10

The Science of Forecasting Presidential Elections

JAMES E. CAMPBELL
University at Buffalo, SUNY

The scientific forecasting of presidential elections is a diverse field that has undergone considerable growth in recent years, as evidenced by the forecasting models in this volume. Since its rebirth in the 1980s, there has been a proliferation of forecasting models and serious scrutiny of their design and accuracy. Even so, there remains significant skepticism about the accuracy and value of the forecasting enterprise. In the 1980s, as interest in the field was being rekindled, one wag dubbed it "recreational political science." After several of the better-known models erroneously predicted a Bush reelection in 1992, some said that forecasters had much to be humble about, and political pundit and part-time wit William Schneider propounded Schneider's Law: "The models work, except when they don't work" (American Survey, 1995-1996, p. 32).

It would be tempting for the forecasting community to respond defensively to these criticisms. How is nonrecreational political science doing after all? Traditional quantitative empirical studies are safely post hoc and, as generally practiced, seldom run an immediate risk of public falsification. Although generally quantitative, most of this research is not particularly interested in precision or even the magnitude of relationships but in arriving at estimates of relationships that are probably not due to random chance. What an ambitious standard! Both punditry and qualitative political commentary are as a rule too vague (or incomprehensible) to be falsifiable at all. And as for the much-vaunted deductive

work, it seems to be too often barely tethered to political reality and smugly indifferent to whether the tether holds or not. It is, after all, high theory, though its distinctiveness from traditional armchair theorizing in some cases seems largely symbolic. From this comparative perspective, the state of election forecasting looks quite healthy. It is specific, rigorous, and usually measures up fairly well and observably to political reality.

Although this response is certainly tempting, to paraphrase President Nixon, it would be wrong. Wrong not because its claims about other areas of the discipline are inaccurate but because the contribution of forecasting ought to be measured against absolute standards and not against the shortcomings of other research. So, as former Mayor of New York Ed Koch used to ask, "How are we doing?" My answer is that the scientific area of presidential election forecasting is doing quite well. Although the difficulties entailed in predicting any human behavior can indeed be humbling, when taken as a whole, scientific election forecasting has made progress. In addition to advances in the specific goal of accurately forecasting presidential elections, election forecasting research has the potential to contribute to advances in theories regarding elections and presidential campaigns, to political methodology, and to political science more generally. This is not to say that the area does not also confront a variety of problems. It does. However, even in exploring these analytical and data problems, election forecasting brings to these issues a different and useful perspective.

The Accuracy Issue

Although there are many dimensions by which forecasting can be judged, and Lewis-Beck has nicely specified a number of these dimensions (Lewis-Beck, 1985; Lewis-Beck & Rice, 1992a), the paramount virtue of a forecast is its accuracy. However early the forecast is made, however simple and theoretical the model, an inaccurate forecast is a bad forecast. So how accurate are the forecasts?

As forecasters would readily concede to the skeptics, election forecasting has not and will never reach perfection. There are too many sources of imperfection in the indicators from which the forecast models are constructed. Sampling error alone in polls measuring rates of presidential job approval or the presidential preferences of likely voters introduces error. Moreover, sampling error is hardly the sole source of error in these polls. One would well expect significant individual errors from respondents placed in an artificial environment and asked by a stranger to characterize their attitudes, of which they may be barely conscious of, into preset categories in the survey. Similarly, anyone who has examined the array of general economic indicators such as the gross domestic and gross national products or per capita real disposable income knows that there is no perfect general indicator of the economy and that the indicators are continually revised well after the quarter in which they are released and used in the forecasting

models.[1] Moreover, the impact of any given level of economic growth may be distributed differently and therefore have different political consequences. In short, both public opinion and economic measures are aggregate assessments that imperfectly measure the attitudes and transactions of millions of Americans; even if we were absolutely certain that we had the single "right" model for forecasting elections, as a practical matter, the forecasts would be imperfect because the data used in the models are imperfect.

Although perfection for forecasting is not attainable, the record of the forecasting models looks quite good by the opposite absolute standard, the naive null model. If you knew nothing about the candidates, parties, public opinion, or the economy in a particular election or the history of voting for in-party candidates and were asked to predict the in-party's share of the two-party vote, the logical guess would be 50%. On average, since 1948, this prediction would have missed the actual two-party vote by 5.1 percentage points (a median absolute error of 4.7 percentage points). Each of the forecasting models easily outperforms this null model. The mean out-of-sample absolute errors of the forecasting models range from 1.5 to a bit over 2 percentage points. For example, my trial-heat and economy forecast model has an average out-of-sample error of 1.5 percentage points, and Abramowitz's (1988) popularity, economy, and third-term model has an average out-of-sample error of 1.7 percentage points.

That forecast models are less accurate than perfection but more accurate than a logical but totally ignorant guess unfortunately tells us precious little about the overall accuracy of the forecasts, though conventional empirical analyses generally find beating random chance enough of a reason to celebrate. In examining the state of presidential election forecasting, we need a meaningful and attainable benchmark against which the accuracy of the forecasts can be compared.

