CHAPTER 19

FORECASTING US PRESIDENTIAL ELECTIONS[1]

THOMAS HOLBROOK

Never make predictions, especially about the future
Attributed to Casey Stengel[2]

We should have gotten our brains beat
Karl Rove, following the 2000 election[3]

As suggested by Casey Stengel's admonition and Karl Rove's observation on the 2000 presidential race, predicting the future can be a difficult and perhaps surprising enterprise. As one who has proffered very public and occasionally very wrong forecasts of presidential election outcomes, I can also attest to the fact that predicting is not always for the faint of heart.[4] Yet occasional misfires by a couple of election forecasting models do not render the entire enterprise of no use. In fact, given the lead time with which most forecasts are made, along with the limited

[1] This chapter was written in the spring and summer of 2008, well before the outcome of the presidential election was known.

[2] http://thinkexist.com/quotation/never_make_predictions-especially_about_the/154388.html

[3] Berke (2001).

[4] See Lewis-Beck (2005) for an example of the type of "citizen response" forecasters were exposed to during and after the 2000 campaign.

number of predictors used (see below), the vast majority of academic forecasting models are surprisingly accurate.

In this chapter I provide an assessment of the current state of presidential election forecasting models. Although congressional election models are also quite prevalent (see Abramowitz 2002, 2006; Bardwell and Lewis-Beck 2004; Erikson and Bafumi 2002; Klarner and Buchanan 2006a, 2006b; Lewis-Beck and Rice 1992), and while the history and evolution of academic election forecasting is interesting and instructive (see Campbell and Lewis-Beck 2008; Lewis-Beck 2005; Lewis-Beck and Rice 1984, 1992), I direct my attention to presidential forecasting models from the last three election cycles. I begin the chapter with a look "under the hood" and discuss the specifics of the most widely known models from the 2004 election; I then consider the predictions made by these models and evaluate the determinants of forecasting accuracy from 1996 to 2004; next, I examine lessons learned from the 2000 campaign; finally, I examine alternatives to the dominant aggregate-national forecasting models: electronic markets, citizen forecasts, and state-level forecasting models.

Election forecasters have encountered some level of criticism over both method and (lack of) theory (Beck 1992; Colomer 2007; Greene 1993; Tetlock 2005). Campbell (1993, 2008a, 2008b) provides a spirited defense of the forecasting enterprise, pointing out that the forecasting community has generally been very responsive to methodological criticisms and that most models are based on a particular theoretical framework: retrospective voting (Fiorina 1981). Indeed, the charge that forecasting models are void of theory strikes me as particularly ludicrous. Given that most models (below) are based on the assumption that the electorate rewards the incumbent party for good times and punishes it for bad times, it would appear that forecasters are on fairly solid footing. Having said this, I do think that forecasters are subject to criticism regarding their selection of indicators of important theoretical concepts. In the end, as Campbell (2008b) points out, the goal of forecasting models is not to generate or test grand theories but to predict election outcomes. Nevertheless, it seems to me that one way to have both interesting and accurate forecasting models is to base them on theoretically plausible assumptions about the electorate, which is the case for most models discussed in the next section.

The Forecasting Models, 2004

Perhaps the best way to get a sense of what constitutes contemporary forecasting models is to examine a group of models developed and made public prior to the 2004 presidential election. Table 19.1 summarizes many of the key aspects of nine

Table 19.1 A Summary of Presidential Forecasting Models, 2004

Author	Economic Variables	Political Variables	Trial-Heat Variables	# Obs.	Predicted Vote Bush	Error
Abramowitz	GDP Change (first 2 qtrs. of election year)	Approval		1948–2000 14 obs.	53.7%	2.4
		Time for Change				
Campbell	GDP Growth (2nd qtr)		Early Sept. Poll	1948–2000 14 obs.	53.8%	2.5
Fair (October)	GDP Growth (first 3 qtrs. of election yr)	President's Party		1916–2000 22 obs.	57.5%	6.3
		Incumbency				
	Inflation	Time for Change				
	GoodNews (# qtrs. with growth > 3.2%)	War				
Holbrook	Personal Finances, adjusted for Economic news	Approval		1956–2000 12 obs.	56.1%	4.9
		Time for Change				
Lewis-Beck and Tien	GNP Growth (4th to 2nd qtr. of elect. yr)	Approval		1952–2000 13 obs.	49.9%	1.3
		Incumbency				

	GNP Growth (x) Incumbent					
	Job Growth					
Lockerbie	Prospective Personal Finances (1st qtr)	Time for Change		1956–2000 12 obs.	57.6%	6.4
Norpoth		Primary Support		1912–2000	54.7%	3.5
				23 obs.		
		Partisan Baseline				
Wlezien and Erikson	Leading Economic Indicators (LEI) Growth: through 13th qtr of sitting President's term	Approval	August Trial Heat Polls (only in 2nd model)	1952–2000 13 obs.	52.9%	1.7
					51.7% (model 2)	0.5
Cuzán and Bundrick	GDP Growth (first 3 qtrs.)	Incumbency		1916–2000 22 obs.	51.1%	.14
		Time for change				
	Inflation					
	All News (Fair's Good News not adjusted for war years)					

different models developed by forecasters who have proffered their prognostications for the past several presidential elections.[5] All of the models are popular vote models and focus on predicting the two-party division of the national votes. While there is much overlap among the forecasting models, there is also considerable diversity. One overriding characteristic of the models—alluded to above—is that most of them are retrospective in nature; that is, most of them incorporate measures of economic or presidential performance and assume an electorate that rewards presidents for good outcomes and punishes presidents for bad outcomes. Indeed, four of the nine models include a measure of presidential approval, and eight out of the nine incorporate at least one measure of economic performance. Another predictor held in common by five of the nine models is some form of Abramowitz's (1988) "time for change" variable, which indicates whether or not the incumbent party has held the White House for at least two consecutive terms. Somewhat less common are measures of trial-heat poll results (two models) and controls for whether or not the incumbent president is running (three models). The one model that stands out from the others is Norpoth's (2004), which does not share any variables in common with the other models. Instead Norpoth's forecast is based on how well the candidates fared in their party's New Hampshire primary, lagged presidential election results, and a partisan baseline.

