Public Choice (2006) **127:** 207–223
DOI: 10.1007/s11127-006-1259-3

Using state polls to forecast U.S. Presidential election outcomes

SOUREN SOUMBATIANTS[1], HENRY W. CHAPPELL, JR.[2,*]
& ERIC JOHNSON[3]

[1]*Department of Business, Lincoln Memorial University, Harrogate, TN 37752;* [2]*Department of Economics, University of South Carolina, Columbia, SC 29208;* [3]*Department of Economics, Kent State University, Kent, OH 44242*
*(*Author for correspondence: e-mail: chappell@moore.sc.edu)*

Accepted: 22 July 2005

Abstract. This paper uses pre-election polls to forecast U.S. Presidential election outcomes in the states and the Electoral College. The approach is notable in three ways. First, we employ state-level polls to predict voting outcomes in the states; second, we associate probabilities with alternative election outcomes, and third, we identify states most likely to be pivotal in the Electoral College. Using information available on the day before the election, we estimated that the probability of a Republican victory in the Electoral College in the 2004 election was 47.27%.

In the midst of U.S. Presidential election campaigns, polls are frequently conducted to assess candidates' chances for victory. Although poll results are widely reported in the popular press, they are rarely accompanied by anything more than a superficial analysis of the candidates' victory probabilities. In this paper we investigate the use of poll data in forecasting presidential election outcomes. Our approach is notable in three ways. First, our raw input data consists of state-specific polls, which are used in making forecasts of state-specific election results and, subsequently, Electoral College outcomes. Second, in addition to producing forecasts, we are able to associate probabilities with alternative outcomes. In doing so, we account for both national and state-specific "shocks" that cause election outcomes to differ from the polls that precede them. Third, our forecasting method can also be used to assess the likelihood that individual states might be pivotal in the Electoral College.

To illustrate the use of our method, we produce "forecasts" of the 2004 Presidential election based on pre-election data. In addition to point forecasts, we report probabilities associated with various popular and Electoral College voting outcomes at alternative time horizons before the election.

The methods we develop should be of interest to anyone tracking the "horserace" aspects of Presidential election campaigns, but they also have potential utility for descriptive analyses of political behavior. As elections approach, polls provide information that can guide campaign decisions. To allocate campaign effort optimally, candidates must assess where additional campaigning is likely to have the highest marginal payoffs in terms of election win probabilities. For this purpose, candidates need to know which states are most likely to be both closely contested and pivotal in the Electoral College setting. Our method correctly identifies Florida and Ohio as key states using data available prior to the 2004 election.

1. The Literature on U.S. Presidential Election Forecasting

Two distinct areas of the academic literature on elections are related to Presidential election forecasting. The first area includes studies that describe the estimation of "vote functions," regression equations that that attempt to explain aggregate election outcomes over time using explanatory variables that are fundamental determinants of voter behavior. In pioneering studies (Kramer, 1971; Fair, 1978) found that rapid economic growth in the period preceding a presidential election led to higher vote totals for the incumbent party candidate.[1] Erikson (1989), Fair (1982, 1988, 1996) and Hibbs (2000) have also estimated vote functions that confirm the importance of economic conditions in voters' decisions.

A second branch of the literature focuses more narrowly on forecasting. Although vote functions can be used for forecasting, they often expressly exclude some explanatory variables with predictive potential. For example, Fair's vote function excluded pre-election polls as explanatory variables. In 1992, his model predicted an easy win by incumbent George Bush, even though pre-election polls made it clear that this outcome was unlikely. Adding a pre-election poll as an explanatory variable in Fair's vote function would presumably have improved its forecast, but would not have provided any fundamental explanation for voters' dissatisfaction with the incumbent. This paper falls into the branch of the literature that makes forecasting the primary objective; its analysis relies heavily on poll data in making forecasts.

A number of previous studies have made use of polls in attempting to produce accurate election forecasts. Among the simplest are models proposed by Campbell and Wink (1990) and Lewis-Beck and Rice (1992) to explain national vote shares. Campbell and Wink's model includes just two predictor variables, a trial-heat poll and second quarter GDP growth in the year of the election. Lewis-Beck and Rice use a similar specification, but add variables capturing recent partisan trends. Both Campbell and Wink and Lewis-Beck and Rice report that their models produce accurate forecasts 60 days prior to the election. Specifically, out-of-sample forecast errors for the Campbell-Wink model had a mean absolute error of just 1.3% over the 1948–1992 period (Campbell, 1996). Other contributions in this genre include (Abramowitz, 1992, 1996; Brown & Chappell, 1999; Erikson & Wlezien, 1996; Holbrook, 1996; Lewis-Beck & Tien, 1996).

Because election winners are determined in the Electoral College, some forecasters have predicted vote shares and winners in each of the states and then aggregated Electoral votes to predict an overall winner. Rosenstone (1983), Holbrook (1991) and Campbell (1992) have each taken this approach, examining election outcomes across both states and time, using a mixture of national- and state-level variables as explanatory variables. Campbell's (1992) model is especially notable; it rivals Holbrook in terms of parsimony of specification and Rosenstone in terms of forecast accuracy. Campbell's model includes an early-September national trial-heat poll and second quarter GDP growth as explanatory variables. In addition, the model includes prior state deviations from national voting outcomes and other state- and region-specific indicators of partisan strength and economic performance.