Two benchmarks that should set a high standard of accuracy for comparison to forecasts are provided by the final preelection preference polls that have been typically conducted within a few days of the election and the American National Election Study's (ANES) postelection survey that is conducted in the weeks following the election.[2] Because these are direct readings of the intentions of voters *immediately prior* to the election and reports of their actual votes *after* the election, one might expect them to be significantly more accurate than forecasts made 2 to as long as 4 months *before* the election. This, however, is not the case. In fact, the election forecasts made months before the election generally have as good or better track records for accuracy than either the polls taken within a few days of the election or the ANES's reported presidential vote from its postelection survey.

The record of November Gallup preference polls for elections since 1948 is presented in Table 10.1. From one perspective, the preelection Gallup preference polls have a fairly good record. The mean absolute error for the November poll was 2.66 percentage points. This is about half the error of the null model.

TABLE 10.1 The Accuracy of the November Gallup Preference Poll and the American National Election Study's (ANES's) Postelection Reported Vote, 1948-1996

				November Preference Poll		*ANES Postelection Reported Vote*	
Year	*In-Party*	*In-Party Candidate*	*Actual Two-Party Vote (%)*	*In-Party Preference*	*Error*	*Reported Vote*	*Error*
1948	Democratic	Truman	52.32	47.34[a]	4.98	—	—
1952	Democratic	Stevenson	44.59	45.98	–1.39	41.96	–2.64
1956	Republican	Eisenhower	57.75	59.38	–1.63	59.68	–1.93
1960	Republican	Nixon	49.92	49.48	0.44	51.04[a]	–1.12
1964	Democratic	Johnson	61.34	68.82	–7.48	67.63	–6.29
1968	Democratic	Humphrey	49.60	48.78	0.82	46.15	3.44
1972	Republican	Nixon	61.79	63.54	–1.75	64.18	–2.39
1976	Republican	Ford	48.95	50.54[a]	–1.59	49.00	–0.05
1980	Democratic	Carter	44.70	48.35	–3.65	43.64	1.05
1984	Republican	Reagan	59.17	59.38	–0.21	58.29	0.88
1988	Republican	Bush	53.90	55.91	–2.01	52.85	1.05
1992	Republican	Bush	46.55	42.47	4.08	41.64	4.90
1996	Democratic	Clinton	54.74	59.30	–4.57	57.99	–3.25
	Number of plurality vote winners wrongly predicted				2		1
	Mean absolute error				±2.66		±2.42
	Median absolute error				±1.75		±2.16
	Largest absolute error				±7.48		±6.29

NOTE: The ANES began its series of presidential election national surveys in 1952. All vote and preference percentages are of the two-party vote. The Gallup Polls are those released in November before election day. The ANES reported vote percentages are of those who also claimed some party identification (including pure independents). This differs only very slightly from the percentages reported in Miller and Traugott (1991).

a. Indicates that the national plurality vote winner was incorrectly identified.

TABLE 10.2 A Comparison of the Accuracy of the National Preference Polls to the Election Forecasting Models, 1996

	Last Preelection National Preference Polls	*Predictions Made by the APQ Forecasting Models*
Mean absolute error from the actual two-party vote percentage	2.46	2.13
Median absolute error from the actual two-party vote percentage	2.09	2.36
Standard deviation of the absolute errors	1.79	1.19
Largest absolute error	5.49	3.85
Mean number of days before election day	2	88
Number of polls or models	8	7

NOTE: The polls examined are "last pre-election, media-sponsored preference polls" as listed in *USA Today* on election day, November 5, 1996 (p. 12A). The eight polls listed are those conducted by or for *USA Today*/CNN/Gallup, ABC News, *New York Times*/CBS, *Wall Street Journal*/NBC, Reuters/Zogby, Pew Research Center, Harris, and Hotline/Battleground. According to the *USA Today* report, all of the polls "were taken through Saturday or through Sunday" before the Tuesday election. The mean number of days before the election for the polls is estimated as 2 based on *USA Today*'s statement and a Sunday poll release date. Because the errors for all of the forecasts and all but one of the polls were to overestimate rather than underestimate the Clinton vote percentage, the mean errors look very similar to the mean absolute errors. The seven forecasting models are those listed in the "Presidential Vote Forecast Compendium" of the October 1996 issue of *American Politics Quarterly*. These are the models that have been updated for this volume.

Moreover, if the wild overestimation of the 1964 Johnson vote and the miscall of the 1948 election (Gallup stopped polling in late October 1948) are dropped, the average November poll error declines to 2.01 points. Given the standard sampling error and some further slippage in exactly who will or will not vote, this is a very respectable record.

Although the November Gallup Polls have been quite accurate, the scientific forecasts made months before have been even more so. All seven forecasting models in this volume have smaller average errors than the November polls, and several have smaller errors than the polls even after discarding the November polls' worst errors.

