*Forecasting the Presidential Vote in the States**

James E. Campbell, *Department of Political Science, Louisiana State University*

This research examines an equation developed to forecast the two-party presidential vote in the states. The equation is composed of 16 independent variables measured at the national, regional, and state levels. It includes national trial-heat polls from early in the campaign, the growth in the national economy, presidential incumbency, the state's voting record in the previous two presidential elections, home state and regional advantages for the candidates, the partisan division of the state legislature following the last midterm election, an index of state ideology based on the roll call voting of its House delegation, and five regional trend variables to take into account partisan shifts over time. The equation is estimated over 531 state elections from 1948 to 1988. The overall fit of the equation is quite good. It accounts for nearly 85% of the variance, leaves a standard error of less than four percentage points, and has an average error of just three percentage points of the vote. This level of accuracy is comparable to that of Steven Rosenstone's equation in his *Forecasting Presidential Elections* (1983). Moreover, the equation is quite parsimonious and yields forecasts by early September of the election year.

Forecasting elections has been of widespread interest ever since popular elections were first held. Candidates and anyone involved in running campaigns have obvious personal interests in election forecasts, since strategies depend on electoral prospects. The media, activists, and the general public have their own forecasting interests. Aside from the routine curiosity of spectators, election forecasts are of broad interest because they shape a number of decisions. Depending on the forecast, the media may determine the nature of the coverage that it devotes to a campaign. Campaign volunteers, activists, and even the average citizen may decide whether or not to participate or the extent of their participation depending on their expectations about the election's likely outcome. Regardless of the use to which they are put, substantial amounts of attention and resources in modern political campaigns are devoted to election forecasts, mostly through polling. In fact, so much attention now focuses on forecasting, that there is substantial evidence to support the complaint that forecasts themselves unduly interfere with the substance of campaigns (Patterson and McClure 1976; Patterson 1980; Sigelman and Bullock 1991; Newman 1988).

Until recently, most serious scholars of elections have not shared the fascination in election forecasting of nonacademics. One noteworthy exception to this neglect was Louis Bean (1948, 1972). Although certainly not sophisticated by modern social scientific methodological standards, Bean's research was among

*I would like to thank Mark Wattier, Tom Holbrook, Mike Lewis-Beck, and the reviewers for their helpful comments on this paper.

American Journal of Political Science, Vol. 36, No. 2, May 1992, Pp. 386–407

the first to offer a systematic approach to forecasting beyond the simple reporting of poll results. Unfortunately, for some time there was little effort to build on Bean's work. Within the last few years, however, forecasting work has restarted. Several election scholars have offered models forecasting the national presidential vote (Lewis-Beck and Rice 1984, 1992; Fair 1988; Abramowitz 1988; Campbell and Wink 1990).[1] Several of these forecasting equations appear to predict the national vote with a good deal of accuracy.

Rosenstone's *Forecasting Presidential Elections*

While most rigorous forecasting analyses focus on the national presidential popular vote, Rosenstone's *Forecasting Presidential Elections* (1983) is an exception. Rosenstone's model forecasts the presidential vote in the individual states. Since electoral votes are awarded in the states based on the state vote, forecasts of the presidential vote in the states are of obvious practical significance.

Like the national vote equations, Rosenstone's forecasting equation of the state presidential vote is quite accurate. Examining the 343 state presidential elections from 1948 to 1972, the equation accounted for 93% of the variance in the presidential vote, leaving a standard error of just 4.5 percentage points of the vote (Rosenstone 1983, Table 4.1). The equation's mean absolute error is just 2.9 percentage points of the state vote. In terms of presidential candidates carrying or winning a plurality of a state's vote, the determinant of a candidate's winning the state's electoral votes, Rosenstone's equation correctly predicted the winning presidential candidate in 91% of the cases, yielding an average electoral vote error of only 33 out of 535 (District of Columbia excluded) electoral votes (Rosenstone 1983, 103).

Although Rosenstone's equation closely fits the state returns examined, several aspects of the equation may be questioned. First, Rosenstone includes several variables involving the candidates' issue positions. He measures the candidates' positions on New Deal social welfare issues and racial issues using a 1977 poll of 43 historians and political scientists (1983, 174). As measured, these issue positions would seem, at best, extremely cumbersome to calculate and, at worst, somewhat subjectively determined. Even if we could safely assume that the professional judgments of this panel of experts entirely checked their possible unrepresentative personal political judgments, there is still the matter of hindsight. How much faith should we put in the ability of a panel of political scientists in 1977 (or any other panel in 1977 for that matter) to evaluate Tom Dewey's racial positions in 1948 in a way remotely like that of voters at the time? While Rosenstone took meticulous care in his use of these data (e.g., standardizing

[1]Strictly speaking, Fair's work was designed to explain rather than to forecast the presidential vote. However, the emphasis placed on his equation's completeness in accounting for vote variance might place it within the group of forecasting models.

each respondent's placement of the candidates) and while there seems to be no obvious misplacement of candidates, it is difficult not to be wary of the composition of these issue variables and skeptical of their value in actually forecasting elections.[2]

Second, as Rosenstone notes in his analysis of the 1980 election, data for some of his variables are unavailable until after the election they are to forecast. In particular, his measures of economic growth are not available before the election (Rosenstone 1983, 122). As Lewis-Beck (1985) rightly contends, a very important property of any forecasting model is that it permits a forecast well before the event. This is not so easily accomplished if we first must accurately forecast the economy before we can forecast the election.

Third, the inclusion of some variables may be regarded as a bit ad hoc. While every election is affected by variables peculiar to it (given its unique political conditions and candidates), when are these considerations so idiosyncratic that they should be left out of an equation proposed for *general* forecasting use? Moreover, is there any objective *before-the-fact* indication that these variables should be included in the forecasting equation?

Possible objections to two variables may be raised on the grounds of their idiosyncratic nature. First, in recognition of Kennedy's Catholicism, Rosenstone included in his equation the Roman Catholic population of the state only for the 1960 election. No other candidate's religious background or, for that matter, any other personal characteristic has triggered the use of a variable of this sort. Should the Catholicism of vice-presidential candidates have been a consideration in forecasting the 1972 (Shriver) or the 1984 (Ferraro) votes or in the forecasts of future Catholic presidential candidates? Second, to a lesser extent, the same objection might be raised about the "mismanagement of wars" variable included only in forecasts of 1952 and 1968 (and not included in 1972 while the Vietnam War continued). Also, in the case of the 1980 election alone, Rosenstone included three additional variables: poll data concerning the party thought best for prosperity, the party best for peace, and the extent of party factionalism as revealed by delegate voting at the national party conventions.