In this paper, we follow Campbell in developing a model to explain election outcomes across states and time. Our model differs from that of Campbell in several ways. First, we employ state-level poll data to forecast state-specific outcomes. Comprehensive state-level polling data has become abundant only in recent years and has not previously been widely used in forecasting.[2] Second, we specify an error structure for the model that includes an election-specific shock (shared across states) and shocks that are state-specific. As Crain, Messenheimer, and Tollison (1993) have shown, the specification of the model's error term may be an important issue when assessing probabilities; specifically, they note that variation in the shared national component of an election's error term tends to account for much of the variation in electoral outcomes.[3] Third, our forecasting model does *not* include a measure of economic growth as an explanatory variable. This is in part a practical necessity; in forecasting we employ a sample that covers only four past elections (providing just four observations on pre-election GDP growth).[4] To the extent that poll responses already incorporate voters' evaluations of economic conditions, the impacts on election outcomes will be accounted for in the model.

To illustrate the use of our procedure, we employ data from the 1988, 1992, 1996, and 2000 elections to estimate models that relate poll results to election outcomes in the states. We then use the estimated models to provide "forecasts" for the 2004 presidential election. The closeness of the 2000 and 2004 elections makes the calculation of several outcome probabilities particularly interesting. Specifically, we estimate win probabilities for each of the candidates, and also the probabilities associated with alternative Electoral College outcomes (including the probabilities that each candidate might win the election while losing the popular vote).

Before proceeding, we should note that in recent years, another source of forecast information has been provided by the Iowa Electronic Markets. In these markets, traders exchange securities whose values reflect market perceptions about upcoming presidential election outcomes. However, the Iowa markets have been structured to provide a forecast of the winner of the two-party national vote, rather than the outcome in the Electoral College.[5]

2. The Forecasting Model

We propose the following econometric model to relate state pre-election polls to state election outcomes:

$$V_{it} = \alpha_i + \gamma P_{it} + u_t + \varepsilon_{it}, \quad t = 1, \ldots, 5; \quad i = 1, 2, \ldots, 51. \qquad (1)$$

In this equation, t indexes the elections in our sample (for 1988 through 2004) while i indexes the U.S. states (including the District of Columbia). The dependent variable, V_{it}, indicates the share of the two-party vote going to the Democratic candidate in state i in election t. The single explanatory variable, P_{it}, is the share of the two-party vote intention favoring the Democratic candidate in a single pre-election poll for state i.[6] Parameters α_i are state-specific intercepts, while u_t and ε_{it} are independent random errors with zero means and constant variances. In estimation, the α_i are treated as fixed state effects, while the u_t are random election-specific effects.

The specification of the composite error reflects both the nature of polling technology and the behavior of voters in processing information as campaigns proceed. Polls do not perfectly predict election outcomes because of sampling error (not all voters are polled), because poll samples are not perfectly representative of the voter population, and because poll respondents are not perfectly truthful. In addition, polls are administered in advance of elections. Even if polls provided a perfect representation of voter sentiment on the day administered, voters could change their minds in the period between the polling date and the election. Sampling errors and some errors resulting from imperfections in polling technologies may be state- and poll-specific, and therefore independent across states. However, other polling error components are likely to be shared across states.

For example, suppose that a poll is administered 30 days before the election and that a candidate debate will occur 15 days before the election. The debate might produce a clearly superior performance by one candidate, perhaps the Democrat, who then benefits with gains in vote share in each state on election day. Polls administered at the 30-day pre-election horizon will not capture the impact of the debate; in Equation (1) the debate effect would be captured in u_t, the error component shared across states.

The state poll data employed in our study were collected and published by the *Hotline Weekly* for the years 1988, 1992, and 1996. Poll data for 2000 and 2004 were published by the *National Journal* on its web site.[7] In each year, the published information was intended to be a comprehensive collection of reputable polling data for the states.

For most states, there were multiple pre-election polls, with polls produced by different polling organizations and collected at different times prior to the election. In our econometric model, P_{it} is the Democratic share of the two-party vote intention in the most recent state poll available at a given pre-election forecasting horizon. Suppose that a forecast is to be generated 15 days prior to the election. In order to make such a forecast, we would estimate equation (1) using sample data that includes the most recent polling outcome for each state, *subject* to the restriction that each poll be completed at least 15 days prior to the election.[8]

3. Estimation and Forecasting

We have estimated equation (1) at two pre-election time horizons (1 day and 15 days before the election). The results of these estimations using election data from 1988 to 2004 are presented in Table 1. In both estimations, the coefficient of the poll variable, P_{it}, is positive, significantly different from zero, and significantly less than one.

We now describe how the estimated empirical model can be used for forecasting and for generating probability assessments of candidates' chances. Our forecasts are based on Monte Carlo simulation methods that could easily be adapted to alternative econometric models of electoral outcomes. Here, however, we continue to make use of the simple model described in the preceding section.

The general approach we take involves using our estimated Equation (1) to produce "forecasts" for elections in years 1988, 1992, 1996 and 2000. We then calculate observed forecast errors for each state in each of those elections. Next, these empirical forecast errors are used to simulate outcomes for the year 2004 election. We carry out multiple simulations in a Monte Carlo framework, so that the frequencies of observed outcomes can be interpreted as probabilities.