Table 19.1 also shows some variation in how far in advance of the election the independent variables are measured, as well as the number of elections upon which the forecasts are made. In terms of timing, the Fair and Cuzán and Bundrick models measure gross domestic product (GDP) growth over the first three quarters, meaning that their data are measured through the end of September and are not available until late October. At the other extreme, Norpoth stands out again, with his independent variables measured by late January. Most of the variation in the number of data points used across the models is a function of whether or not the model includes survey-based measures, such as presidential approval, economic perceptions, or trial-heat polls, most of which are not available for elections prior to 1948, at the earliest. Those that do not use such measures (Fair, Cuzán and Bundrick, and Norpoth) incorporate data from many more elections than those that do.

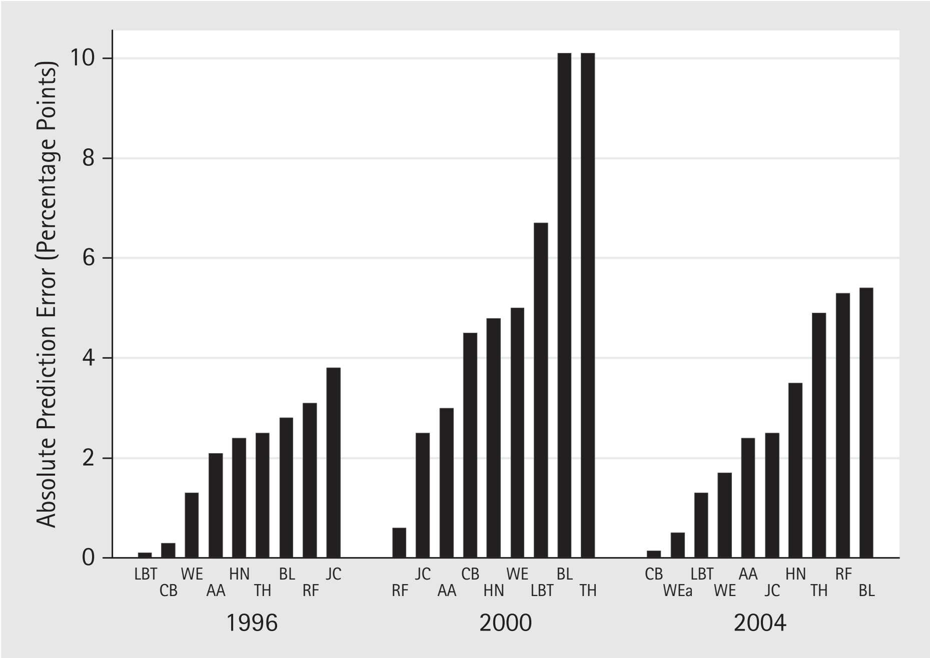
[5] With the exception of the Fair (2004) and Cuzán and Bundrick (2005) models, all others were presented at a roundtable discussion at the 2004 meeting of the American Political Science Association and published in the pre-election issue of *PS: Political Science and Politics* (Abramowitz 2004; Campbell 2004; Holbrook 2004; Lewis-Beck and Tien 2004; Lockerbie 2004; Norpoth 2004; Wlezien and Erikson 2004).

Finally, there is also appreciable variation in the accuracy of the prediction generated by the forecasting models. The most accurate forecasts in 2004 were Cuzán and Bundrick and Wlezien and Erikson, off by .14 and .5 percentage points, respectively; and the least accurate predictions were generated by Fair and Lockerbie, off by 6.3 and 6.4 percentage points, respectively. Overall, the average absolute error for the 2004 models was 2.96 percentage points.

Forecasting Accuracy, 1996–2004

Figure 19.1 provides a broader descriptive look at forecasting accuracy from the 1996, 2000, and 2004 presidential elections. The top panel of the figure displays the absolute error in the forecasts, and the bottom panel displays the real value of the error. Several interesting findings are revealed here. First, there is considerable variation in forecasting accuracy both within and across elections: the average absolute error and variation in absolute error were greatest in 2000 and lowest in 1996. The bottom panel reveals another interesting pattern regarding the direction of error: in all three election cycles, the vast majority of the predictions (twenty-five out of twenty-eight) *overestimated* the expected vote of the incumbent party. One possibility here is that there is some sort of systematic bias in the models that leads to erroneously positive expectations for the incumbent party. The source of this bias remains a mystery and is something about which forecasters should be aware. One other finding that is not apparent from these data is that the models do a very good job of calling the popular vote winner: in twenty-seven of twenty-eight cases the forecasting models predicted the correct winner of the two-party vote.[6] Even in 2000, when the average error was relatively high, every model called the popular vote for Gore. This may seem like a minor point, since presidential elections are won with electoral votes but, but in all fairness forecasting models do focus on predicting the outcome of the two-party popular vote.

[6] Lewis-Beck and Tien's 2004 model predicted a Bush vote share of 49.9 percent of the two-party vote.