Although most national polls do not have Gallup's long history and therefore do allow for comparisons across time, we can compare the November results of eight of the more prominent national polls to the forecasts in the 1996 election. On election day in 1996, *USA Today* published a summary of the "last preelection, media-sponsored preference polls." I have computed the two-party division of preferences for each poll and calculated its error from the actual November vote. Table 10.2 presents the summary statistics of these errors and com-

pares them to similar statistics for the forecasts of the seven models in this volume (originally published in October 1996). Despite the fact that the forecasting models typically produced a forecast more than 12 weeks before the last poll, the mean forecast error was actually one third of a percentage point smaller than the mean error of the last polls (though the median forecast error was slightly larger).[3] The standard deviations also indicate that there was a greater consensus among the various forecasts in 1996 than among the various preelection polls. In short, the forecasts look good compared to the November polls. They not only provided much earlier vote predictions but also predictions that are at least as accurate or more accurate than the final preelection polls.

The accuracy of the forecasts also measures up well against the ANES postelection report of the presidential vote. The right-hand columns of Table 10.1 present the aggregated national two-party division of the presidential vote in the ANES postelection surveys since their inception in 1952. The aggregate accuracy of the reported vote in the ANES surveys is quite comparable to that of the November preelection Gallup Polls. Both the Gallup Polls and the ANES postelection vote divisions have been fairly accurate, though neither has been as generally accurate as the forecasting models. The mean absolute error of ANES for elections since 1952 has been about 2.4 percentage points of the two-party vote. By comparison, the mean out-of-sample errors for most of the forecasting models, again offering predictions well in advance of the election and not after it, are approximately 2 percentage points or less. In short, by the two most demanding and realistic standards of accuracy, the November polls and the ANES postelection surveys, the forecasting models generally have performed exceedingly well.

THE CERTAINTY ISSUE

Although forecasting models have generally achieved an impressive degree of accuracy relative to attainable benchmarks, forecasting remains an imperfect science. As such, it is important to assess the amount of certainty accompanying any vote prediction from a model. The degree of certainty is especially important to presidential election forecasts because they are commonly based on models estimated with as few as a dozen elections. A data-mining approach to forecasting could easily produce a model that might correspond to the series of recent elections but would collapse when applied to future elections, much like the bogus bellwethers of which baseball league pennant winner (National or American) won the World Series or whether the hemlines of women's dresses were going up or down.

Evaluations of the certainty or uncertainty of forecasts produced by the models have become both more common and more sophisticated in recent elections. Not so long ago, critics were applying the conventional 95% confidence inter-

vals to the point forecasts using in-sample standard errors (Beck, 1992a; Greene, 1993). Not surprisingly, given the level of certainty specified and the small number of cases involved, the confidence interval commonly crossed the 50% vote point, indicating that one should not place much confidence in the forecast having predicted the winning candidate.

Although recognizing that uncertainty estimates are an integral part of the forecast, forecasting has moved beyond the 95% confidence interval approach in several important respects. First, recognizing the potential sensitivity of the model estimates to a few aberrant cases, the models have examined out-of-sample errors as well as other statistics that assess the robustness of their estimation.[4] The distribution of out-of-sample errors offers a more appropriate (and in no part tautological) estimate of the likelihood of errors of various magnitudes.[5] Second, recognizing that 95% confidence is essentially an arbitrary level, forecasting has attempted to gauge and report the specific level of certainty in a forecast. It is more useful to know that the forecast is 90% confident that it has correctly predicted the winning candidate than to be told that it is not 95% confident of the predicted winner. Finally, in contrast to the confidence interval approach, it is clear that the direction of possible errors matters, and this dictates a one-tailed rather than two-tailed analysis of uncertainty. When we read a forecast of a 55% vote for a candidate, we want to know the probability (based on past errors) that the model could have overestimated the candidate's vote by 5 percentage points or more (i.e., the probability that the other candidate could actually win). We are much less interested in whether the forecast underestimated the candidate's vote by 5 percentage points (i.e., that the candidate will win with a 60% plus landslide). We want to know the likelihood of the vote forecast being 5 percentage points or more *high* and not merely the likelihood that it is 5 percentage points or more off in either direction.[6]

Real progress in the forecasting field has been made in the way that the confidence in particular vote forecasts is evaluated, but other issues regarding certainty remain. These issues involve how much confidence should be placed in the various forecasting models themselves rather than the level of confidence in the vote forecast generated by a model for a particular election. Although the record of accuracy documented above would seem to be sufficient to convince all but an O. J. Simpson jury that the models warrant a high degree of confidence, there is a question of how to assess the individual forecasting models apart from their in-sample goodness-of-fit statistics (standard error of the estimate, mean absolute error, adjusted R^2). The initial approach to this issue is to examine the out-of-sample errors, whether the mean, median, or, as Neal Beck suggests, a dispersion measure of these errors such as the Press statistic.[7] In addition, examination of the models with robust regression techniques could detect sets of influence points that might go undetected by the simple out-of-sample statistics (Rousseeuw & Leroy, 1987).

As useful as these approaches are to gauging the level of confidence in a model, they only go so far. They do not incorporate the indirect evidence that can affect confidence in a model. For instance, as Campbell and Mann (1992) and Greene (1993, p. 19) observe, before the 1992 election, the third-term variable in the Abramowitz model overlapped quite closely with whether the sitting president was personally seeking reelection and that, because of this collinearity, the Abramowitz model and a nearly identical model with a personal incumbency variable would have had equally strong pre-1992 records for accuracy but would have offered very different forecasts for 1992.[8] We could only determine after the 1992 election that the Abramowitz model was stronger than its alternative (because Bush was in the unusual position of being an incumbent who was seeking more than a second term for his party), but before 1992, the strength of the very close alternative model should have signaled some additional reservations about confidence in the Abramowitz model at that time.