The use of empirical forecast error distributions in the simulation stage is a useful strategy because it leads to forecasts and probability inferences that will be robust across a variety of possible misspecifications of the original estimating equation. For example, at the estimation stage we might incorrectly assume that error distributions are normal or that heteroscedasticity is absent. Analytically derived forecast error distributions would then also be incorrect, and simulations based on them would be misleading. Generally, in calculating analytical forecast errors one assumes that the original model is the correct one – forecast errors occur only because of randomness in nature. But empirical forecast errors reflect errors in specification as well as random effects that are a part of the model. Because we employ empirical distributions of forecast errors in our simulations, our probability estimates should be robust to a variety of imperfections of the econometric model.

We next describe the specific steps involved in our forecasting simulations. Our task is to use the historical data provided by polling and election data from years 1988 through 2000 in order to forecast the election for year 2004. First

Table 1. Regression coefficients summary

	1 day before election		15 days before election	
Variable	Coefficient	Standard error	Coefficient	Standard error
Poll	0.19	0.03	0.18	0.03
District of Columbia	72.46	3.37	73.96	3.39
Alabama	33.28	2.33	33.64	2.39
Alaska	29.75	2.23	30.17	2.27
Arizona	35.78	2.46	36.66	2.47
Arkansas	40.16	2.50	41.13	2.50
California	44.30	2.52	44.88	2.57
Colorado	38.21	2.43	38.82	2.47
Connecticut	44.03	2.57	44.63	2.63
Delaware	43.01	2.49	44.04	2.48
Florida	37.34	2.46	38.43	2.44
Georgia	35.79	2.34	36.50	2.36
Hawaii	47.26	2.51	47.75	2.57
Idaho	26.41	2.24	27.23	2.23
Illinois	44.67	2.53	45.40	2.57
Indiana	33.75	2.33	34.39	2.35
Iowa	42.76	2.43	43.84	2.42
Kansas	32.46	2.28	33.06	2.31
Kentucky	35.99	2.40	37.13	2.37
Louisiana	38.78	2.40	39.40	2.43
Maine	43.80	2.44	44.09	2.53
Maryland	45.33	2.51	46.03	2.55
Massachusetts	50.15	2.69	51.32	2.69
Michigan	41.88	2.47	42.57	2.51
Minnesota	43.83	2.52	45.10	2.49
Mississippi	33.20	2.36	34.00	2.36
Missouri	40.41	2.44	41.39	2.43
Montana	35.25	2.35	36.20	2.33
Nebraska	28.64	2.26	29.58	2.24
Nevada	37.43	2.45	38.27	2.47
New Hampshire	37.87	2.51	38.81	2.52
New Jersey	42.20	2.53	43.02	2.55
New Mexico	40.63	2.49	41.61	2.49
New York	48.73	2.56	49.41	2.61
North Carolina	35.51	2.39	36.29	2.40
North Dakota	31.22	2.35	32.02	2.36

(*Continued on next page*)

Table 1. (*Continued*)

Variable	1 day before election Coefficient	1 day before election Standard error	15 days before election Coefficient	15 days before election Standard error
Ohio	39.43	2.40	39.57	2.50
Oklahoma	32.23	2.29	33.01	2.29
Oregon	42.72	2.49	43.49	2.51
Pennsylvania	42.26	2.46	42.75	2.52
Rhode Island	50.59	2.66	51.35	2.71
South Carolina	33.46	2.34	34.23	2.34
South Dakota	34.89	2.33	35.77	2.32
Tennessee	37.38	2.41	38.11	2.44
Texas	34.91	2.25	35.37	2.28
Utah	24.23	2.22	24.59	2.27
Vermont	46.24	2.55	47.46	2.52
Virginia	36.17	2.35	36.38	2.43
Washington	44.61	2.42	44.43	2.56
West Virginia	41.43	2.46	42.29	2.47
Wisconsin	42.05	2.43	42.16	2.54
Wyoming	29.74	2.15	30.48	2.15
Standard Deviation of u_t	2.92	0.94	2.86	0.92
Standard Deviation of ε_{it}	2.68	0.12	2.72	0.12

recall our estimating equation (1):

$$V_{it} = \alpha_i + \gamma P_{it} + u_t + \varepsilon_{it}, \quad t = 1, \ldots, T; \quad i = 1, 2, \ldots, 51. \qquad (1)$$

1. Estimate Equation (1) using poll data for a given horizon before the election (e.g., polls available 15 days prior to the election). Initially use data for election years $t = 1, \ldots, 3$ (1988, 1992, and 1996) in estimation, omitting data for year $t = 4$ (2000). Save the estimated coefficients, designating these as $\hat{\alpha}_{i4}$ (for $i = 1, \ldots, 51$) and $\hat{\gamma}_4$. The subscript "4" indicates the omitted year).
2. Use the estimates obtained in step 1 to calculate forecast values for voting outcomes for each state in election $t = 4$:[9]

$$\hat{V}_{i4} = \hat{\alpha}_{i4} + \hat{\gamma}_4 P_{it}$$

3. Calculate forecast errors for each state for the year $t = 4$ election. The forecast error is the actual vote share outcome minus the forecasted vote