Finally, Rosenstone's equation includes quite a large number of variables. The equation is estimated over 343 cases yet consists of a total of 74 independent variables, including 49 dummy variables for separate state intercepts. While

[2]Even if the panel of experts were able to look dispassionately at the past in identifying the issue positioning of candidates, there still remains the question (an important question for forecasting purposes) of whether such a panel could offer the same dispassionate observations in the heat of an ongoing campaign. Moreover, setting aside the subjectivity concern, constructing the issue indices in a particular election is a very complicated task, as his Appendix A (Rosenstone 1983, 155–80) indicates. Additionally, the effects of these issues are specified to have different effects in different regions in different election years. As a result, forecasts are based on nine different coefficients for the two general issues.

these state dummies were included to deal with estimation problems (see Rosenstone 1983, Appendix B), they make the model anything but parsimonious. Moreover, by their sheer number, they account for a substantial amount of variance.

Recently, Holbrook (1991) constructed an alternative model to explain presidential voting in the states across time. Although explicitly not intended for use as a forecasting model, Holbrook clearly builds on Rosenstone's earlier work. He develops a 13 variable equation of the incumbent party's presidential vote and examines the equation over the 350 state elections held from 1960 to 1984. Holbrook's equation includes seven state variables (two economic indicators, two ideological variables, a partisanship index, and two home state advantage variables), two regional variables, and four national variables (two economic variables, an incumbency variable, and a measure of the president's popularity). While much more parsimonious than Rosenstone's equation and, again, though not intended for forecasting purposes, Holbrook's equation appears less accurate than Rosenstone's. Although comparisons between models estimated with different samples pose some hazards, the average absolute errors might suggest the relative strengths of the two equations.[3] As noted above, the average absolute error of Rosenstone's model is just 2.9 percentage points while the average absolute error of Holbrook's equation is 4.4 percentage points.

This research attempts to build a forecasting model with the accuracy of Rosenstone's model and the parsimony of Holbrook's model. Moreover, to the extent possible, the proposed model uses objective predictor variables available prior to the election. The proposed model takes as its starting point the variables used so successfully to forecast the national presidential vote. Both Lewis-Beck and Rice (Lewis-Beck 1985, 58) and Campbell and Wink (1990) found that a combination of early September trial-heat polls and second quarter real GNP growth produced highly accurate forecasts of the national vote. In addition to the these two variables, the equation includes prior state deviations from the national vote as well as several other indicators of partisan strength and short-term conditions in the state and its region of the country. All indicators are objective and available by early September of the election year—the traditional kick-off of the general election campaign.

[3]Since Holbrook's dependent variable is oriented in terms of the incumbent presidential party, a comparison of the Holbrook and the Rosenstone R-squares would be meaningless. A comparison of average absolute errors, like a comparison of standard errors of the estimate, runs the risk of confusing the strength of the model with the variance of the dependent variable in the two samples. That is, even a weak model may have relatively small errors if the variance of the dependent variable is itself small. In this particular case, the variance of the two samples are quite comparable. The standard deviation of the state presidential votes from 1948 to 1972 (Rosenstone) is 11.0, while the standard deviation of the vote from 1960 to 1984 (Holbrook) is 10.3. Thus, if anything, the smaller errors of the Rosenstone model, given the greater variance of the vote for his sample, are more impressive.

The Dependent Variable

The dependent variable is the Democratic presidential candidate's share of the two-party statewide vote.[4] The two-party rather than total vote percentage simplifies matters by essentially excluding minor-party votes. Given that third-party votes can be drawn from either party's supporters in any given election and that the emergence of significant third-party candidates in elections is sporadic, it would be extremely cumbersome to attempt to predict generally the strength of third-party candidates in presidential contests. Moreover, since third-party candidates rarely carry states, the parsimony afforded by using the two-party vote can be purchased at a minimal expense to accuracy. The accuracy of predicting the dependent variable will be assessed in two ways: the difference between the expected and actual vote and the difference between the party expected to win and the party that actually carried a majority of the two-party vote in the state. Since it is ultimately the states' electoral votes that elect the president, and since all but one state award their electoral votes as a unit to the candidate with a plurality of the individual state's popular vote, this second assessment is of great practical importance.[5]

Independent Predictor Variables

The forecasting equation is built from 16 predictor variables. There are three types of variables, differing in the temporal and spatial dimensions: contemporaneous national variables, contemporaneous and historical state variables, and variables taking regional trends into account. It should be emphasized at the outset that in specifying the model, I have drawn heavily from previous theoretical explanations of the vote: the notion of the "normal vote," economic voting, incumbency effects, "friends and neighbors" voting, and regional patterns of partisan realignment.[6] Also, while the equation shares a few variables in common with the Rosenstone and Holbrook equations, it differs from those models in

[4]The dependent variable in Rosenstone's analysis was also the Democratic share of the two-party popular vote. However, Rosenstone altered this dependent variable in two of the elections: (1) for the 1948 vote he combined votes for minor-party candidates Henry Wallace and Strom Thurmond with Truman's vote and (2) for the 1968 vote he counted Humphrey's share of the total vote and did not count third-party candidate George Wallace's vote or use two-party Humphrey vote. For the sake of consistency, the analysis here uses only the two-party vote share and addresses the third-party vote problem by estimating the equation both with and without states that had substantial third-party votes.

[5]The one exception is Maine, which awards its votes according to the votes within the congressional district. However, given that the state has had only two congressional districts since the 1960s, the state's electoral laws have not yet produced a split in its electoral vote.

[6]There are, of course, also differences in the selection of variables for the purpose of prediction as opposed to explanation. Most notably, while trial-heat polls may be very valuable information for prediction and may encompass a wide variety of unspecified influences on the vote, the fact that these influences are unspecified makes them useless for explanation purposes.

several respects. The similarities and differences of these three equations are summarized in Table 1.

Current National Variables

The first type of predictor variable indicates the national political climate of the current campaign. These variables have been used to predict the national presidential vote. Since the national vote is a simple aggregation of the individual state presidential votes, it may help predict the state votes. Moreover, compared to state-level indicators, those at the national level are available more regularly and are released more quickly (an important consideration in forecasting). In addition, voters may react to these general national circumstances as well as to their more parochial local concerns.

