% / 50% / 97.5% / $\Pr\{\text{Win}\}$	2.5% / 50% / 97.5% / $\Pr\{\text{Win}\}$	2.5% / 50% / 97.5% / $\Pr\{\text{Win}\}$	2.5% / 50% / 97.5% / $\Pr\{\text{Win}\}$
510	136 / 229 / 325 / 0.213	135 / 225 / 321 / 0.1897	130 / 221 / 317 / 0.165	136 / 226 / 322 / 0.1934
1020	124 / 214 / 310 / 0.134	123 / 210 / 304 / 0.1121	117 / 203 / 297 / 0.0868	120 / 205 / 299 / 0.0917
2040	109 / 196 / 289 / 0.0635	109 / 191 / 282 / 0.0463	102 / 183 / 274 / 0.0309	107 / 186 / 276 / 0.0342
4080	93 / 176 / 289 / 0.0196	95 / 170 / 255 / 0.0107	88 / 162 / 247 / 0.006	95 / 166 / 249 / 0.0069
5100	89 / 169 / 257 / 0.012	91 / 163 / 246 / 0.0056	84 / 156 / 238 / 0.003	92 / 159 / 239 / 0.0034
7650	81 / 158 / 243 / 0.0041	85 / 152 / 230 / 0.0013	79 / 145 / 222 / 0.0006	87 / 148 / 222 / 0.0007

winning the presidency. We have presented a model for determining the probability distribution of the number of Electoral College votes that a candidate will win. From this distribution, one can compute directly the probability of winning the presidency. We have shown how to derive the necessary parameters for this model via Bayesian analysis of state-by-state voter surveys, and we have illustrated our methods with polls conducted during the 2000 presidential campaign. We have also examined the problem of how to efficiently allocate a sample across the states, developed a very simple Marginal Allocation Algorithm for solving this problem, and illustrated via recourse to the 2000 and 1988 contests.

It remains to apply these methods prospectively in a future presidential election. In particular, our model seems highly amenable for use with exit polling data on election night itself. There might also be applications in the realm of campaign strategy. Our model could offer an approach to estimating the incremental gain in the probability of winning the presidency that would result from particular gains in popular vote shares in individual states. Such a “value added” criterion might prove useful in allocating campaign effort over time.

ENDNOTE

1. The estimate of Gore’s strength based on the September PollingReport.com data is lower than that in the concurrent AMR poll (with a median of 285 electoral votes versus 340). The difference reflects methodological divergences between the two sampling procedures that, while important, are not germane to the main point of this paper.

REFERENCES

- Box, G. E. P., G. C. Tiao. 1973. *Bayesian Inference in Statistical Analysis*. Addison-Wesley, Reading, MA.
- Erikson, R. S., K. Sigman. 2000. A simple stochastic model for close U.S. presidential elections. Working paper, Department of Political Science, Columbia University, New York.
- Jackman, S., D. Rivers. 2001. State-level election forecasting during Election 2000 via dynamic Bayesian hierarchical modeling. Working paper, Department of Political Science, Stanford University, Stanford, CA.
- Jeffreys, H. 1961. *Theory of Probability*, 3rd ed. Clarendon Press, Oxford, U.K.
- Raiffa, H., R. Schlaifer. 1968. *Applied Statistical Decision Theory*. MIT Press, Cambridge, MA.