In the above case, consideration of ancillary evidence might have raised doubts about the model in question, but in other cases, the consideration of ancillary evidence could bolster confidence. From a very narrow perspective, the perspective dealing only with the model specification strictly defined, an analyst could examine inside-out my trial-heat and economy model estimated over the 13 elections to determine how much confidence to place in it. However, this conventional examination would unfortunately not take into account some very important information that should increase the level of confidence. As I noted in the 1996 "Polls and Votes" article, the similarity of coefficients for national variables in a state-level forecasting model, the "well-behaved" pattern of trial-heat and economy coefficients for the model estimated with polls at different times in the campaign, and the stability of the model using different general economic indicators should provide some additional measure of confidence that the strong record of accuracy for the model is not a matter of chance. The model in this larger sense is really not limited to two independent variables and 13 cases. At this point, this corroborating evidence for a model sits on the sideline and is not formally incorporated into appraisals of certainty. But it is nevertheless important information that in some way needs to be taken into account.

The issue of corroborating evidence can be cast more generally. Although some may take the perspective that the various forecasting models are in competition with each other and that over time we will be able to settle on the "true" or "right" model, from another perspective there may be any number of right forecasting models. Moreover, to the degree that these successful models have common or similar elements, our confidence in each should increase. In fact, most, if not all, of the forecast models in this volume appear to offer corroborating evidence for each other. Although 10 different economic measures were used in the seven presidential forecasting models in this volume, the important point is that each of the models included some measure of the election year economy.[9] Most

of the models (all except Norpoth and Lockerbie) include some measure of mid-summer or early fall public opinion. Finally, most of the forecasting models include presidential incumbency in some form either explicitly (as in the cases of Abramowitz, Holbrook, Lockerbie, and Norpoth) or implicitly (in the constant as a result of orienting the analysis in terms of the in-party). The fact that these models have these conceptual commonalities while using different operationalizations (e.g., GDP vs. GNP, etc.) should enhance our confidence in each of the models. Somehow, this and other corroborating as well as contrary ancillary evidence ought to find its way into systematic measures of confidence in the forecasting models.

CONTRIBUTIONS TO THE STUDY OF PRESIDENTIAL ELECTIONS

Beyond their success in reliably and accurately forecasting presidential elections, the forecasting models are positioned by their failures as well as their successes to contribute to explanations of electoral behavior and presidential campaigns. There is a complex relationship between forecasting and explaining elections. Forecasting models have certainly drawn on electoral research, drawing on the large body of research concerning both the impact of economic conditions on voting behavior (Erikson, 1989; Fiorina, 1981; Hibbs, 1982, 1987; Kiewiet, 1983; Lewis-Beck, 1988; Tufte, 1978) and the combined effects of party, issues, and candidate image assessments on candidate preferences (Kelley, 1983; Kelley & Mirer, 1974; Stokes, Campbell, & Miller, 1958). However, both with its concern for a high degree of accuracy and with the many individual differences, complexities, and timings in vote deliberations as one moves more causally distant from the vote decision itself, the forecasting enterprise has understandably focused on factors that are causally proximate to the vote itself, having direct or nearly direct effects on the vote rather than affecting the vote indirectly.[10] Forecasting, in effect, works backwards from the vote, whereas explanatory research is often more concerned about more removed or root causes and the ways in which they come to influence the vote.

The fact that forecasting models concentrate their attention near the end of "the funnel of causality," as Campbell et al. (1960) termed it, does not mean that they are removed from theoretical issues. In fact, the forecasting models offer several insights for explanatory electoral research. Both Gelman and King (1993) and Holbrook (1996), for example, have used forecasting research to explore how presidential campaigns affect election results. Although there is a good deal more to say about what the forecasting models reveal about the effects of general election campaigns, there are other insights about the electoral process gained from experience with the forecasting models. In particular, the forecasting research suggests that there is a significant presidential incumbency ad-

vantage that has not been fully appreciated by explanatory electoral research and that although partisanship may be extremely important in shaping many opinions pertinent to the vote choice, it probably has a negligible *direct* impact on election results.

Presidential Incumbency

First, although congressional incumbency advantage has been studied for the past two decades at excruciating length, this has not been the case with respect to presidential incumbency. From one standpoint, presidential approval ratings may be understood to measure the variable electoral impact of presidential incumbency, and these ratings have certainly been well studied (Brace & Hinckley, 1992; Brody, 1991; Brody & Sigelman, 1983; Sigelman, 1979). In addition, retrospective evaluations of the incumbent's performance in office are strongly affected by perceptions about the state of the economy, and, as already noted, economic influences on presidential elections have also been thoroughly studied. What is unclear, however, is whether incumbent presidents generally or certain kinds of incumbent presidents (seeking a second term consecutive for their party as opposed to a third term or more) are accorded a general advantage in seeking reelection. If presidential incumbency, like congressional incumbency, presents the candidate with opportunities to exploit for a reelection bid, are these opportunities so readily exploitable that all or almost all incumbents receive this boost?