There are three national-level variables in the model: early Gallup trial-heat poll results, the state of the national economy, and the incumbency status of the presidential candidates. First, the trial-heat variable is the percentage of the two-party preference for the Democratic presidential candidate in the national Gallup poll conducted in early September following both parties' national conventions. Second, the state of the economy variable is the second-quarter change in the real GNP. This index takes into account real economic growth while discounting for inflation.[7] These data are available in the August election year issues of the Department of Commerce's *Survey of Current Business*. Since there is substantial research that the incumbent presidential party is held accountable for the state of the economy, the rate of economic growth is multiplied by a variable that indicates the party of the administration. This administration variable is plus one for Democratic administrations and minus one for Republican administrations.

Like both Rosenstone's and Holbrook's equations, this equation incorporates a presidential incumbency advantage. By inertia alone, the incumbent president, as a proven popular vote getter, ought to have an advantage over a challenger who has not previously run and won across the entire nation. The presidential incumbency variable simply indicates the party of the current president and whether he is seeking reelection. It is coded plus one for a Democratic president seeking reelection, minus one for Republican president seeking reelection, and zero for elections in which the president is not seeking reelection or when an unelected president is seeking election (President Ford in 1976).[8]

[7]It would be preferable to include, as Rosenstone does, real disposable income per capita to discount for tax burdens; however, that data for periods close to election are not available until after the election. Thus, it is not useful for before-the-fact forecasting.

[8]Rosenstone (1983, 74) finds presidential incumbency to provide the incumbent with a four percentage-point advantage. Holbrook (1991, 103), quite surprisingly, estimates the presidential incumbency advantage at a full 12 percentage points. Given the inclusion of trial-heat ratings in the present equation (and incumbents score higher on these ratings) and the past performance of the

Table 1. Specifications of Three State Presidential Vote Equations

Variables	Equation		
	Current	Rosenstone	Holbrook
Dependent variable:			
State democratic percentage of the two-party presidential vote	Yes	Yes[a]	Incumbent party
Independent variables:			
National trial-heat poll results	Yes	No	No
National presidential popularity	No	No	Yes
Social welfare issues	No	Yes(3)	No
Racial issues	No	Yes(6)	No
National economic indicator	Yes	Yes	Yes(2)
Presidential incumbency	Yes	Yes(2)	Yes
Vice-presidential incumbency	No	Yes	No
Mismanagement of war (1952 & 1968)	No	Yes	No
Catholic population (1960)	No	Yes	No
Southern home region advantage: president	Yes	Yes	Yes[b]
Southern region	No	No	Yes
Regional trends	Yes(5)	No	No
State presidential voting history	Yes(2)	No	No
Home state advantage: president	Yes	Yes	Yes
Home state advantage: vice-president	Yes	Yes	Yes
Home state advantage: 3rd-party cand.	No	Yes	No
Midterm state U.S. house vote	No	Yes(6)	No
Midterm state legislative results	Yes	No	No
State economic indicator	Yes	No	Yes(2)
State ideological inclination	Yes	No	Yes(2)
State partisanship	No	No	Yes
State dummy variables	No	Yes(49)	No
Total coefficients (excluding constant)	16	74	13
Elections included	1948–88 (N = 531)	1948–72 (N = 343)	1960–84 (N = 350)

Note: The number in parentheses indicates how many variables of this type or interactions that were included in the equation. Also, there are some differences of variables within a type (e.g., different national economic indicators, different codings of the southern home region variable).

[a]Except for 1948 and 1968 (not two-party).

[b]Holbrook's presidential home region advantage variable is not restricted to the southern states.

State Variables

A second group of variables is composed of state characteristics, either indicating the partisan predisposition of a state based on its history or suggesting partisan reactions to current circumstances (such as the current condition of the state's economy). There are seven state variables: the state's deviation from the national presidential vote in the previous two presidential elections, the home state advantages of both presidential and vice-presidential candidates, the partisan division of the state legislature, the rate of economic growth in the state, and the state's ideological leanings as revealed by the voting records of its congressional delegation to the House of Representatives.

An obvious place to look for clues about how a state will vote in the next election is how it voted in the past, its usual or normal vote (Converse 1966). Of course, past elections have their own set of short-term forces that do not necessarily pertain to the current election. For this reason, it would be fruitless to seek forecasting help from a state's simple presidential vote in a prior election. Even the most Republican-oriented state might vote for the Democratic candidate in a Democratic landslide. However, if we assume a general national uniformity in short-term forces, deviations of past state votes from the election's national vote may be informative of the partisan disposition of the state. If a state has consistently exceeded the national Democratic presidential vote in past elections, we ought to expect that Democratic candidates will do relatively well in the next election. A preliminary analysis suggests that this is indeed the case. A state's deviation from the national presidential vote is strongly correlated with its deviation from the last election's national presidential vote ($r = .68$). Since elections prior to the last election may also provide clues to the state's next vote, the equation includes state deviations from the national vote in the previous two presidential elections.[9]

Since these previous deviations from the national vote may be influenced by some short-term forces peculiar to a particular state rather than common to the nation as a whole, three adjustments to these deviations were made. The state deviations were corrected for past temporary advantages of presidential and vice-presidential candidates in their native states and, in the case of southern Democratic presidential candidates, their region (see below).[10]

---

incumbent in the state, I expect the incumbency advantage to be a good deal smaller than either prior estimate. Also, while a case could be made for grouping Ford with elected incumbents, since he enjoyed the perquisites of office, his unprecedented nonelectoral route to the presidency sets him apart from other incumbents and suggests that he did not have many of the electoral advantages of an incumbent who seeks reelection.

[9]Preliminary analysis included the deviation of the state's vote in the previous three presidential elections. However, the earliest of these elections, held 12 years before the predicted election, did not provide additional predictive power and was dropped from the final model.

[10]The adjustment was arrived at by an iterative process: estimating temporary home state and regional advantages in the full model and then using those estimates to adjust prior state

The model also includes two variables to take into account the home state bonus vote for presidential and vice-presidential candidates. Presidential tickets may win added "friends and neighbors" votes in their home states. Rosenstone (1983, 74), Lewis-Beck and Rice (1983), and Garand (1988) find presidential candidates to win approximately an extra four percentage points of the vote in their home states. These studies also indicate a vice-presidential home state advantage of slightly smaller proportions. A party's ticket is estimated to win a bonus of about 2.5 percentage points in its vice-presidential candidate's home state. While candidates generally win added support in their home states, the extent of this advantage apparently differs according to the size, diversity, and population of the state. As one might expect, as Lewis-Beck and Rice (1983, 555) found and as preliminary analysis also suggested, candidates from large, populous states receive a smaller percentage point boost in their home states.[11] This difference between the home state advantage in large and small states is incorporated in the coding of both the presidential and vice-presidential home state variables. The Democrat's home state is coded plus one and the Republican's home state is coded minus one, except when the candidate is from a large state. The home state advantage variables for candidates from New York, Illinois, and California are coded as half the value of other states.[12]

Another likely short-term influence on a state's vote is its economy. The incumbent presidential party is generally held accountable for the state of the economy. While the national economic indicator, discussed above, may suggest general economic circumstances, economic conditions may differ a good bit from state to state. The forecasts may, therefore, benefit from including a measure of *state* economic conditions. The index of state economic growth used here is the growth in a state's total personal income between the prior year's fourth quarter and the first quarter of the election year. This index is available in the Commerce Department's *Survey of Current Business* far enough before the election to be useful for forecasting yet taps economic conditions close enough to the election

---

deviations from the national vote. The eventual weighting was 6.6 for the presidential home state advantage, 2.5 for the vice-presidential home state advantage, and 7.3 for the southern Democrat home region advantage.