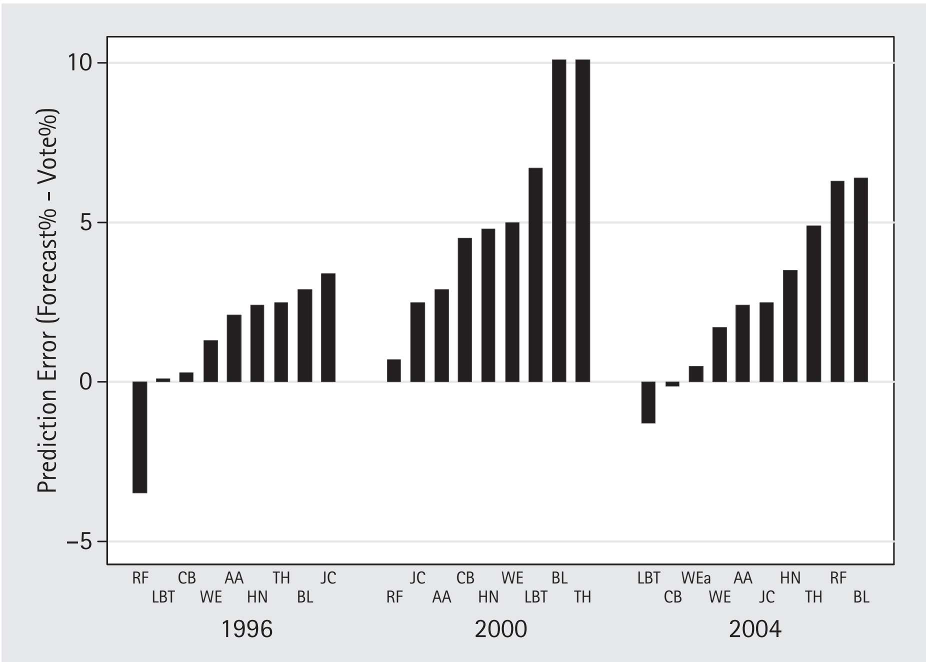


Figure 19.1 Forecasting accuracy in US presidential elections, 1996–2004

Note: AA=Alan Abramowitz; BL= Brad Lockerbie; CB=Cuzán and Bundrick; HN=Helmut Norpoth; JC=Jim Campbell; LBT=Lewis-Beck and Tien; RF=Ray Fair; TH=Tom Holbrook; WE(a)=Wlezien and Erikson

What Makes for a Good Model?

The data in Figure 19.1 also show that there is no single model that is always the "best" model. Given the above description of the error rates in 2004 (Table 19.1), one might be tempted to conclude that the Cuzán and Bundrick and Wlezien and Erikson models are the "best" models, while the Fair and Lockerbie models are the "worst." However, this requires us to pin everything on only the most recent election when we know the models show different levels of accuracy in other elections. For instance, in the difficult-to-call election of 2000, Fair's model came within seven-tenths of a percentage point of the actual vote while Cuzán and Bundrick and Wlezien and Erikson missed by 4.5 and 5 percentage points, respectively. The point here is that judging models based on single elections is hardly a sound strategy for understanding what constitutes a good forecasting model.

Lewis-Beck (1985, 2005) argues that accuracy, parsimony, reproducibility, and lead time (with more weight assigned to lead time) should all be taken into account when evaluating a model. I have no problem with this, in principle, as these are all important characteristics. However, Campbell (2008b) argues that Lewis-Beck's indicator of quality is imperfect, primarily because rearranging the same components and assigning more weight to accuracy than lead time produces very different results. For instance, using Lewis-Beck's conceptualization, Campbell's 2004 forecast is ranked sixth (out of seven forecasts), whereas assigning more weight to accuracy than to lead time vaults Campbell's forecast to first among the seven models. Again, though, I would argue that the larger problem is not how you arrange the components but the fact that we are evaluating forecasting models on the basis of single election outcomes.

One alternative to focusing on which single model is "best" is to instead turn to an analysis of the types of model characteristics that lead, on average, to greater accuracy in predictions. In other words, given the variation in forecasting accuracy, why are some models more accurate than others? This is the task to which I now turn. However, in order to avoid the pitfalls of relying on results from just one election I examine accuracy across the last three elections.

Determinants of Accuracy

In order to assess the determinants of forecasting accuracy I use the absolute errors in forecasting presidential elections from 1996 to 2004 (presented in the top panel of Figure 19.1) as a dependent variable and then model that error as a function of model characteristics that are expected to provide a good statistical account of it. Although this analysis is somewhat limited due to the number of forecasts, and also because of the general similarities among models, there is enough variation both in error and model characteristics that a plausible model can be developed and evaluated. In addition, it should be pointed out that while some of the models retained the same characteristics across all three elections, many of them incorporated modest changes from year to year.

Lead time. It is important to consider lead time because models based on data measured more proximate to the election should generally be the most accurate, as they are more likely to capture the actual conditions surrounding the election. There is appreciable variation in lead time, ranging from the Fair and Cuzán and Bundrick October models, which include a measure of GDP growth in the first three quarters (measured in September), to Norpoth's 2004 model which relies on the outcome of the New Hampshire primary, measured in January. Here, the Fair and Cuzán and Bundrick models are given a score of 2 on lead time since the latest time of measurement for their model occurred two calendar months prior to the election, and Norpoth's 2004 model is given a value of 10, since his latest measurement was ten months prior to the election. The median lead time (time of last measurement) across all twenty-eight estimates is 3.5 and the mean is 3.96.

Retrospective Elements. As previously mentioned, most of the models are retrospective in nature, including, in some combination, measures of the economy, presidential approval, and "time for change". In order to assess the importance of these variables for forecasting accuracy, I include separate dichotomous indicators for whether the models included a measure of presidential approval or a measure for time for change.[7] Across all three election cycles 46 percent of the models included a measure of presidential approval and 54 percent included a "time for change" measure.

Data Points. Theoretically, models based on more data should involve less error, since they incorporate a wider range of electoral experiences and the impact of single elections is minimized. Once again, there is considerable variation in sample size, ranging from ten observations for Lockerbie's 1996 model to twenty-three observations for Norpoth's 2004 model. The mean and median numbers of observations (sample size) across all three years are 15.2 and 13 observations, respectively.

Year of Election. Of course, the number of observations available for analysis increases naturally with each election cycle. Because of this, and also to control for year-specific factors not captured in the models, I also add dummy variables for 1996 and 2000.

[7] Norpoth's 2000 and 2004 models are the only ones that did not include a measure of the economy, so adding a control for the presence of an economic variable would be equivalent to a control for Norpoth.

Table 19.2 Determinants of error in forecasting models, 1996–2004

	b	s.e.	t-score	$p \leq$ (one-tailed)
Lead Time	0.423	0.201	2.11	0.024
Time for Change	1.568	0.836	1.88	0.038
Approval	−1.548	1.036	−1.49	0.075
Sample Size	−0.225	0.119	−1.89	0.037
1996	−1.045	1.021	−1.02	0.159
2000	1.957	0.993	1.97	0.031
Constant	4.734	2.475	1.91	0.035
N=28 S.E. = 2.15 Adj. R^2=.346				

Note: Variables described in text

Table 19.2 presents the findings from the forecasting accuracy model.[8] Here we see that some of the expectations were borne out and some not. The slope for lead time suggests that for every additional month prior to the election that a model's last measurement was taken, error can be expected to increase by .42 percentage points. This means, for instance, that models using data measured in June can be expected, all else held constant, to have 1.3 points more error than models whose last measurement is in September. According to Lewis-Beck (2005) one of the virtues of a good forecasting model is greater lead time. The evidence presented here shows that this emphasis on lead time must be balanced against a very real cost in forecasting accuracy.