share:

$$\hat{e}_{i4} = V_{i4} - \hat{V}_{i4}.$$

4. Repeat steps 1–3, as modified below.

   (a) Repeat steps 1–3, but use data for elections $t = 1, 2, 4$ in estimation, omitting data for election $t = 3$. Forecast outcomes for election $t = 3$ and calculate forecast errors $\hat{e}_{i3} = V_{i3} - \hat{V}_{i3}$. Record estimated coefficients $\hat{\alpha}_{i3}$ and $\hat{\gamma}_3$.
   (b) Repeat steps 1–3, but this time use data for elections $t = 1, 3, 4$ in estimation, omitting data for election $t = 2$. Forecast outcomes for election $t = 2$ and calculate forecast errors $\hat{e}_{i2} = V_{i2} - \hat{V}_{i2}$. Record estimated coefficients $\hat{\alpha}_{i2}$ and $\hat{\gamma}_2$.
   (c) Repeat steps 1–3, but this time use data for elections $t = 2, 3, 4$ in estimation, omitting data for election $t = 1$. Forecast outcomes for election $t = 1$ and calculate forecast errors $\hat{e}_{i1} = V_{i1} - \hat{V}_{i1}$. Record estimated coefficients $\hat{\alpha}_{i1}$ and $\hat{\gamma}_1$.

   We now have four sets of forecast errors, $\hat{e}_{i1}$, $\hat{e}_{i2}$, $\hat{e}_{i3}$, and $\hat{e}_{i4}$, one set for each election from 1988 to 2000. We also have four sets of estimated parameters: $\{\hat{\alpha}_{i1}, \hat{\gamma}_1\}$, $\{\hat{\alpha}_{i2}, \hat{\gamma}_2\}$, $\{\hat{\alpha}_{i3}, \hat{\gamma}_3\}$, and $\{\hat{\alpha}_{i4}, \hat{\gamma}_4\}$. Remaining steps will use these forecast errors and parameters in a Monte Carlo simulation to produce candidate win probabilities for election $t = 5$ (2004).
5. Randomly draw a set of parameters from the four sets produced above; i.e., select randomly from $\{\hat{\alpha}_{i1}, \hat{\gamma}_1\}$, $\{\hat{\alpha}_{i2}, \hat{\gamma}_2\}$, $\{\hat{\alpha}_{i3}, \hat{\gamma}_3\}$, and $\{\hat{\alpha}_{i4}, \hat{\gamma}_4\}$. Designate the selection as $\{\hat{\alpha}_{is}, \hat{\gamma}_s\}$
6. Randomly draw a year-set of forecast errors from the four sets produced above; i.e., randomly select randomly from $\hat{e}_{i1}$, $\hat{e}_{i2}$, $\hat{e}_{i3}$, and $\hat{e}_{i4}$. Designate the selection as $\hat{e}_{iq}$.
7. In the preceding step, $\hat{e}_{iq}$ defines 51 state-level forecast errors. Now for each state i, randomly choose an element j from these 51 forecast errors. Define the selection as $\hat{\xi}_{iq} = \hat{e}_{jq}$.
8. Simulate an election outcome for state i in election $t = 5$ according to:

$$\hat{V}_{it} = \hat{\alpha}_{is} + \hat{\gamma}_s P_{it} + \hat{\xi}_{iq}$$

   and determine the simulated election winner in each state.
9. Repeat steps 6–8 10,000 times. Accumulate frequencies with which the respective parties win in the national vote total and in the Electoral College to determine win probabilities.

In developing our estimating equation, we stressed the importance of distinguishing state- and election-specific error components. In forecasting this is also important – when states share an election-specific error component, then forecast errors will be correlated across states within a year. We account for this phenomenon in our simulation method as well. Specifically, steps 6 and 7 above insure that in simulating we first select an election year, and then select state forecast errors from the empirical distribution for that year. Drawing from within-year error distributions leads us to correctly replicate empirical cross-state correlations in our simulations.

Tables 2 and 3 summarize the results of the Monte Carlo experiment undertaken to forecast the results of the 2004 presidential election. The results indicate that at 15 days before the election, the predicted Democratic vote share was 49.73%, the predicted Democratic Electoral College vote share was 48.30%, and the probability of a Democratic victory in the Electoral College was 51.06%. At 1 day before the election, the predicted Democratic vote share was 50.05%, the predicted Democratic Electoral College vote share was 49.68%, and the probability of a Democratic victory was 52.73%. The actual Democratic share of the two-party vote in the 2004 election was 48.75% with 46.84% of the Electoral College vote.[10] By themselves, these figures reveal limitations of point estimates as forecasts: for both time horizons the party with the higher predicted Electoral College share nevertheless has a less-than-50% chance of winning there. The tables also provide state-specific vote share forecasts and record the predicted winner and the associated Democratic win probability for each state. The distribution of possible Electoral vote outcomes is described in Figure 1 by a histogram that depicts probabilities associated with various possible Democratic Electoral vote totals.