The lack of scholarly attention to presidential incumbency as a general advantage may be based on recent electoral history that suggests that presidential incumbency may be more of an electoral liability than an asset. Of the past five president-candidates, three were defeated (Ford, Carter, and Bush). Despite these recent incumbent defeats, the forecasting models point to a significant presidential incumbency advantage, though the exact nature of this advantage is not clear.

The presidential incumbency advantage is most clearly evident in the separate forecasting models of Alan Abramowitz and Helmut Norpoth.[11] A key variable in the Abramowitz model is the "time-for-change" variable, which distinguishes between in-party candidates who are seeking a second term for their party (and usually themselves) from in-party candidates who are seeking a third consecutive term or more for their party. The coefficient for this variable indicates that this is a distinction that makes about a 4 percentage point difference to the vote. All other things being equal, an in-party candidate seeking a second term receives about 4 percentage points more of the vote than an in-party candidate who is seeking a third or fourth or even fifth presidential term for candidates of his party.

The number of consecutive terms distinction is one that could be interpreted in two ways. It may be that voters penalize the in-party seeking more than a second term and think that after 8 years of one party in office that it is "time for a change." On the other hand, voters may reward an in-party seeking a second term, thinking that 4 years is not enough time for a president to fully implement his program and that you "don't change horses in the middle of the stream." This interpretation makes "the third-term penalty" not really a penalty at all but the absence of the advantage that goes with seeking a second term.

Are third-party term seekers penalized for trying to overstay their welcome, or are second-party term seekers at an advantage? It appears to be the latter. An in-party candidate with a neutral presidential approval rating (50%) and a median level of economic growth for an election year who is seeking a second-party term is predicted by the Abramowitz model to receive about 55% of the two-party vote. An in-party candidate who is otherwise in the identical circumstances but is seeking a third term or more for his party is predicted by the model to receive about 51% of the vote. In short, the second-term in-party candidate is at an advantage, whereas candidates who seek to extend their party's occupancy of the White House for more than 8 years are at no distinct advantage or disadvantage over their opponents.

Helmut Norpoth's time-series forecasting model meshes perfectly with this interpretation. Norpoth's model indicates that the immediately preceding presidential vote for a party has a *positive* effect on the predicted vote in the next election, but that the vote for the party two elections before has a *negative* effect of about the same size. All other things being equal, the best situation for the in-party candidate is for his party to have won the last election and to have lost the election before that. This is precisely the situation of the in-party candidate seeking a second presidential term for his party.

The presidential incumbency advantage is evident in other forecasting models, only through the values on the predictor variables that incumbents have rather than through the models' specification. For instance, in my trial-heat and economy model, five of the six in-party candidates seeking a second term for their party since 1948 held poll leads over their opponents going into the fall campaign, and this was reflected in the model's forecast. President Carter in 1980 was the single exception, and Carter trailed Reagan by only a couple of percentage points in the September poll. Typically these incumbents held a commanding 60% to 40% lead in the September polls. On the other hand, the seven in-party candidates who sought a third term or more for their party typically trailed their opponent in the Labor Day polls. Of the seven in-party candidates seeking to extend their party's control of the White House beyond a second term, only Nixon in 1960 and Bush in 1988 had poll leads in the first week of September. Second-party term seekers also had better numbers than the other in-party

TABLE 10.3 The Record of Incumbency, Party Tenure, and the Personal Incumbency Advantage in Presidential Elections, 1868-1996

	Election Results		
Record of In-Party (vs. Out-Party)	*Won*	*Lost*	*Mean Two-Party National Vote*
All in-party candidates	20 (61%)	13 (39%)	52.5
Incumbent president			
Seeking election	14 (67%)	7 (33%)	54.2
Not running	6 (50%)	6 (50%)	49.7
In-party seeking			
Second consecutive term	9 (69%)	4 (31%)	55.3
Third consecutive term or more	11 (55%)	9 (45%)	50.6

NOTE: Because of the complication of the second-place finish of Theodore Roosevelt's third-party candidacy in 1912, the 1912 vote is not included in computing the mean votes, though it is counted as a loss for the in-party. Five incumbents seeking reelection had not previously been elected to the presidency (T. Roosevelt in 1904, Coolidge in 1924, Truman in 1948, Johnson in 1964, and Ford in 1976). With the 1912 election excluded, the difference in the means between both categories for the presidential incumbency and the number of terms sought were statistically significant at $p < .03$, one-tailed.

candidates in Lewis-Beck and Tien's (1996) model. They typically had higher presidential approval ratings (56.5% to 40.0%, medians), stronger first-half year economies (2.4 to 1.9), and stronger scores on the peace and prosperity index (111 to 90).

Given the overlap between the situations of incumbents personally seeking a second term and in-party candidates in general seeking a second-party term, it is not entirely clear whether the incumbency advantage is a personal advantage or tied to the number of terms a party has held the presidency. Table 10.3 presents the record for both perspectives on presidential incumbency for the 33 presidential elections since the end of the Civil War. Both the won-loss records and the average votes for in-party candidates demonstrate what appears to be a distinct incumbency advantage. Nonincumbent and in-party candidates seeking more than a second term for their party have fared no better than their opponents. However, incumbents who have run for election (reelection except for those who attained office by succession) and in-party candidates seeking only a second

term for their parties have won about twice as often as their opponents and averaged about 55% of the two-party vote.