The results also illustrate the virtue of basing forecasts on as many observations of past electoral behavior as possible. For every additional data point used in the models, forecasting error declined by .23 percentage points. One could argue that this coefficient does little more than distinguish those models that rely on survey-based measures, such as presidential approval and trial-heat polls—which are limited to data from 1948 (at the earliest) onward—from the models that rely on aggregate economic and political data (Fair, Norpoth, and Cuzán and Bundrick). However, three points undermine this argument. First, the model also includes a control variable for whether the forecast was based on a measure of presidential approval, so this distinction is already adequately captured. Second, when the variable for number of observations was dropped in favor of a dichotomous

[8] Given the left-censored nature of the dependent variable (absolute errors cannot be less than zero) I also ran the model presented in Table 19.2 as a tobit model. The results of the two specifications were very similar, except that the tobit model produced coefficients with slightly higher t-scores.

variable distinguishing between the two groups of models, the coefficient was slightly less significant (t=-1.82) and the model fit declined slightly (adjusted R^2=.335). Finally, it is hard to argue that more data is not a good thing.

The evidence for the role of retrospective variables is a bit mixed. On the one hand, if one were willing to stretch the bounds of the traditional limits on statistical significance (as one might do with such a small sample), there appears to be a significant influence from presidential approval. The error from models that include a measure of approval is, on average, 1.53 points lower than models that do not control for approval. On the other hand, models that include some form of the "time for change" variable, on average, have about 1.59 percentage points greater error than other models. This finding is a bit at odds with expectations and suggests that the focus on incumbency-oriented effects does not always lead to less error, or perhaps that the "time for change" concept does not fit as squarely within the retrospective framework.

To reiterate, the purpose of this analysis is not to declare one model as "best" but rather to highlight the characteristics of models that generally lead to less forecasting error. Based on these results, models that have a relatively short lead time, include presidential approval, maximize the number of observations from past elections, and do not include a measure of "time for change" typically generate less error in election forecasts.

Lessons from 2000

The analysis above reinforces the exceptional nature of the 2000 presidential election. At least in recent history it stands out as the most difficult election—on average and across models—to predict; and although the data in Figure 19.1 illustrate that the 2000 election was more a disaster for some forecasters (me, for instance) than for others, the average error was large enough that it seems reasonable to be open to lessons from the 2000 experience. Two general explanations have been offered for the discrepancy between results and prediction: one that focuses on campaign strategies and another that focuses on the role of incumbency. I, along with others (Campbell 2001; Holbrook 2001; Wlezien 2001), have argued that at least part of the explanation rests with Al Gore's failure to embrace the Clinton–Gore record and reinforce retrospective voting. At the same time, a case can be made that the retrospective cue may generally be weaker when the president himself is not on the ticket (Campbell 2001; Lewis-Beck and Tien 2001; Nadeau and Lewis-Beck, 2001). Indeed, Campbell (2001, 2008a) argues in favor of only giving half weight to presidential performance variables when the vice president rather

Jan Leighley 19-Leighly-Chapter19 Page Proof page 357 10.8.2009 7:28pm

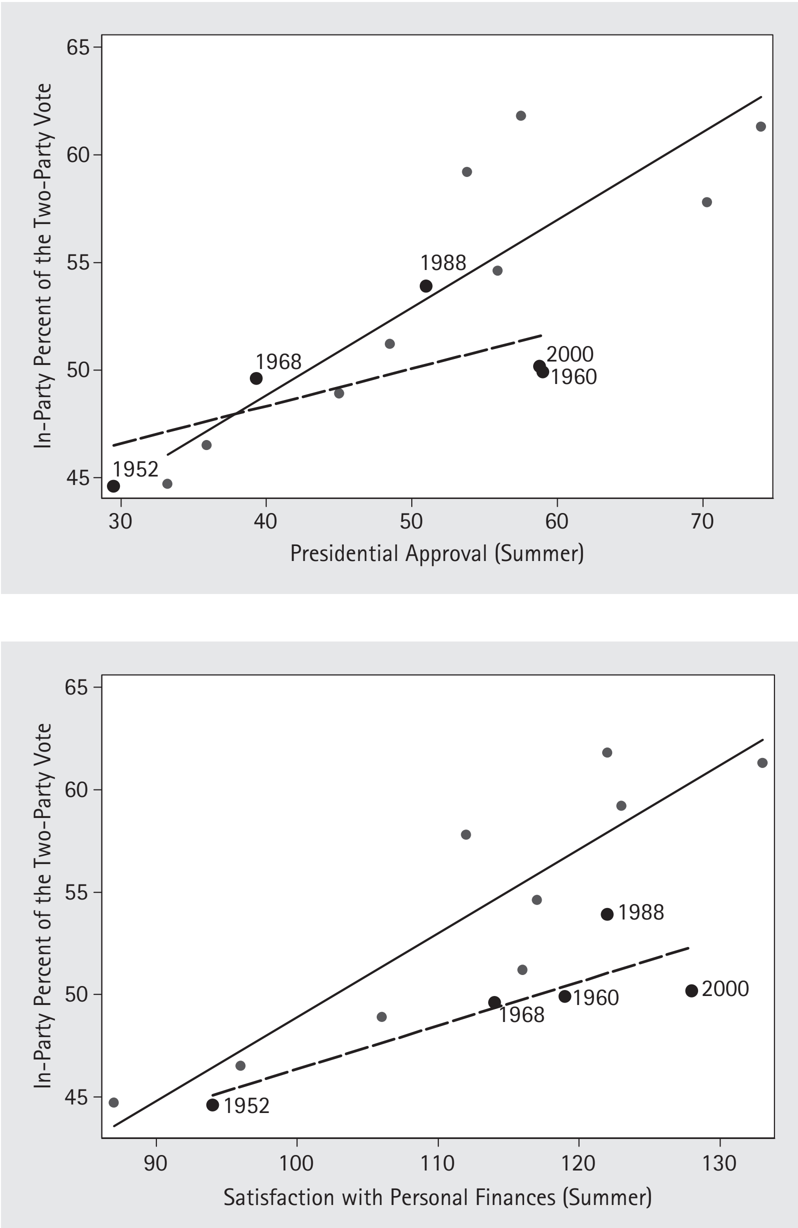


Figure 19.2 The effects of presidential approval and personal finances on support for the incumbent party, 1952–2004

than president is representing the incumbent administration. My own view is that Campbell's general idea is right but that there is no strong reason for assuming the half weight is the right factor. This issue takes on more importance as the 2008 election—an election in which neither the president nor vice president will be on the ticket—draws near. This election represents the first since 1952 in which the race can truly be considered an open race, with no representative from the incumbent administration.