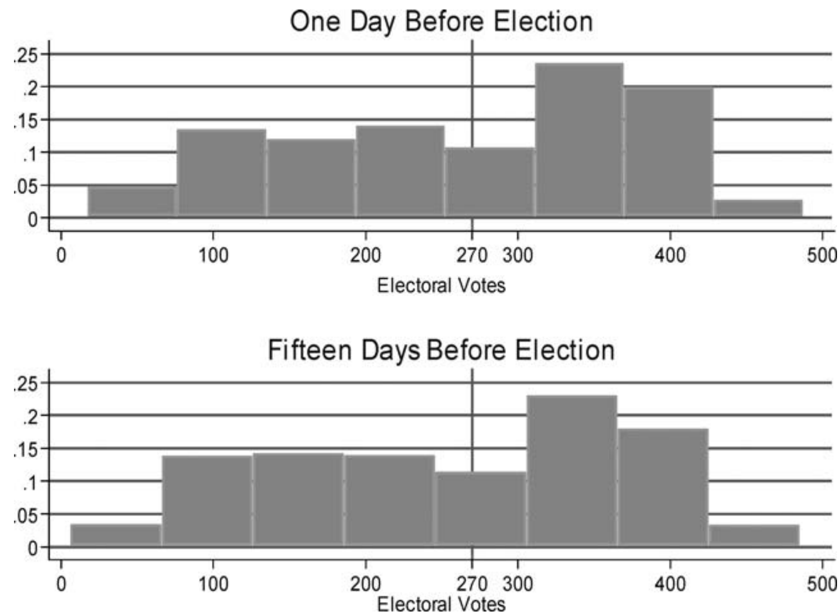

Figure 1. Simulated electoral vote distribtion.

Table 2. Forecasting performance one day before election

State/ constituency	Electoral vote	Percentage of two party vote		Error	Probability of winning	Prediction
		Outcome	Forecast			
District of Columbia	3	90.52	88.84	1.68	100%	Correct
Alabama	9	37.10	42.65	−5.55	7%	Correct
Alaska	3	36.77	36.17	0.60	0%	Correct
Arizona	10	44.72	45.00	−0.28	21%	Correct
Arkansas	6	45.06	50.73	−5.67	63%	
California	55	55.04	54.70	0.34	78%	Correct
Colorado	9	47.63	47.73	−0.10	46%	Correct
Connecticut	7	55.27	55.03	0.24	80%	Correct
Delaware	3	53.83	53.63	0.20	75%	Correct
Florida	27	47.48	46.60	0.88	33%	Correct
Georgia	15	41.65	42.71	−1.06	7%	Correct
Hawaii	4	54.40	56.81	−2.41	85%	Correct
Idaho	4	30.68	33.27	−2.59	0%	Correct
Illinois	21	55.20	55.01	0.19	79%	Correct
Indiana	11	39.58	42.52	−2.94	6%	Correct
Iowa	7	49.66	53.59	−3.93	75%	
Kansas	6	37.13	40.53	−3.40	2%	Correct
Kentucky	8	39.99	44.90	−4.91	20%	Correct
Louisiana	9	42.67	48.63	−5.96	51%	Correct
Maine	4	54.58	53.36	1.22	73%	Correct
Maryland	10	56.56	55.40	1.16	80%	Correct
Massachusetts	12	62.74	63.67	−0.93	96%	Correct
Michigan	17	51.73	52.04	−0.31	69%	Correct
Minnesota	10	51.76	53.68	−1.92	74%	Correct
Mississippi	6	40.06	42.42	−2.36	6%	Correct
Missouri	11	46.38	50.09	−3.71	59%	
Montana	3	39.50	43.53	−4.03	15%	Correct
Nebraska	5	33.15	35.41	−2.26	0%	Correct
Nevada	5	48.68	46.03	2.65	30%	Correct
New Hampshire	4	50.69	46.72	3.97	34%	
New Jersey	15	53.37	51.13	2.24	64%	Correct
New Mexico	5	49.60	50.24	−0.64	59%	
New York	31	59.29	59.32	−0.03	90%	Correct
North Carolina	15	43.76	44.31	−0.55	15%	Correct
North Dakota	3	36.09	38.14	−2.05	1%	Correct

(*Continued on next page*)

Table 2. (Continued)

State/ constituency	Electoral vote	Percentage of two party vote Outcome	Forecast	Error	Probability of winning	Prediction
Ohio	20	48.94	49.08	−0.14	55%	Correct
Oklahoma	7	34.43	39.24	−4.81	1%	Correct
Oregon	7	52.11	52.95	−0.84	73%	Correct
Pennsylvania	21	51.26	51.94	−0.68	68%	Correct
Rhode Island	4	60.58	63.65	−3.07	96%	Correct
South Carolina	8	41.36	42.05	−0.69	6%	Correct
South Dakota	3	39.09	43.97	−4.88	14%	Correct
Tennessee	11	42.81	46.96	−4.15	38%	Correct
Texas	34	38.49	43.18	−4.69	9%	Correct
Utah	5	26.65	30.82	−4.17	0%	Correct
Vermont	3	60.30	55.89	4.41	83%	Correct
Virginia	13	45.87	44.51	1.36	17%	Correct
Washington	11	53.65	54.44	−0.79	78%	Correct
West Virginia	5	43.52	51.63	−8.11	66%	
Wisconsin	10	50.19	52.84	−2.65	71%	Correct
Wyoming	3	29.69	36.58	−6.89	1%	Correct
Totals						88.23%
Electoral votes		252[a]	267	15	52.73%	
Popular vote		48.75	49.68	0.93	52.68%	

[a]The Democratic candidate Kerry actually received 251 Electoral College votes as one elector voted for the Democratic vice presidential candidate, John Edwards for both president and vice president.