[11]A preliminary analysis was conducted without coding the home state advantage in large and small states differently. Errors in the forecast of elections in the 22 home states of presidential candidates over the 11 elections were examined. Favorite sons exceeded expectations in 13 cases and fell short in nine. Most of these nine were elections in large states. In fact, of the nine cases in which the home state candidate did not do as well as expected, seven involve California, Illinois, or New York. Presidential candidates from smaller states quite consistently received more than their expected share of the vote, usually two or three percentage points more.

[12]One of the home state cases to fall short of expectations in the preliminary analysis (outside of California, Illinois, and New York cases) was Texas in 1988. The vote for George Bush fell 5.6% short of the expected vote. This may be explained by Bush's close ties with other states (Connecticut and Maine), the fact that Texas itself is a large state, and the vice-presidential candidacy of Texas Democratic Senator Lloyd Bentsen.

to reflect economic conditions that may influence vote decisions.[13] One problem with this index, however, is that it does not take into account inflation. To remedy this problem, the index is standardized across states within each election year. As such, it indicates, in a relative sense, the states in which the economy is better or worse than average. Like national economic conditions, since economic rewards or punishments are meted out to the incumbent party, the index is multiplied by plus one if the incumbent administration is Democratic and minus one if the incumbent administration is Republican.

The final two state-level variables concern the general partisan and ideological tendencies of the state. All things being equal, traditionally liberal Democratic states might be expected to vote more Democratic than traditionally conservative Republican states. The partisanship variable is the partisan division of the lower chamber of the state legislature after the previous midterm election.[14] The index is the Democratic share of state legislative seats above the 50% mark.[15] The predominant ideological disposition of a state's public is measured by an index constructed by Holbrook and Poe (1987). This state ideology index is calculated from the ADA and ACA ratings of the roll call votes of each state's delegation to the U.S. House of Representatives. To make the index consistent with other indicators, the sign was reversed so that higher values indicated more liberal states.

[13]Unfortunately, no state-level quarterly economic data were available for late 1947 for use in the 1948 election forecast. Annual state data were substituted in its place for this one year. Since the data are, in any case, standardized within the year, this probably has little, if any, effect in the equation's performance.

[14]Two states, Nebraska and Minnesota, held nonpartisan state legislative elections for all (Nebraska) or part (Minnesota) of the period under study. In the case of Nebraska, the state legislative division was estimated based on the ranking of states on Wright, Erikson, and McIver's state partisan rankings based on public opinion data. Using this index, Nebraska was assigned the mean partisan division of the state most similar to it on the public opinion index, the nearly equally Republican state of North Dakota (28% Democratic). The partisan division of the Minnesota legislature in its nonpartisan years (before 1972) is coded as the mean of its partisan division once it reformed to partisan elections (61.57% Democratic). Also, some states do not hold state elections in the midterm election year. In these cases, the partisan composition of the state legislature at the midterm (though usually elected a year earlier) was still used.

[15]Rosenstone included the statewide congressional midterm vote as a comparable partisan index (1983, 89–90) and also included the squared term to take into account diminishing returns. State legislative midterm results may be preferable for several reasons: (1) they are more readily accessible; (2) they do not appear to require adjustments for incumbency and uncontested seat effects; and (3) preliminary analysis suggested that they were stronger indicators of partisan dispositions. The statewide midterm congressional vote, without its various adjustments, was examined and did not significantly improve the forecast. Also, to take into account possible "diminishing returns" of extreme partisan divisions, a squared term (of the Democratic state legislative advantage) was included in the preliminary analysis. This parallels Rosenstone's use of a squared congressional vote. However, this squared state legislative variable failed to strengthen the equation's fit and was dropped from the final equation.

Regional Variables

The third set of predictor variables controls for factors that affect groups of states in a region, principally electoral trends across an entire region. Considerable research (Bullock 1988; Petrocik 1987; Stanley 1988; Carmines and Stimson 1989) indicates regional partisan shifts beginning in the early 1960s, with some regions (most notably the South) moving in the Republican direction while others became more Democratic. Although many of these disruptions to preexisting state partisan dispositions could not themselves have been anticipated and are thus of questionable forecasting value, their inclusion is necessary for a reasonably complete general model over an extended series of elections across a diverse set of states.

There are six regional variables in the equation. Five concern regional partisan trends related to the partisan realignment begun in the 1960s. Each of the five regional trend variables are coded as dummy variables. Three of these five trends concern developments begun in the early 1960s: a strong shift in the Republican direction in the Deep South, a more moderate shift to the Republicans throughout the South, and a shift in the Democratic direction in New England.[16] The remaining two regional trends concern the shift of North Central states toward the Democrats that preliminary analysis suggests began in 1972 and the shift of Rocky Mountain states toward the Republicans that appears to have begun in 1976.[17] These regional trends are essentially those identified by Bullock (1988) in his examination of subpresidential election outcomes. Bullock's analysis finds consistent evidence that states in the South and West have shifted toward the Republicans while those in the Northeast and North Central regions have shifted toward the Democrats. Each regional trend is specified as a dummy variable in a single election (or two elections in the cases of the New England and Rocky Mountain variables). It is unnecessary to specify continuing effects, since the effects of a regional shift in later elections should be reflected in a state's vote deviations in previous elections and its state legislative partisan divisions.

The sixth regional variable is the advantage southern Democratic presidential candidates have in that region. As Rosenstone and others have found, the cultural and historical distinctiveness of the South and its strong regional con-

[16]States are classified as follows: (1) Deep South includes Alabama, Georgia, Louisiana, Mississippi, and South Carolina; (2) South includes the Deep South states plus Arkansas, Florida, North Carolina, Tennessee, Texas, and Virginia; and (3) New England includes Maine, Massachusetts, New Hampshire, Rhode Island, and Vermont. Each of the three variables is coded plus one for the pertinent states in 1964 and zero otherwise. The New England variable is also coded plus one in the 1960 election.