Rather than assuming the in-party candidate gets only half blame (or credit) for presidential performance when the president is not running, it is probably better to let the data decide the appropriate weighting. As a means of assessing the importance of controlling for incumbency, I test a simple forecasting model, based on my own work, that controls for whether the contests involve an incumbent president or not. I use two variables as predictors of the incumbent party share of the two-party vote from 1952 to 2004: the average rate of presidential approval and the average level of consumer satisfaction[9] in the summer before the presidential election.

Following the logic articulated above, the expectation is that the relationship between presidential approval, personal finances, and the vote should be weaker in non-incumbent years than in years when the incumbent is running. The potential for these asymmetric effects is shown in Figure 19.2 where the top panel shows the plot of the relationship between presidential approval (summer) and the percentage vote for the incumbent presidential party. Election years with no incumbent running are highlighted with year labels and regression prediction lines are generated separately for incumbent (solid) and open seat (dashed) contests. At first glance, there appears to be a meaningful difference in the impact of approval on vote shares, with approval translating more readily into votes in incumbent contests than in open contests. One reservation that must be acknowledged however is that much of this difference seems to be driven by the 1960 and 2000 elections. Indeed, the 1952, 1968, and 1988 elections do not appear to follow a distinct pattern. But still, the 1960 and 2000 elections constitute 40 percent of the open-seat contests and can, therefore, hardly be dismissed as oddball, or atypical.

The second panel of Figure 19.2 illustrates the same pattern for the relationship between personal finances and vote shares. Here, we see a more clearly different pattern among the open-seat contests, where there is an appreciably shallower slope (taking into account the interaction) than in the incumbent contests. Again,

[9] The measure of aggregated personal finances is taken from the monthly survey of consumers (University of Michigan) and is measured as 100 plus the difference between the percentage saying they are better off financially than a year ago, and the percentage saying they are worse off financially. The measure of presidential approval is taken from the Gallup Poll and is the percentage responding "approve" when asked, "Do you approve or disapprove of the way is handling his job as president?" Both measures are averaged across June, July, and August. For 1952, there was no summer measure of personal finances, so the value from May was used.

Table 19.3 The conditioning effect of incumbents on retrospective elements of a forecasting model

	b	T	b	t	B	t
Personal Finances	0.41	6.06				
Personal Finances*Open Seat	−0.20	−1.58				
Approval			0.41	5.00		
Approval*Open Seat			−0.23	−1.58		
National Conditions					0.42	6.42
National Conditions*Open Seat					−0.23	−1.98
Open Seat	17.18	1.20	8.85	1.18	14.34	1.57
Constant	7.90	1.03	32.56	7.37	21.20	4.09
N	14		14		14	
Adj. R^2	0.78		0.70		0.80	
S.E.	2.72		3.19		2.62	
Prob. F>	.002		.0003		.0002	

Note: Variables described in text

this scatter-plot suggests a somewhat lower level of accountability for the incumbent party in years in which the president is not on the ticket.

These visual suggestions are put to the test in Table 19.3, which displays the results of a regression analysis in which the independent variables are interacted with a dichotomous variable scored 0 for incumbent contests and 1 for open-seat contests. If there is really less accountability during open-seat contests, the interaction slope should be negative. The first two models present the slopes and t-scores for the approval and personal finance equations run separately. In both cases we see a reduction in accountability by roughly the same proportion: the slopes for both approval and personal finances are cut just about in half in open-seat contests. And, while the interaction slopes are only marginally significant, it might be good to exercise a bit of latitude in interpreting significance, given the small number of cases, as well as the relatively high level of collinearity (tolerances range from .01 to .06).

The third model is a combined model in which both approval and personal finances are transformed so that each observation is expressed as a percentage of its highest value in the data set, and then averaged together to provide a measure of national conditions that reflects both presidential approval and satisfaction with personal finances. Once again, we see that the impact of the personal finances and presidential approval is much greater in incumbent races than in open-seat contests. In this case, the interaction coefficient is safely within the standard limits of statistical significance.

These results fit quite nicely with Campbell's contention that forecasting models need to account for incumbency when using "accountability" variables. It also reflects the same sentiment found in Nadeau and Lewis-Beck's (2001) individual-level work, which found that voters generally assigned less blame or credit to the incumbent party during open-seat contests. At the same time, these results should probably be regarded with some level of caution, as they are based on a single forecasting model, with a limited number of observations. Indeed, when the same analysis is done substituting second-quarter GDP growth for perceptions of personal finances, there is no incumbency effect. This may suggest that conditional effects from incumbency are more likely when using survey-based measures such as presidential approval and perceived personal finances, though this analysis is too limited to make that determination. The bottom line is that forecasters should check to see if the reward–punishment aspects of their models are sensitive to the conditioning influence of incumbency.

Political Markets

Electronic political markets represent an important alternative to traditional forecasting models and are also viewed as an alternative to trial-heat polls. These markets operate similarly to the standard stock markets, with investors able to buy and sell shares of certain political outcomes rather than publicly traded companies. The "granddaddy" of the political markets (modern markets,[10] at any rate) is the Iowa Electronic Markets (IEM), created by a group of faculty at the University of Iowa prior to the 1988 election. Although many other markets have sprung up in the meantime, including those that predict outcomes in sports, current events, and entertainment,[11] the IEM remains the standard for political markets and is the focus of this discussion.

The mechanism for translating market activity into an election forecast is fairly straightforward. Investors buy candidate shares (units) and the payoff for those shares is equal to the percentage of the final vote won by the candidate. If an investor thinks a candidate is likely to win with 55 percent of the vote, she has an incentive to buy that candidate's units when they are available at less than 55 cents per unit and to sell when they cost more than 55 cents per unit. The market aggregates traders' expectations to produce a market price that ends up reflecting the expected (among market participants) percentage of the vote the candidate will receive in the election. Wolfers and Zitzewitz describe the predictive power of electronic markets as deriving "from the fact that they provide incentives for *truthful revelation*, they provide incentives for research and *information discovery*, and the market provides an algorithm for *aggregating opinions*" (2004, 121).