Close election outcomes in 2000 and 2004 have focused renewed attention on the possibility that the winner of the popular vote might lose in the Electoral College. Our calculations imply that the *a priori* probability that the Democrats would win the popular vote but lose in the Electoral College was 1.03% at 15 days before the election and 1.52% at 1 day before the election. The odds that the Republican candidate would win the popular vote while losing the Electoral College vote were estimated to be slightly higher, at 1.50% at 15 days before the election and 1.57% at 1 day before the election.

We have also investigated the strategic importance of individual states in the 2004 election. It is generally recognized that large states, particularly large states where the outcome is expected to be close, are likely to be pivotal. To provide a quantitative assessment of the strategic importance of indiviudal states, we have carried out 51 additional Monte Carlo experiments, one each for states $i = 1, \dots, 51$. In these experiments we augmented the intercept of the vote equation for state i to add three percentage points to the

Table 3. Forecasting performance fifteen days before election

State/ constituency	Electoral vote	Percentage of two party vote		Error	Probability of winning	Prediction
		Outcome	Forecast			
District of Columbia	3	90.52	88.85	1.67	100.00%	Correct
Alabama	9	37.10	42.15	−5.04	4.89%	Correct
Alaska	3	36.77	35.96	0.81	0.15%	Correct
Arizona	10	44.72	45.10	−0.37	21.49%	Correct
Arkansas	6	45.06	50.76	−5.69	62.45%	
California	55	55.04	54.17	0.87	77.10%	Correct
Colorado	9	47.63	46.85	0.79	37.00%	Correct
Connecticut	7	55.27	54.47	0.80	77.76%	Correct
Delaware	3	53.83	53.68	0.15	73.94%	Correct
Florida	27	47.48	46.86	0.61	34.26%	Correct
Georgia	15	41.65	42.94	−1.29	8.65%	Correct
Hawaii	4	54.40	56.55	−2.15	83.90%	Correct
Idaho	4	30.68	33.74	−3.07	0.00%	Correct
Illinois	21	55.20	54.79	0.41	79.47%	Correct
Indiana	11	39.58	42.26	−2.68	4.64%	Correct
Iowa	7	49.66	52.26	−2.60	68.79%	
Kansas	6	37.13	40.37	−3.24	1.22%	Correct
Kentucky	8	39.99	45.40	−5.41	23.69%	Correct
Louisiana	9	42.67	48.58	−5.91	49.30%	Correct
Maine	4	54.58	52.65	1.93	69.94%	Correct
Maryland	10	56.56	55.24	1.32	80.23%	Correct
Massachusetts	12	62.74	63.54	−0.80	97.40%	Correct
Michigan	17	51.73	51.22	0.50	63.20%	Correct
Minnesota	10	51.76	53.91	−2.15	76.39%	Correct
Mississippi	6	40.06	42.44	−2.38	5.91%	Correct
Missouri	11	46.38	50.28	−3.90	59.38%	
Montana	3	39.50	44.01	−4.51	16.44%	Correct
Nebraska	5	33.15	35.86	−2.71	0.00%	
Nevada	5	48.68	45.46	3.22	23.08%	Correct
New Hampshire	4	50.69	46.49	4.20	32.44%	
New Jersey	15	53.37	51.30	2.07	64.26%	Correct
New Mexico	5	49.60	50.29	−0.69	60.11%	
New York	31	59.29	58.86	0.43	89.71%	Correct
North Carolina	15	43.76	44.28	−0.52	14.56%	Correct
North Dakota	3	36.09	38.58	−2.49	0.89%	Correct

(*Continued on next page*)

Table 3. (*Continued*)

State/ constituency	Electoral vote	Percentage of two party vote		Error	Probability of winning	Prediction
		Outcome	Forecast			
Ohio	20	48.94	47.78	1.16	44.94%	Correct
Oklahoma	7	34.43	39.52	−5.09	1.18%	Correct
Oregon	7	52.11	52.60	−0.50	70.62%	Correct
Pennsylvania	21	51.26	51.43	−0.17	64.8%	Correct
Rhode Island	4	60.58	63.10	−2.52	96.64%	Correct
South Carolina	8	41.36	42.04	−0.67	4.87%	Correct
South Dakota	3	39.09	44.22	−5.13	14.82%	Correct
Tennessee	11	42.81	46.90	−4.09	37.39%	Correct
Texas	34	38.49	42.99	−4.50	8.19%	Correct
Utah	5	26.65	30.70	−4.04	0.00%	Correct
Vermont	3	60.30	56.14	4.16	82.57%	Correct
Virginia	13	45.87	44.26	1.61	14.4%	Correct
Washington	11	53.65	53.43	0.21	74.22%	Correct
West Virginia	5	43.52	52.14	−8.62	67.97%	
Wisconsin	10	50.19	50.93	−0.74	62.47%	Correct
Wyoming	3	29.69	36.98	−7.30	0.45%	Correct
Totals						86.27%
Electoral votes		252[a]	260	−8	51.06%	
Popular vote		48.75	49.73	−0.98	50.59%	