[17]The North Central states (coded one in 1972) include Illinois, Iowa, Kansas, Minnesota, Michigan, Nebraska, North Dakota, South Dakota, and Wisconsin. The Rocky Mountain western states (coded one in 1976 and 1980) include Alaska, Arizona, Colorado, Idaho, Montana, Nevada, New Mexico, Utah, and Wyoming.

sciousness are unique and result in regional favorite son treatment of southern presidential candidates unlike that for any other region of the country (Rosenstone 1983, 64–65). The southern regional Democratic candidate advantage variable is coded plus one for southern states in Jimmy Carter's 1976 and 1980 elections and zero otherwise.[18]

Data

The equation composed of the above 16 variables is estimated over a total of 531 state-level presidential elections from 1948 to 1988. Over this period there were a total of 544 state-level presidential election results.[19] This total includes 11 elections in each of the 50 states, except Hawaii and Alaska which began voting in the 1960 election. Of this total, 13 state elections were excluded. Four elections in Alaska and Hawaii were excluded because they lacked the electoral history (the past deviations from the national vote) used as independent variables in the forecasting equation. The remaining nine omitted cases are the result of idiosyncrasies in presidential voting in Alabama in 1948 (Truman was excluded from the ballot), in Mississippi in 1960 (39% of the vote went to unpledged electors), and in Alabama in 1964 (Johnson was excluded from the ballot). These problems resulted in omitting the three idiosyncratic cases and another six cases in which the prior state presidential vote was not then available as an independent variable.

The equation is estimated with OLS regression.[20] Although spatial and temporal autocorrelation are potentially problems for OLS estimation of pooled time series data, analysis of residuals failed to reveal any general autocorrelation

[18]The coding of the southern regional advantage variable differs from Rosenstone's coding (1983, 185). Rosenstone coded the variable plus one in 1948 and minus one in 1968. This difference is a result of his counting Strom Thurmond votes in 1948 as Democratic votes and of counting George Wallace votes in 1968 as non-Democratic votes. Recall that Rosenstone's dependent variable in 1968 was the Humphrey vote as a percentage of the total vote (Wallace included) rather than the two-party vote. Also, Rosenstone omits Texas and Florida from the South on this variable (1983, 185) but they are counted as in the South here.

[19]The equation was also estimated over those state elections in which there was no significant third-party vote (less than 10%) in the presidential contest ($N = 502$). The substantial third-party vote cases are restricted to three election years: 1948, 1968, and 1980. Strom Thurmond received 10% or more of the 1948 vote in nine states. George Wallace received 10% or more of the 1968 vote in 23 states. In 1980 John Anderson received a substantial vote in nine states, and that election's Libertarian candidate, Ed Clark, received more than 10% of the vote in one state (Alaska). The results of this restricted analysis closely correspond to the full analysis; therefore, the analysis of all state elections regardless of third-party activity is reported here.

[20]Although the two-party vote is bounded (0 to 100%), which might suggest logit analysis, its distribution is well within these bounds. Given the straightforwardness of the linear regression model, it was used. However, a logit analysis was conducted on the dichotomous dependent variable of which party carried the state. The logit estimation produced expected winners that were just as accurate as those obtained by the OLS estimate of the vote itself, assigning winning and losing parties by the expected two-party vote above or below the 50% mark.

problem. With the elections sorted by state and election year, the equation's Durbin-Watson statistic indicated no significant autocorrelation ($DW = 1.88$).

Findings

Predicting the Popular Vote

The estimated equation is presented in Table 2. Each independent variable has its expected effect and is statistically significant ($p < .01$). In fact, with the exception of the vice-presidential home state advantage, each coefficient is at least three and a half times the magnitude of its standard error. Of the 16 variables, four stand out as particularly strong. The two variables used in national vote forecasts (the early September national trial-heat poll results and the second-quarter growth in the national economy) are the strongest clues to election results as indicated by the standardized coefficients. Nearly as helpful are the two previous state deviations from the national presidential vote. As expected, a party's presidential candidate should expect to do better than average in states in which the party's previous presidential candidates had done especially well.

Of the remaining 12 variables, nine appear to be of roughly equal importance in forming the forecast (betas of .11 to .15): the incumbency advantage, the presidential home state and southern regional advantages, the partisan (state legislative division) and ideological leanings of the states, and four of the five regional trend variables. The three least powerful predictor variables are the regional shift of North Central states toward the Democrats in 1972, the relative growth in state economies, and the vice-presidential home state advantage.[21]

By almost any standard, the equation produces a close fit to the data. It accounts for nearly 85% of the variance in the state presidential vote over this period of four decades. While the standard deviation of the presidential vote is nearly 10 percentage points (9.74), the standard error of the estimate is only 3.86 percentage points. The average absolute error of the state vote is just three percentage points. In terms of "goodness of fit," the equation performs just about as well as Rosenstone's.[22]

The distribution of errors (grouped to the closest percentage point) are depicted in the bar chart in Figure 1. As the figure shows, most errors are fairly small. About half of the elections are missed by less than 2.5 percentage points, and 85% of the elections are missed by less than 5.5 percentage points. Fewer

[21]Dropping these three weakest variables only slightly weakens the equation (adjusted R-square = .832 and standard error = 3.990).

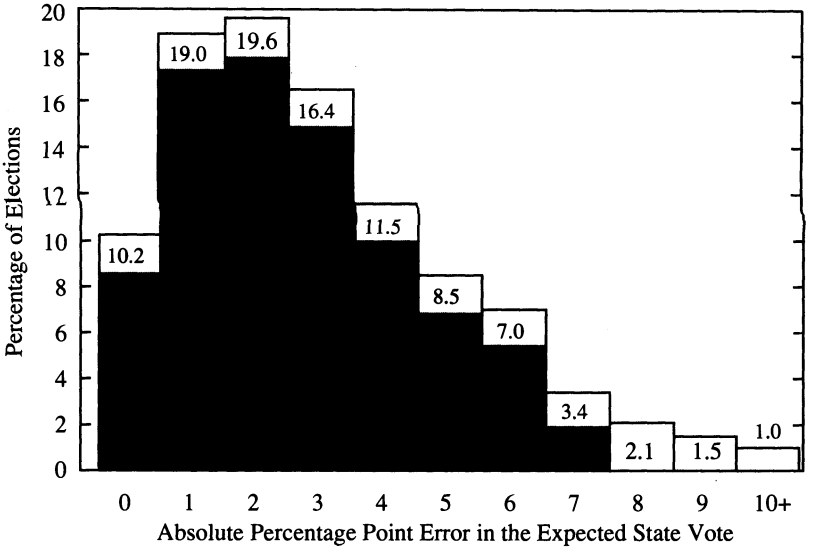
[22]While studies based on different samples with different variances in the dependent variable make comparisons of R-squares and the standard error of estimates problematic, they nonetheless may suggest the models' relative strengths. The R-squares are quite high in both equations (.85 vs. .93 for Rosenstone). The standard error of the estimates is a bit better in the current model (3.86 vs. 4.5 for Rosenstone). In terms of the average absolute error, the equations are about on par (3.0 vs. 2.9 for Rosenstone).