One of the central claims of market proponents is that the market estimates are actually better predictors of vote shares than the results of trial-heat polls (Berg, Nelson, and Rietz 2008). And indeed, when comparing the accuracy of the market estimates of candidate vote shares with trial-heat polls, the markets do dominate. However, Erikson and Wlezien (2008) point out that this comparison is based on a naive treatment of trial-heat polls. What Erikson and Wlezien suggest is that rather than using the level of candidate support from trial-heat polls on day *t* as our estimate of the eventual election outcome, we should be *projecting* the expected election-day level of candidate support based on the historical relationship between polls on day *t* and the election-day results. Erikson and Wlezien are able to generate these estimates based on data from their "poll of polls" (Wlezien and Erikson 2002), in which they calculate a weighted daily average of all trial-heat polls from 1952 to 2004, from as far out as 200 days prior to the election. These daily estimates are then used to generate election-day predictions from daily observations of poll results, based on the historical trends in the data. Erikson and Wlezien's analysis of poll projections shows the polls dominating the 1988 and 1992 contests, edging out market predictions in 2004, and losing to market projections in 1996 and 2000. Erikson and Wlezien go on to show that when the winner-take-all market projections are compared to poll-based projections, the polls dominate in every election cycle.

As I see it, there are several points to take away based on the work of Erikson and Wlezien (2008) and Berg et al. (2008). First, election predictions drawn from the vote-share market are generally very close to the eventual outcome, especially as the election draws near. Second, it's not clearly the case that these predictions "dominate" the polls, especially if poll-based projections are used rather than simple poll-based vote shares.[12] Second, one important point that should not be lost is that trial-heat polls are expressions of vote intention, not of the respondents' perceptions of the expected vote share for each of the candidates. So, at some very real level, a comparison of raw trial-heat numbers to investors' expectations of the eventual outcome is not a fair fight. Finally, although the poll projections generally offer a higher level of accuracy, one of the chief advantages of the vote-share market is that market prices can be translated directly into expected vote shares without much effort. By comparison, the poll projections require Erikson and Wlezien's "poll of polls"-based daily regression slopes to translate the trial-heat results into expected vote totals. This advantage in ease-of-use makes the market a more accessible and, probably, preferred source for interested consumers of "horse-race" information.

Still, I wonder if "forecast" is the appropriate term for translating market shares into expected election outcomes. After all, the traditional understanding of forecasting, and the approach used by the models discussed earlier (Table 19.1), is that the historical relationship between predictor variables (measured prior to the election) and past election outcomes is used to forecast future outcomes. While one of the attractive features of the vote-share market is that the prices translate directly into expected outcomes, a behavioral model that estimates the relationship between market prices and actual outcomes, similar to traditional forecasting models, might eventually provide even better predictions. For instance, à la Erikson and Wlezien, instead of simply using something like the average market share price during, say, the last two weeks of the campaign as our estimate of the eventual outcome, it might be more interesting to ask what the expected outcome is based on the historical relationship between market prices during the last two weeks and the outcome on election day.

Unfortunately, the IEM has only been in existence since 1988 and there are only five data observations for any given pre-election time period from which to develop

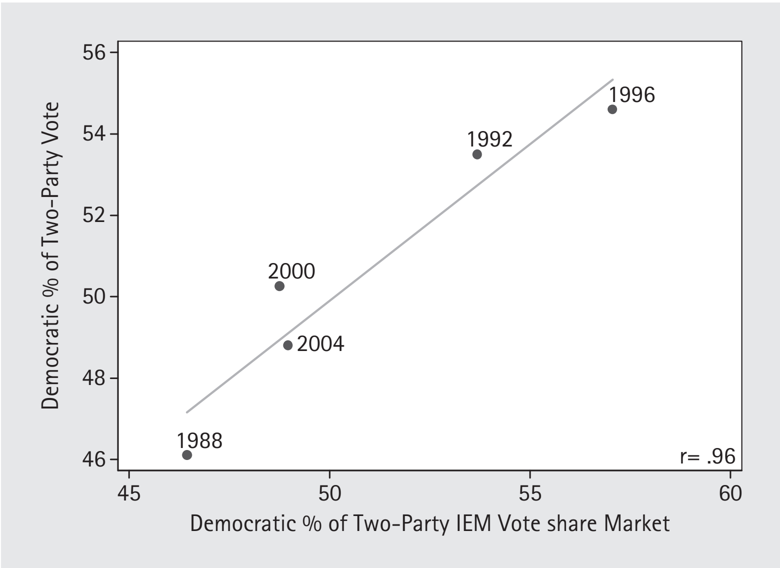


Figure 19.3 The linear relationship between the IEM vote share market (last two weeks of the campaign) and election outcomes, 1988–2004

such a behavioral model, so it is a bit risky to bet too much on estimates generated from such a model. Figure 19.3, which plots the relationship between the IEM vote-share prices during the last two weeks of the campaign and actual election results, demonstrates the potential for a behavioral model that translates vote-share prices into vote projections in a manner similar to Erikson and Wlezien's projections using poll-based vote shares. There is a very strong relationship between the vote-share market and election outcomes, one that follows a predictable, linear pattern. Again, there are probably too few observations to get a reliable sense of the predictive power of the relationship, but when a regression slope is estimated using observations from 1988 to 2000 and used to generate an out-of-sample prediction of the 2004 election,[13] the prediction error is .4 percentage points, whereas the error based on a direct translation of the market share price from the last two weeks of the 2004 campaign is 1.17 percentage points.

Citizen Forecasts

One of the explanations for the relative accuracy of the IEM vote-share market is that the market provides incentives for investors to gather information and weigh it objectively. One question this raises is whether investors are really better at predicting election outcomes than non-investors. Do investors out-perform casual observers? Put another way, and borrowing from Lewis-Beck and Skalaban (1989) and Lewis-Beck and Tien (1999), do investors out-perform "citizen forecasters"?