[a]The Democratic candidate Kerry actually received 251 Electoral College votes as one elector voted for the Democratic vice presidential candidate, John Edwards for both president and vice president.

expected Democratic vote share, replicated the procedure described earlier, and calculated the resulting gain in the probability of Democratic victory in the Electoral College. Table 4 shows that the three percent expected vote gain produced the greatest probability increment in Florida (1.5%), with gains in Ohio (1.01%), and Pennsylvania (0.91%) also notably high.[11] Information in this form should be of particular value to candidates as they make real-time decisions about the allocation of campaign resources.[12]

4. Conclusions

We have developed a method for forecasting U.S. presidential elections that employs state-level poll data as a source of information and associates probabilities with alternative possible outcomes. Our method employs Monte Carlo simulations that draw from distributions of empirical forecast errors from an

Table 4. Probability impact of a state specific increase in expected vote share[a]

State/constituency	Electoral vote	Percentage of two party vote[b] Outcome	Forecast	Increase in probability of winning
District of Columbia	3	90.52	91.85	0.00%
Alabama	9	37.10	45.15	0.20%
Alaska	3	36.77	38.96	0.00%
Arizona	10	44.72	48.10	0.33%
Arkansas	6	45.06	53.76	0.23%
California	55	55.04	57.17	0.35%
Colorado	9	47.63	49.85	0.45%
Connecticut	7	55.27	57.47	0.10%
Delaware	3	53.83	56.68	0.07%
Florida	27	47.48	49.86	1.50%
Georgia	15	41.65	45.94	0.26%
Hawaii	4	54.40	59.55	0.04%
Idaho	4	30.68	36.74	0.00%
Illinois	21	55.20	57.79	0.29%
Indiana	11	39.58	45.26	0.16%
Iowa	7	49.66	55.26	0.26%
Kansas	6	37.13	43.37	0.01%
Kentucky	8	39.99	48.40	0.32%
Louisiana	9	42.67	51.58	0.42%
Maine	4	54.58	55.65	0.05%
Maryland	10	56.56	58.24	0.22%
Massachusetts	12	62.74	66.54	0.01%
Michigan	17	51.73	54.22	0.63%
Minnesota	10	51.76	56.91	0.26%
Mississippi	6	40.06	45.44	0.17%
Missouri	11	46.38	53.28	0.39%
Montana	3	39.50	47.01	0.03%
Nebraska	5	33.15	38.86	0.00%
Nevada	5	48.68	48.46	0.31%
New Hampshire	4	50.69	49.49	0.18%
New Jersey	15	53.37	54.30	0.53%
New Mexico	5	49.60	53.29	0.24%
New York	31	59.29	61.86	0.15%
North Carolina	15	43.76	47.28	0.43%
North Dakota	3	36.09	41.58	0.00%

(*Continued on next page*)

Table 4. (*Continued*)

State/constituency	Electoral vote	Percentage of two party vote[b] Outcome	Forecast	Increase in probability of winning
Ohio	20	48.94	50.78	1.01%
Oklahoma	7	34.43	42.52	0.01%
Oregon	7	52.11	55.60	0.26%
Pennsylvania	21	51.26	54.43	0.91%
Rhode Island	4	60.58	66.10	0.00%
South Carolina	8	41.36	45.04	0.11%
South Dakota	3	39.09	47.22	0.01%
Tennessee	11	42.81	49.90	0.58%
Texas	34	38.49	45.99	0.86%
Utah	5	26.65	33.70	0.00%
Vermont	3	60.30	59.14	0.00%
Virginia	13	45.87	47.26	0.52%
Washington	11	53.65	56.43	0.28%
West Virginia	5	43.52	55.14	0.24%
Wisconsin	10	50.19	53.93	0.37%
Wyoming	3	29.69	39.98	0.00%

[a]Increase in the one day before election probability of a Democratic win in the Electoral College resulting from a three percent increase in the expected vote share in a state.
[b]Forecast vote share plus three percentage points.

econometric model. The simulation method itself is very general, and could easily be applied to alternative econometric models that predict voting outcomes in individual states.

In an application of our method, we produced a “forecast” of the 2004 election using current pre-election polls and model estimates obtained from sample data from the 1988, 1992, 1996, and 2000 elections. Our 1-day-ahead forecast predicted a Democratic party win probability of 52.73%, even though the expected Democratic share in the electoral College was slightly below 50% The probability that one of the candidates would win in the Electoral College while losing the popular was estimated to be 3.09%.

Because our forecasts are derived from state-level outcomes and Electoral College totals, our procedure can be used to identify states where given vote gains would produce the largest increments to the overall Electoral College win probability. Using pre-election data, our calculations identified Florida and Ohio as states likely to be pivotal in the 2004 election. If candidates allocate campaign resources rationally, then the geographic dispersion of campaigning should be responsive to estimated “marginal productivities,” suggesting that the procedures developed here might also be useful in future empirical research investigating candidates' campaign strategies.

Acknowledgments

We acknowledge helpful comments from McKinley Blackburn and seminar participants at the University of South Carolina.