Table 2. The Forecasting Equation of the Two-Party Presidential Popular Vote Percentage in the States, 1948–88

	Coefficient	SE	beta
Dependent variable:			
Democratic two-party state presidential vote percentage			
Independent variables:			
Constant	26.239	1.212	—
National variables:			
Democratic trial-heat percentage	.449	.026	.409
2d-qtr. GNP growth × incumbent party	2.318	.160	.322
Elected incumbent seeking reelection	1.550	.318	.118
State variables:			
Prior state vote deviation from national vote ($t - 4$) (adjusted)	.321	.033	.263
Prior state vote deviation from national vote ($t - 8$) (adjusted)	.263	.028	.244
Presidential home state advantage	6.671	.995	.119
Vice-presidential home state advantage	2.485	.942	.046
State legislature party division ($t - 2$)	.047	.010	.115
Standardized 1st-qtr. state economic growth × incumbent party	.632	.173	.064
State liberalism index (ADA & ACA)	.035	.005	.126
Regional variables:			
Pres. home region (southern) advantage	7.308	.994	.150
Southern state (1964)	−8.646	1.715	−.115
Deep South state (1964)	−17.030	2.748	−.131
New England state (1960 & 1964)	7.739	1.193	.118
Rocky Mountain West state (1976 & 1980)	−6.938	.977	−.132
North Central state (1972)	5.412	1.350	.072
Number of cases	531		
R^2 (adjusted R^2)	.848 (.843)		
Standard error of estimate	3.857		
Mean absolute error	3.020		

Note: Partisan divisions (e.g., presidential vote %) involve only the two major parties. Except for regional trends, positive values on each variable favor Democratic presidential candidates. Regional trends are dummy variables (1 for states in the region for the specified year[s] and 0 otherwise). The adjustment to the prior presidential vote deviations include the temporary prior state and regional (southern) presidential and vice-presidential advantages.

Figure 1. Distribution of Errors in the Expected State Two-Party Presidential Vote, 1948–88



Note: The expected percentage of the vote was calculated from the equation in Table 2. The absolute differences between these expected votes and the actual votes were then calculated and grouped to whole percentages (with .5 rounded up).

than 3% of the expected votes are off by more than 8.5 percentage points. Given the numerous idiosyncrasies that may intrude in elections over this stretch of time, the frequency of large errors is rather small.

Residual Analysis

The average errors for election years and for individual states are presented in Tables 3 and 4, respectively. In terms of the errors in particular years, the only election that deviates much from the average is 1952. When we examine the residuals, it appears that Eisenhower's candidacy had a greater than otherwise expected appeal in farm belt and Rocky Mountain states. The Eisenhower vote exceeded expectations by at least five percentage points in Arizona, Idaho, Iowa, Montana, Nebraska, Nevada, North Dakota, South Dakota, Texas, and Utah. While this may suggest Eisenhower's absolute strength in this section of the country, it probably says more about his appeal in this region compared to that of the dapper New York Governor Thomas Dewey, the GOP standard-bearer in the 1944 and 1948 elections. Because of Dewey's appeal to the eastern, or "Wall-Street," wing of the Republican party, the deviations of the previous votes from the national vote may have been less a clue to the vote in 1952 than usual.

The average errors in individual states (Table 4), while again demonstrating

Table 3. Average Forecast Error by Election Year, 1948–88

Election	Average Absolute Vote Percentage Error
1948	3.27
1952	4.00
1956	3.53
1960	2.89
1964	2.45
1968	2.44
1972	2.89
1976	3.44
1980	2.72
1984	2.74
1988	2.87
Total	3.02

the general accuracy of the equation, also indicate where the fit might be improved. Before examining a few of these areas, we might note that both the above equation and the Rosenstone equation, despite their many differences, had trouble predicting the vote in the same states. Of the 10 most troublesome states for each equation, seven are the same: Georgia, Hawaii, Maine, Mississippi, Oklahoma, South Carolina, and Texas. These average state errors along with the examination of the individual state election errors may help identify weaknesses in the above model. The largest individual errors are listed along with their actual and expected votes in Table 5.

Aside from errors associated with the swing of the Republican party away from its "Wall Street" wing in 1952, there appear to be several patterns to the errors. Three in particular seem to be worth closer examination. Several sizable errors appear to be associated with the Dixiecrat revolt of 1948, reactions in the 1960 election to Kennedy's Catholicism, and, more generally, the spill-over of regional and state influences into bordering states.

1. The Dixiecrat candidacy of Strom Thurmond in 1948 appears to complicate forecasts in several southern states in 1948, especially Thurmond's home state of South Carolina. Thurmond actually carried four states (Alabama, Louisiana, Mississippi, and South Carolina) and won a significant portion of the vote in another eight.[23] Obviously, under these circumstances the two-party vote may, on occasion, misrepresent voter sentiment (though dropping all significant third-party state elections had little effect on the results). Moreover, since forecasts

[23]Truman was denied designation as the Democratic presidential nominee in the four southern states that Thurmond carried (Garson 1974, 311).