Lewis-Beck and Skalaban (1989) and Lewis-Beck and Tien (1999) demonstrate that, in aggregate, the American public is pretty good at picking the winners of US presidential elections. Lewis-Beck and colleagues are not able to use citizen-based estimates of the actual vote *share* for each candidate, but they are able to use a survey question from the pre-election wave (September through election day) of the National Election Study (NES) that asks respondents, "Who do you think will be elected President in November?" Aggregated responses to this question show an uncanny ability on the part of the general public to sniff out the likely winner, as shown in Figure 19.4. Here we see a strong, positive, linear relationship ($r = .87$) between citizen forecasts in the last two weeks of the campaign and actual election outcomes; and if the criterion is simply whether the citizen forecasts called the correct winner, they were only off the mark in 1976, 1980, and 2000. Since two of these elections (1976 and 2000) saw razor-thin margins of victory, this is a strong record during this period.

[13] $\text{Vote} = 12.0 + .76X$; $N=4$; $r^2=.91$

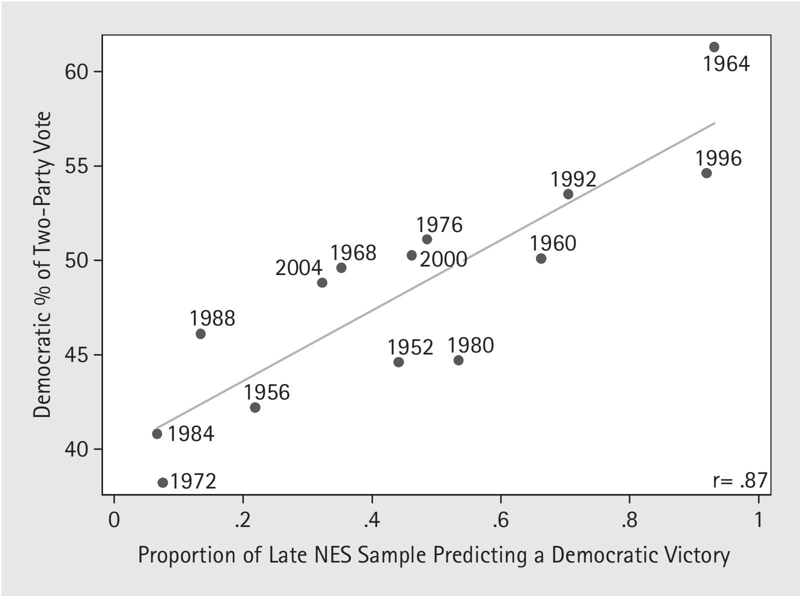


Figure 19.4 Citizen forecasts (last two weeks of campaign) and election outcomes, 1952–2004

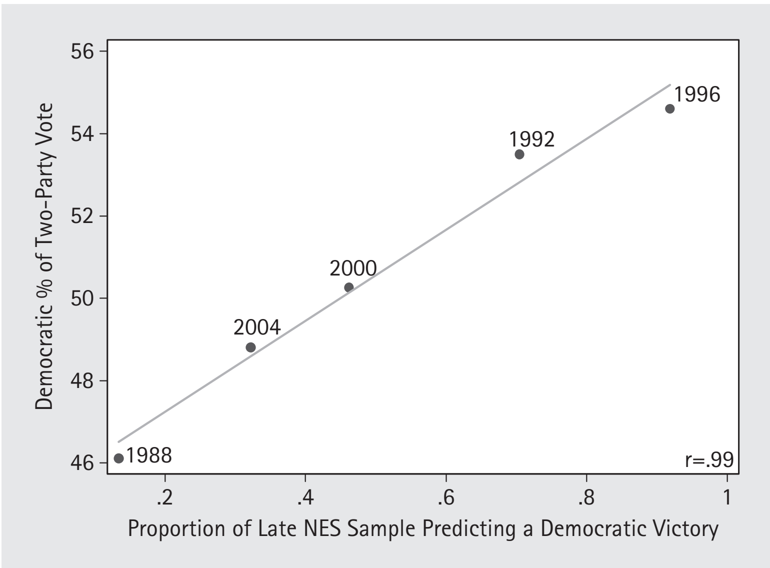


Figure 19.5 Citizen forecasts (last two weeks of campaign) and election outcomes, 1988–2004

But how do citizen forecasts in the last two weeks of the campaign stack up to those made by investors? A direct comparison based on Figures 19.3 and 19.4 is unfair since Figure 19.4 encompasses a longer time period and we don't know what the investor predictions would have been for the years 1952–1984. Instead, the best way to compare the two groups is to look at the relationship between citizen forecasts and votes for the same years (1988–2004) covered in Figure 19.3; these data are provided in Figure 19.5. From these data we can say that during the time period when both IEM- and NES-based citizen forecasts have been available, both sources of prediction track very closely with the actual vote received, and the relationship between citizen forecasts and actual votes is actually a bit stronger. So, again, the question is, do we need complex voting markets to generate accurate predictions of election outcomes? The evidence presented here suggests that surveys of the general public's sense of the likely election outcome may be just as useful.

State-Level Models

All of the preceding discussion has focused on forecasting mechanisms for predicting the *national* vote. But while the popular vote winner generally wins the election, American presidential elections are determined by winning a majority of the electoral votes, which are parceled out among the states. Given this feature of the American electoral system, it is somewhat surprising that more attention has not been given to forecasting state-level presidential election outcomes. This is not to say that this area has been completely neglected. Indeed, a number of scholars (Campbell 1992; Campbell, Ali, and Jalazai 2006; Desart and Holbrook 2003; Holbrook and Desart 1999; Rosenstone 1983; Soumbatiants, Chappell, and Johnson 2006) have assembled state-level forecasting models that have yielded impressive results.

In this section, I focus on those state-level models that utilize trial-heat polls to forecast election outcomes. This approach, used by Desart (2005), Desart and Holbrook (2003), Holbrook and Desart (1999), and Soumbatiants, Chappell, and Johnson (2006), has yielded impressive results and should assume a more prominent position in forecasts, as state-level polls grow in abundance.[14] The Holbrook and DeSart and the Soumbatiants et al. models are similar in spirit (trial-heat polls predict election outcomes) but different in implementation. Soumbatiants et al.

[14] This points to the one limitation in using state-level trial-heat polls: the polls themselves have only become plentiful in the last few election cycles, and the earliest usable data for all states is the 1988 election cycle.