Whether the incumbency advantage is a personal advantage or an advantage linked to an in-party still new enough to office that it is able to exploit the twin campaign themes of "change" and "stability" in seeking a second term, the forecasting research demonstrates that there has been an incumbency advantage over and above what might accrue from a favorable election year economy, and electoral theories seeking to explain voting behavior ought to take a closer look at this phenomenon.

Partisanship

A second contribution to electoral theory from the forecasting models comes in an indirect way from something that forecasting scholars have not found. Despite often-repeated claims about the decline of partisanship in the electorate, the evidence is now quite clear that most American voters are partisans and that most partisan voters vote loyally for their party's presidential candidate (Keith et al., 1992). On average, only about 8% of presidential voters are pure independents, and they have never numbered more than 10%. Among the 90% plus of the electorate that is partisan, about 80% of Democrats and 90% of Republicans typically vote for their party's presidential candidate.[12] Although the impact of partisanship on the vote of individual voters is unrivaled, partisanship is quite notably absent from the forecasting models. Not one of the seven presidential election forecasting models includes a partisanship variable among its predictor variables, yet each model has a strong forecasting record with expected votes corresponding quite closely to the actual election results. This forecasting record poses an interesting question for electoral research: If partisanship is so important to the vote choice of individual voters, how can election results be so accurately predicted without taking some form of macro-partisanship or the normal vote into account?

There are several possible explanations for why partisanship exerts such a compelling force on the choice of individual voters but can be omitted from the forecasting models without any loss in their accuracy. It may be that after taking turnout and loyalty differences between Democrats and Republicans into account, the parties are nearly evenly balanced. If so, the aggregate effect of partisanship may be negligible because one party's loyalists counterbalance the opposition's loyalists. Alternatively, it may be that the impact of partisanship is almost entirely indirect in nature, affecting the vote by affecting voter reactions to candidates and issues. Like Tip O'Neil's oft-quoted aphorism that all politics is local, it may be that all electoral politics is ultimately short term. The impact of partisanship may be embedded in the public opinion measures in the forecasting models, measures of presidential approval, or the preference polls. Finally, it

may be the case that partisanship affects the vote choice of early deciding voters and plays less of a role in the decisions of late deciders who may have delayed reaching a decision because of some dissatisfaction with their party's candidate.

Prediction and Explanation

Although forecasting has much to offer theoretically minded explanatory research on elections and voting behavior, and although that research has much to offer to research into forecasting models, it is important for those engaged in forecasting research to bear in mind that forecasting and explaining elections are *not* the same enterprise. For an explanation to be theoretically very interesting, it must offer an explanation that is conceptually separated from the phenomenon being explained. It is not particularly interesting from an explanatory standpoint to offer as an explanation of an election outcome, for instance, that a majority of voters voted for a candidate because they liked him more than his opponent. To understand the election, we want to know more. We want to know why they liked one candidate more than the other.

From a prediction standpoint, however, the analytical separation between the dependent and predictor ("independent") variables is not important. Forecasting is more concerned about the accuracy of the prediction (and making a forecast early) than in being theoretically interesting. If voters in September tell you they like one candidate more than his opponent and if there is a record indicating that these preferences tend to hold well through the campaign, then these September preferences are very useful to prediction. Those interested in a deeper understanding of the vote will want to dig deeper into the reasons for voter preferences in September (and forecasters might also want to dig deeper to obtain an earlier forecast), but if you are interested in accurately forecasting the November vote, you would not ignore what voters in September were telling you. To do so would be the equivalent of a weather forecaster refusing to look over the horizon to see a storm system heading his way. There is no reason to forecast with one hand tied behind your back in a mistaken belief that a good forecasting model must also be a good explanatory model.

To make this point more concretely, if the accuracy of a forecasting model can be improved by incorporating measures of public opinion into it (e.g., presidential approval or trial-heat polls), it would be a mistake to exclude the public opinion variable because the forecaster considered it too conceptually close to the vote itself. This might be a proper concern for an explanatory model, but it misses the point of forecasting. In this regard, Fair has excluded, I believe wrongly, measures of public opinion from his model in an attempt to make the model serve both explanatory and predictive purposes. Lockerbie also appears not to draw the prediction-explanation distinction (see page 136).

My point is simply this: Good forecasting is not the same as good explanation, but each can learn from the other. Explanatory theory has guided forecasting work, and in return, forecasting research may contribute in many ways to explanatory theories of elections—from an enhanced understanding of presidential campaign effects to an appreciation of some of the subtleties of the presidential incumbency advantage to a more refined understanding of how partisanship affects the vote.