Table 4. Average Forecast Error by State, 1948–88

Rank	State	Mean Absolute Error	Rank	State	Mean Absolute Error
1.	Indiana	1.089	26.	Arizona	2.820
2.	Alabama (5)	1.264	27.	New Jersey	2.857
3.	Michigan	1.654	28.	Kentucky	2.890
4.	North Carolina	1.680	29.	Nebraska	2.933
5.	Missouri	1.689	30.	South Dakota	2.972
6.	Ohio	1.818	31.	New York	2.979
7.	Illinois	2.200	32.	Idaho	3.075
8.	Wyoming	2.261	33.	Nevada	3.237
9.	Minnesota	2.279	34.	North Dakota	3.382
10.	Delaware	2.308	35.	New Mexico	3.513
11.	Connecticut	2.316	36.	Oregon	3.550
12.	Montana	2.327	37.	Arkansas	3.610
13.	Pennsylvania	2.331	38.	Iowa	3.645
14.	Tennessee	2.380	39.	Rhode Island	3.724
15.	Virginia	2.472	40.	Texas	3.842
16.	Washington	2.512	41.	Massachusetts	4.000
17.	Utah	2.583	42.	California	4.065
18.	Alaska (6)	2.587	43.	Georgia	4.152
19.	Florida	2.621	44.	Vermont	4.352
20.	Maryland	2.632	45.	Hawaii (6)	4.444
21.	Wisconsin	2.651	46.	Maine	4.569
22.	Colorado	2.664	47.	South Carolina	4.596
23.	West Virginia	2.715	48.	Louisiana	4.722
24.	New Hampshire	2.742	49.	Oklahoma	5.272
25.	Kansas	2.811	50.	Mississippi (8)	5.329

Note: Unless the number in parentheses indicates otherwise, each mean is based on 11 elections.

rely significantly on previous state votes, this 1948 complication reverberates in the 1952 and 1956 forecasts. Given this, it is not surprising that the South Carolina forecasts for 1948, 1952, and 1956 are among the largest errors of the series, as listed in Table 5.[24]

2. Several large errors result from not taking into account (as Rosenstone did take into account) the special appeal of Kennedy's Catholicism in the 1960

[24]The 1948 Dixiecrat vote may have also contributed to the larger than average errors in 1952 (Table 3) by adding error to the previous vote deviation variable used in forecasting the 1952 vote. The equation was reestimated after eliminating the 1948, 1952, and 1956 elections in the four southern states carried by Thurmond in 1948 (1952 and 1956 elections were dropped, since the forecasts depend on the states' 1948 votes). As reestimated, the standard error dropped from 3.857 to 3.703, and the mean absolute error declined from 3.02 to 2.93.

election. Kennedy did especially well in states with sizable Roman Catholic populations, far exceeding expectations in Illinois (+4.8%), Louisiana (+12.3%), and New York (+5.8%). Conversely, Kennedy's vote fell well short of expectations in states with relatively small Roman Catholic populations, including Kentucky (−5.2%), Oklahoma (−9.3%), South Carolina (−6.0%), and Tennessee (−5.5%). Overall, the more Roman Catholic a state's population in 1960, the more Kennedy's vote was underestimated by the forecast model ($r = .36$). Obviously, though excluding the state's Catholicism in 1960 purchased additional parsimony and obviated one objection of post hoc fitting, the price was a loss of some accuracy.

3. An inspection of the errors also suggests, more generally, that some may be associated with regional trends. The specified regional trends, in several cases, appear in somewhat diluted form in states bordering those regions. Regional political trends rarely cut cleanly at state lines. For instance, in some re-

Table 5. Largest State Forecast Errors

Rank	Year	State	Actual Vote %	Expected Vote %	Absolute Error %	
1.	1952	South Carolina	50.7	65.8	15.1	R
2.	1956	Mississippi	70.4	56.5	13.9	D
3.	1960	Louisiana	63.8	51.5	12.3	D
4.	1972	Hawaii	37.5	47.5	10.0	R
5.	1984	Iowa	46.3	36.8	9.5	D
6.	1972	California	43.0	33.6	9.5	R
7.	1960	Oklahoma	41.0	50.3	9.3	R
8.	1948	South Carolina	86.4	77.1	9.3	D
9.	1956	Maine	29.1	37.9	8.8	R
10.	1948	Oklahoma	62.7	53.9	8.8	D
11.	1980	New Hampshire	33.0	41.7	8.7	R
12.	1952	Georgia	69.7	61.1	8.6	D
13.	1972	Oregon	44.6	36.1	8.6	D
14.	1948	Arkansas	74.6	66.1	8.5	D
15.	1956	Georgia	66.6	58.3	8.3	D
16.	1956	Rhode Island	41.7	50.0	8.3	R
17.	1968	Massachusetts	65.7	57.5	8.3	D
18.	1972	Oklahoma	24.6	32.8	8.2	R
19.	1956	South Carolina	64.3	56.1	8.2	D
20.	1952	Idaho	34.5	42.6	8.1	R

Note: Cases are ordered by the magnitude of the error. The votes are the Democratic presidential two-party votes. Expected and error percentages are rounded. All errors of 8.0% or larger are included.

spects, Oklahoma may have voted like a southern state. In the 1964, 1968, and 1972 elections, Democrats fell anywhere from four to eight percentage points short of their expected vote in the state and in the 1976 election, with Carter benefiting from the southern regional advantage, the Democratic vote in the state exceeded expectations by more than five percentage points. Elsewhere, northeastern states outside of New England (i.e., New York, New Jersey, and Pennsylvania) may have experienced some of the same shift toward the Democrats. Democrats in 1964 received a larger than expected vote in each of the states, winning nearly eight percentage points more than expected in New York. Similarly, though there is no consistent evidence of a home region advantage for presidential candidates (outside of that for southern candidates), some candidates may win additional "friends and neighbors" votes in nearby states. For instance, as Lewis-Beck (1985) suggested, the spill-over of the home state advantage may help explain why Iowa's 1984 vote for Mondale well exceeded expectations (+9.5%).

Predicting the Electoral Vote

While the error in the vote percentage is one measure of the equation's strength, the equation can also be assessed by its ability to forecast the party that wins the plurality of the state's vote and, therefore, all of its electoral votes. To assess this aspect of the forecast, the actual and expected presidential votes were converted to indicate only which party actually carried the state or was expected to carry it. The cross-tabulation of the actual and expected state presidential vote winners is presented in Figure 2.

As Figure 2 indicates, the equation correctly identified state presidential vote winners in 88% of the cases, missing in just 66 of the 531 elections. Over half of these misses occurred in the three closest national elections of the series: 1948 (11 misses), 1960 (14), and 1976 (10). As one might expect, most of the errors also involved very close state votes. Of the 66 errors, 43 involved state elections decided by a margin of 52% to 48% or closer and an additional 10 errors were in elections decided by 53% to 47% or less. A companion logit analysis produced similarly accurate results.