Notes

1. Kramer's study focused on Congressional rather than presidential elections, but he did make use of data for both election types in his work.
2. DeSart and Holbrook (1999) used state level polls for 1992 and 1996 to estimate a model that was otherwise similar to that of Campbell and Wink (1990).
3. Crain, Messenheimer, and Tollison (1993) note that independent state-specific errors tend to offset one another, and therefore have little impact on aggregate outcomes. Although Crain, Messenheimer, and Tollison do calculate election win probabilities, they adopt a longer term perspective that relies on historical patterns rather than poll data in making forecasts.
4. Although there is some variation in economic activity across states, Strumpf and Phillippe (1999) show that the cross-state variations have much smaller impacts on election outcomes than the cross-time variations; we therefore do not include a measure of state-specific economic growth.
5. It also seems likely that Iowa traders rely heavily on published polls. For example, on election day in 2004, Iowa market prices fluctuated dramatically in response to the release of exit polls. The relative accuracy of the market forecasts and those that are based on econometric analyses of polls remains an open question.
6. We initially considered a model that included the lagged voting outcome, V_{it-1}, as a an explanatory variable, but this variable was insignificant and was subsequently dropped from the model.
7. Polls were published under the Poll Track area of the web site at http://nationaljournal.com and were available only to subscribers.
8. In some states, polls were not conducted frequently. For example, in Washington, D.C. in 1996, at the forecasting horizon of 15 days before the election, the most recently available poll was in fact completed 87 days before the election. However, polls were conducted more frequently in large states where the outcome was expected to be close. Across all years, 55% of the polls used in the 15 days before the election estimation are conducted within 30 days of the election, and 90% within 53 days of the election.
9. Vote shares must be between zero and one, but predicted values from a linear regression are not constrained to this interval. As a practical matter, none of our forecasts fall outside the appropriate range (see Tables 2 and 3).
10. Kerry won 252 (46.84%) electoral votes, although one elector defected and voted for John Edwards, leaving Kerry with 251 in the formal count.
11. The estimated Democratic probability gain is also high for Texas. Our model does not account for home-state advantages (apart from their impacts captured by state polls) and this leads to a large forecast error for Texas.
12. The allocation of campaign resources would also depend on the technology for producing vote gains. For example, it is likely to be less costly to gain a given percentage vote share

in a small state than a large state, and this would also affect the allocation of campaign effort.

References

Abramowitz, A. (1996). Bill and Al's excellent adventure: Forecasting the 1996 presidential election. *American Politics Quarterly*, *24*, 434–442.

Abramowitz, A. (1992). An improved model for predicting presidential outcomes. *PS: Political Science and Politics*, *21*, 843–847.

Brown, L., & Chappell, H. (1999). Forecasting presidential elections using history and polls. *International Journal of Forecasting*, *15*, 127–135.

Campbell, J., & Wink, K. (1990). Trial-heat forecasts of the presidential vote. *American Politics Quarterly*, *18*, 251–269.

Campbell, J. (1992). Forecasting the presidential vote in the states. *American Journal of Political Science*, *36*, 386–407.

Campbell, J. (1996). Polls and votes: The trial heat presidential election forecasting model, certainty, and political campaigns. *American Politics Quarterly*, *24*, 408–433.

Crain, W., Messenheimer, H., & Tollison, R. (1993). The probability of being president. *The Review of Economics and Statistics*, *75*, 683–689.

DeSart, J., & Holbrook, T. (1999). Using state polls to forecast presidential election outcomes in the American states. *International Journal of Forecasting*, *15*, 137–142.

Erikson, R. (1989). Economic conditions and the presidential vote. *American Political Science Review*, *83*, 567–573.

Erikson, R., & Wlezien, C. (1996). Temporal horizons and presidential election forecasts. *American Politics Quarterly*, *24*, 492–505.

Fair, R. (1996). Econometrics and presidential elections. *The Journal of Economic Perspectives*, *3*, 89–102.

Fair, R. (1978). The effect of economic events on votes for president. *Review of Economics and Statistics*, *60*, 159–173.

Fair, R. (1982). The effect of economic events on votes for president: 1980 results. *Review of Economics and Statistics*, *64*, 322–325.

Fair, R. (1988). The effect of economic events on votes for president: 1984 update. *Political Behavior*, *10*, 168–179.

Kramer, G. (1971). Short-term fluctuations in U.S. voting behavior. 1896–1964. *American Political Science Review*, *65*, 131–143.

Hibbs, D. (2000). Bread and peace voting in U.S. presidential elections. *Public Choice*, *104*, 149–180.

Holbrook, T. (1991). Presidential elections in space and time. *American Journal of Political Science*, *35*, 91–109.

Holbrook, T. (1996). Reading the political tea leaves: A forecasting model of contemporary presidential elections. *American Politics Quarterly*, *24*, 434–442.

Lewis-Beck, M., & Rice, T. (1992). *Forecasting Elections*. Washington, D.C.: Congressional Quarterly Press.

Lewis-Beck, M., & Tien, C. (1996). The future in forecasting: Prospective presidential models. *American Politics Quarterly*, *24*, 468–492.

Rosenstone, S. (1983) *Forecasting Presidential Elections*. New Haven: Yale University Press.

Strumpf, C., & Phillippe, J. (1999) Estimating presidential elections: The importance of state fixed effects and the role of national versus local information. *Economics and Politics*, *11*, 33–50.