CONTRIBUTIONS TO THE SCIENCE OF POLITICAL SCIENCE

By example, presidential election forecasting may also have some beneficial side effects on the development of scientific political science over time. As conventionally practiced, empirical political science tends to be inordinately fixated on threshold tests of statistical significance of estimated effects. Although the certainty of any finding is a very important matter, as discussed above, certainty is not an either/or matter, and greater attention should be paid to the specifics of the findings themselves.[13] The extent of an effect matters. Empirical political science as conventionally practiced is too often unnecessarily and inexcusably soft and unspecific. Empirical political scientists should pay more attention to precision in their findings, the estimated coefficients of effects, and the level of certainty in those estimates. The estimated effects of one variable on another is a matter of some measurable degree. We should be no more satisfied with a statement that the coefficient is most probably nonzero than we would be with a statement that the measure of any variable is probably nonzero. The concern for precision extends to precision in estimates of uncertainty, in how much confidence ought to be placed in the findings. Precision remains an undervalued virtue in much of modern empirical political science research.[14] Experience with election forecasting in which specificity is a necessity may help to rectify this shortcoming.[15]

A second potentially salutary effect of forecasting on scientific political science is that it may stimulate greater interest in real-world, observable, aggregate political phenomena. Whether a result of borrowing from social-psychological research, methodological innovations, the availability of survey research data, or perhaps a belief that less directly observable phenomena are more theoretically interesting, empirical political science research in electoral politics has paid disproportionate attention both to unobservable political subjects (opinions rather than behaviors) and to individual-level subject matter. It sometimes seems that many analysts believe that the only reason to examine aggregate data is to make ecological inferences about individual-level phenomena

when individual-level data are unavailable. It would be healthy to redress these imbalances, to examine aggregate as well as individual-level political phenomena (as well as their interrelationship), and to examine overt behavior as well as unobservable attitudes and beliefs. Understanding elections is as important as understanding the vote choice and the attitudes of individual voters. The focus of forecasting on the very real and aggregate subject of election results may be a partial corrective to this problem, encouraging political science, to at least some degree, to "get real." Generating greater interest in real aggregate political phenomena may also introduce more of the "political" in a field that sometimes seems perilously close to becoming a subsidiary of either social-psychology or, heaven forbid, economics.

To the extent that election forecasting models influence the way in which other research in political science is conducted, political science research will more often examine concrete political subjects, will be more specific in this analysis, will become more sophisticated in its appraisal of the certainty of its findings, and will be more complete in its explanation of the phenomena under scrutiny. The scientific forecasting of elections has the important potential to influence the development of a more hard-nosed and scientific political science.

NOTES

1. The revision in economic indicators raises an interesting unresolved question for the forecasting models. The question is whether the original at-the-time economic measures should be used in estimating the models or whether the later revised measures should be used. The former have the virtue of being obtainable at the time of an actual forecast, but the latter presumably have the virtue of being more accurate about what actual economic conditions were at the time of the forecast (Fair, 1996).

2. Though less formal, another benchmark for comparison is the set of predictions made by pundits before the election. In his analysis of predictions by pundits, Rosenstone (1983) generally found that as a group, their record for accuracy was not very strong. As I noted in my 1996 analysis, the trial-heat and economy forecast of the vote made 2 months before the election was more accurate than 9 and as accurate as another 5 of 15 pundits surveyed within a few days of the election (Broder, 1992). As a response to Schneider's Law, this suggests a new law regarding forecasts by pundits: "Punditry is usually plausible, but seldom very accurate."

3. Although the difference of mean errors for the November polls and the forecast models suggests that the forecasts may be more accurate, the difference is not statistically significant. By the same token, the difference of the median errors favored the polls slightly. In short, the most prudent conclusion in examining the 1996 evidence is that the forecasts made months before the election are at least as accurate as the polls taken in the last few days before the election.

4. I have examined my model with least median squares regression, a form of robust regression analysis with accompanying diagnostics for influence cases (see Rousseeuw & Leroy, 1987). Lewis-Beck and Tien (in this volume) examine Thiel's *U* and employ the Kruskal-Wallis test in revising an earlier model.

5. In using out-of-sample errors to assess the degree of uncertainty in any model, there are several remaining issues. First, there is the issue of what distribution of the out-of-sample errors ought to be used, an assumed normal distribution or the actual observed distribution.

In the first approach, a standard error is computed from the out-of-sample errors and applied to an assumed normal distribution to calculate the probability of the forecast being wrong by whatever number of percentage points. In the second approach—the approach that I favor because of its weaker assumption of the error distribution—the size of the hypothesized error is compared to the known distribution of errors. For example, in a forecast of 55%, the model would have to be wrong by 5 percentage points or more to have incorrectly identified the winner. The approach offers a clear way to assess confidence in a vote forecast. To judge how likely an error of this magnitude is, we would examine the history of the model's errors to determine how often it had been off by 5 percentage points or more. If 1 out of 12 out-of-sample errors was this large, then we would be 92% confident (11 divided by 12) that the model had predicted the winning candidate. Standard interpolation techniques could also be used to fill in the distribution and obtain more precise estimates of uncertainty. Second, the consideration of out-of-sample errors in estimating levels of uncertainty should take into account in some way systematic variation in these errors. As Neal Beck points out in his essay and as I observe in my 1996 forecast article (see note 16), because of uncertainty in the unstandardized coefficients, errors tend to be greater as an election's values on the independent variables move further away from the mean values of those variables. If the second approach using the distribution of known errors is used, these errors could themselves be "modeled" to determine when they are likely to be larger or smaller than average.