The strength of the forecast model can be assessed by the number of electoral votes correctly predicted and also by the ultimate criteria: the correct prediction of the winning presidential candidate. The average net electoral college error was 36 votes, three more than Rosenstone's equation.[25] The only election in which the aggregate expected electoral votes were wrong in predicting the

[25]The mean gross electoral college vote error was 65 votes. In calculating these errors, "faithless elector" votes were reassigned to the candidate winning the state's popular vote and electoral votes for third-party candidates were not counted. The largest errors, as expected, were in the closest elections: 1948, 1960, and 1976. The gross electoral vote errors in these three elections each exceeded 100 votes (1948, 101; 1960, 156; and 1976, 144).

Figure 2. Expected and Actual State Votes, 1948–88

		Expected Presidential Vote Winner: Democrats	Expected Presidential Vote Winner: Republicans	
Actual Presidential Vote Winner	Democrats	130 correct	34 wrong	164
	Republicans	32 wrong	335 correct	367
		162	369	$N = 531$

Correct expected winner = 465 (88%)
Expected winner incorrect = 66 (12%)

Note: Expected presidential vote winner was determined by the equation in Table 2. Expected Democratic presidential votes over 50% are counted as expected Democratic wins and under 50% are counted as expected Republican wins.

winning presidential candidate was 1960. In that election Kennedy won several large states (including Illinois, Michigan, New Jersey, and New York) that the model wrongly predicted that Nixon would carry.

Overview

Presidential election results in the states can be predicted quite accurately early in the campaign with a very parsimonious group of objective indicators. Based on polling data, economic indicators, elementary candidate characteristics (e.g., incumbency, etc.), state political history, and a few regional partisan trends, the forecast equation constructed here accounts for nearly 85% of the variance in statewide presidential votes in the more than 500 state returns of presidential elections held since the end of World War II. The average error is just three percentage points, and the equation correctly identified the probable state electoral vote winner in 88% of the 531 elections examined.

While direct comparisons to Rosenstone's earlier forecasting equation are difficult because of differences in the elections examined and in their measurements of the dependent variable in two election years (1948 and 1968), a case can be made that the equation examined here compares favorably in several respects. Although the equation is no more accurate than Rosenstone's equation, it appears to be in the same league—whether measured by the average error or the standard error of the estimate. Both models are quite accurate, though both also leave some room for improvement.

Where the above equation compares most favorably to the Rosenstone equation is in its ability to yield before-the-fact forecasts (without forecasting the

predictor variables themselves) and in its parsimony. First, unlike Rosenstone's equation, all indicators are available by mid-September of the election year. Second, whereas Rosenstone's equation involves the estimation of 74 coefficients and the construction of complicated issue dimensions, the equation here requires the estimation of only 16 coefficients (only three more than the Holbrook equation). Moreover, it does not necessitate the prior construction of intricate issue indices.

The parsimony of the equation is achieved in three ways. First, the equation relies on trial-heat polls to tap popular impressions of the two specific candidates rather than the issue indices. Second, it uses state legislative election results rather than the congressional vote (in six variables) to measure state partisanship. The more numerous state legislative elections should allow candidate-specific factors to cancel out, rather than requiring a battery of controls. Third, perhaps most importantly, the equation makes use of prior state votes. These prior state votes along with regional trend variables serve in the place of the 49 individual state intercepts. Together the prior state votes and regional trend variables measure the usual or expected state vote, given the state's past vote and any regional partisan shift. The improvement of including prior state votes should be of special importance, since the use of state intercepts implies a static state vote, an assumption that becomes increasingly dubious as equations are estimated over an extended span of electoral history.

Manuscript received 10 April 1991
Final manuscript submitted 10 August 1991

REFERENCES

Abramowitz, Alan I. 1988. "An Improved Model for Predicting Presidential Outcomes." *PS* 21:843–47.

Bean, Louis H. 1948. *How to Predict Elections*. New York: Knopf.

———. 1972. *How to Predict the 1972 Elections*. New York: Quadrangle Books.

Bullock, Charles S., III. 1988. "Regional Realignment from an Officeholding Perspective." *Journal of Politics* 50:553–74.

Campbell, James E., and Kenneth A. Wink. 1990. "Trial-Heat Forecasts of the Presidential Vote." *American Politics Quarterly* 18:251–69.

Carmines, Edward G., and James A. Stimson. 1989. *Issue Evolution: Race and the Transformation of American Politics*. Princeton: Princeton University Press.

Converse, Philip. 1966. "The Concept of the 'Normal Vote.'" In *Elections and the Political Order*, ed. Angus Campbell, Philip E. Converse, Warren E. Miller, and Donald E. Stokes. New York: Wiley.

Fair, Ray C. 1988. "The Effect of Economic Events on Votes for President: 1984 Update." *Political Behavior* 10:168–79.

Garand, James C. 1988. "Localism and Regionalism in Presidential Elections: Is There a Home State or Regional Advantage." *Western Political Quarterly* 41:85–104.

Garson, Robert A. 1974. *The Democratic Party and the Politics of Sectionalism, 1941–1948*. Baton Rouge: Louisiana State University Press.

Holbrook, Thomas M. 1991. "Presidential Elections in Space and Time." *American Journal of Political Science* 35:91–109.

Holbrook, Thomas M., and Steven C. Poe. 1987. "Measuring State Political Ideology." *American Politics Quarterly* 15:399–416.

Lewis-Beck, Michael. 1985. "Election Forecasts in 1984: How Accurate Were They?" *PS* 18:53–62.

Lewis-Beck, Michael S., and Tom W. Rice. 1983. "Localism in Presidential Elections: The Home State Advantage." *American Journal of Political Science* 27:548–56.

———. 1984. "Forecasting Presidential Elections: A Comparison of Naive Models." *Political Behavior* 6:9–21.

———. 1992. *Forecasting Elections*. Washington, DC: Congressional Quarterly Press. Forthcoming.

Newman, Edwin. [1984] 1988. "Polls, Polls, Polls." In Edwin Newman, *I Must Say.* New York: Warner Books.

Patterson, Thomas E. 1980. *The Mass Media Election*. New York: Praeger.

Patterson, Thomas E., and Robert D. McClure. 1976. *The Unseeing Eye*. New York: Putnam's.

Petrocik, John R. 1987. "Realignment: New Party Coalitions and the Nationalization of the South." *Journal of Politics* 49:347–75.

Rosenstone, Steven J. 1983. *Forecasting Presidential Elections*. New Haven: Yale University Press.

Sigelman, Lee, and David Bullock. 1991. "Candidates, Issues, Horse Races, and Hoopla: Presidential Campaign Coverage, 1888–1988." *American Politics Quarterly* 19:5–32.

Stanley, Harold W. 1988. "Southern Partisan Changes: Dealignment, Realignment, or Both?" *Journal of Politics* 50:64–88.