6. The first noted change, moving to the out-of-sample rather than in-sample error distribution, is a conservative change because out-of-sample errors are larger. The third change, moving from two-tailed to one-tailed estimates of certainty, is a liberal change because it increases confidence expressed in the point estimate.

7. One should carefully note that the Press statistic suggested by Beck is a measure of dispersion in the errors *from zero* and not from the mean error as would be produced by a conventional measure of variance.

8. Greene (1993) examines the effect that small specification changes make in several models and rightly concludes that our confidence in the reliability of a model should suffer if small specification changes greatly affect a model's accuracy or forecast. My concern about Greene's warning is that it is unclear what constitutes a "small change" in a model's specification. For instance, when Greene reanalyzed my state-level presidential election forecasting model after dropping an entire set of regional variables, I regarded his reanalysis as testing an essentially different model and not making a small change in my model.

9. The 10 economic indicators were either objective indicators of economic activity or indicators of public attitudes about the economy. The most common indicator was the

gross domestic product, though it was measured alternatively as growth from the second quarter of the previous year to the second quarter of the election year (Norpoth), growth in the first half of the election year (Abramowitz), and growth from the first to second quarter of the election year (Campbell). Other objective measures were first-half gross national product (GNP) growth (Lewis-Beck and Tien), real disposable income growth from the second quarter of the previous year to the second quarter of the election year (Lockerbie), the inflation rate from the second quarter to the second quarter (Norpoth), and a cumulative index of the leading economic indicators (Wlezien and Erikson). The survey-based measures of public reactions to the economy are an index of opinion about personal finances (Holbrook), an index that includes attitudes regarding the parties' abilities to ensure national prosperity (Lewis-Beck and Tien), and an index of "personal prospective economic outlooks" (Lockerbie).

10. Forecasting models have not fared well when they have attempted to push into more causally distant explanatory variables, as Lewis-Beck and Rice (1992a) did in including variables for the in-party's unity, as measured by a dichotomous index of its nominee's percentage of primary votes, and the relative strength of the parties, as measured by the in-party's House seat losses in the prior midterm election. The Lewis-Beck and Rice model, which also included the July presidential approval rate and the first-half election year GNP growth rate, was off the mark in predicting that Bush would probably win the 1992 election with 51.5% of the two-party vote. It could be argued that their indicators of party unity and the relative strengths of the parties were not appropriate measures. The change in the reliance on primaries after party reforms of the 1970s made the party unity measure dubious. On its face, classifying Lyndon Johnson's 1964 coronation as Democratic standard-bearer as a case in which the candidate did *not* have his party united behind him should raise serious doubts about the measure's validity across time. Given that midterm seat losses are most affected by the size of the prior presidential victory (Campbell, 1997), reading them as signs of a party's strength is highly questionable. Even if proper measures of the party unity and party strength concepts were used, however, they probably would not be appropriately included in the forecast model. Although both party unity and party strength undoubtedly affect the vote, their impact is quite likely to be indirectly felt through the presidential approval measure. The impact of an in-party that is torn apart over a candidacy or is weakened relative to the opposing party will most likely be registered in low presidential approval ratings.

11. The incumbency advantage is also explicitly recognized in the models of Fair (1996) and Lockerbie (this volume) and in the Lichtman's (1996) nonstatistical "keys to the presidency" forecasting scheme.

12. The distribution and loyalty rates of partisans have been computed from the presidential election year American National Election Studies from 1952 to 1996. As Keith et al. (1992) persuasively argue, so-called "leaning independents" are counted as partisans.

13. The neglect of the exact coefficients estimates can be illustrated by three commonly limited forms of analysis. Although less frequently the case now, it was once common for empirical analysts to focus on standardized coefficients to the exclusion of the unstandardized coefficients. Similarly, when logit was first coming into common usage in the late 1970s, findings were routinely reported in a way in which one could not directly determine the effects of the independent variable on the probability of a case mov-

ing a category in the dependent variable. Finally, it was often the case that the statistical significance of coefficients would be reported as an either/or matter, often using a simple asterisk to report a coefficient as being significant at some preordained level.

14. An increased concern for precision would also have the beneficial effect of raising awareness of and attention to measurement problems and differences. Because measurement errors tend to diminish the accuracy of forecasts, forecasters are especially concerned about improving the reliability and accuracy of the measurement of variables used in the models. This concern for specificity and measurement accuracy is also consistent with forecasting's interest in aggregate, real-world phenomena because there are often independent and observable alternative measures of real-world phenomena. For instance, only by examining the actual national presidential vote results would one know that the ANES study in 1996 overstated the Clinton vote percentage by fully 3.6 percentage points of the total vote.

15. Another beneficial potential influence of forecasting on empirical political science is its emphasis on completeness in explaining variance in the dependent variable. Some research focuses on an independent variable or on a hypothesized effect of an independent variable on the dependent variable. The result of this is a scattered, unintegrated, and incomplete research on the dependent variable. Election forecasting focuses attention squarely on the dependent variable of the election results.