(2006) use slopes based on state-level trial-heat polls from one day and fifteen days prior to elections, from 1988 to 2000, to generate predicted outcomes for the 2004 presidential election. In addition to providing point estimates, they also use simulation techniques to generate estimates of the probability of alternative state-level outcomes, as well as the probability of Electoral College outcomes. Using this methodology, Soumbatiants et al. are able to predict 86 percent of the 2004 state outcomes correctly from fifteen days out and 88 percent correctly from one day out.

The Holbrook and DeSart model predicts the Democratic share of the two-party popular vote in the states as a function of the Democratic vote in the previous two elections, and the statewide Democratic share of the two-party vote in trial-heat polls averaged across of all publicly available polls taken in September. Using data from 1992–2000 to forecast the 2004 state outcomes generates correct predictions in forty-seven (94 percent) of the fifty states; and, on average, the out-of-sample projections from 1992–2004 estimated 93 percent of the state outcomes correctly (Desart 2005). Although the lagged vote variable makes an important contribution to the model, the real power comes from the September poll average. In the 130 instances from 1992–2004 when one of the candidates had a lead in the September state-poll average outside the margin of error, the leading candidate went on to win 126 times (96.9 percent). If the margin of error is ignored and projections are based simply on which candidate is ahead in the September poll average, the September polls themselves predict 87.5 percent of all outcomes correctly.[15]

Even just these few examples of state-level forecasting models demonstrates that they offer an interesting and, I think, underutilized alternative to the standard national-aggregate model. Whether based mostly on aggregated state characteristics, similar to Campbell and colleagues, or taking advantage of state-level polling, similar to Soumbatiants et al. and Holbrook and DeSart, state-level models can generate highly accurate forecasts that, through aggregation, can also be used to generate predictions of the national outcome.

Conclusion

A number of recommendations flow directly from the analysis of forecasting models presented in this chapter. First, it is not terribly useful to declare one model or approach the "winner" based on accuracy in a single election period.

[15] Calculated with data presented in Holbrook and Desart (1999), Desart and Holbrook (2003), and Desart (2005).

Instead, it makes a good deal more sense to evaluate model characteristics that generally lead to greater accuracy across election cycles. Second, while we should not pin too much on single elections, we should not be closed off to learning from them. One lesson from the 2000 election that shows some promise is the idea of incorporating the conditioning effect of incumbency into forecasting models. The analysis presented here suggests just such a conditioning influence on the effects of presidential approval and perceived economic conditions. Finally, the dominant approach to forecasting presidential elections is to use national, longitudinal models, such as those presented in Table 19.1. While these models have been relatively successful over the past three election cycles, a number of interesting alternatives also exist, including electronic markets, "citizen" forecasts, and state-level forecasting models. Among these, the state-level models represent the closest kin to the dominant approach and (although they have already shown great success) also represent an area where much more work can be done—work that will hopefully help us better understand and forecast US presidential elections.

But I think it would also be fruitful to consider other innovations, not directly touched on in this essay. One possibility is to consider differential weighting of forecasting data to assign more emphasis to recent elections than to those in the relatively distant past. This would seem to be most important for models such as those used by Fair, Norpoth, and Cuzán and Bundrick, whose models include data from elections from the early twentieth century. Intuition suggests that the relationships in the model have evolved over time and that more weight assigned to recent elections could improve accuracy.

Another strategy that I think could bear fruit is to focus on more of a meta-analytic approach to forecasting. With each passing election cycle there is an increasing number of forecasts and approaches to forecasting, almost all of which are on sound theoretical ground, and most of which have an established track record. One interesting approach is to not pin everything on a single model or single approach, but instead to take advantage of the collective wisdom of the forecasting community and consider something like meta-forecasting. This "combining" approach is advocated by Armstrong (2001) and is applied to election forecasting at PollyVote (http://www.forecastingprinciples.com/PollyVote/). Although combining does not have much of a track record in election forecasts, it did provide an accurate prediction of the 2004 presidential election (Cuzán, Armstrong, and Jones 2005) and has improved the accuracy of prediction in other domains (Armstrong 2001). This approach warrants monitoring in the next couple of election cycles in order to get a sense of its advantages and disadvantages relative to traditional models.

Update: The 2008 Forecasts

The 2008 election provides an opportunity to update and evaluate many of the points made in this chapter. As in the past, most models are set up to predict the expected level of support (usually share of the two-party vote) for the incumbent party candidate, so the reference point here is McCain's 46.6 percent of the two-party popular vote and his 173 Electoral College votes. Nine forecasters published eleven pre-election predictions[16] in the October issue of *PS: Political Science and Politics* (Campbell 2008a), ranging from predictions of a narrow McCain victory to an Obama blowout. The average absolute error of the predictions (with Ray Fair's model added to the mix as the twelfth forecast) was 2.86 percentage points, making 2008 a better year than either 2000 (5.26 points) or 2004 (2.96 points). Similar to the data presented earlier, there was a tendency to over-predict the incumbent party's vote share, though the pattern was not as strong as in the past (seven of twelve models over-predicted McCain's vote share). The accuracy of the incumbency interaction model developed in this chapter was slightly better than the average model, missing the actual vote by just 2.3 points. Also similar to previous years, nearly every model (ten of twelve) predicted the correct winner of the popular vote. Alternatives to the national econometric models also fared well: the average IEM vote-share market during the last two weeks of the campaign predicted McCain would get 45.6 percent of the two-party vote, just one point less than he received, and the market's prediction on the day before the election was 46.9 percent for McCain. State-level models were also fairly accurate: DeSart and Holbrook's September poll model predicted 48.2 percent of the national (two-party) popular vote and 202 electoral votes for McCain, whereas their October model did somewhat better, predicting 46.1 percent of the popular vote and 189 electoral votes for McCain; and Klarner's (2008) state-level model nearly called the national popular vote on the nose, predicting 47 percent of the two-party vote for McCain, and also predicted 192 electoral votes for McCain.

In sum, from a forecasting perspective, the 2008 election outcome was business as usual. Some models were more accurate than others, as is always the case, but the average error was somewhat lower than in the past two elections cycles.

References

Abramowitz, A. I. 1988. An Improved Model for Predicting Presidential Election Outcomes. *PS: Political Science and Politics*, 21:843–7.

[16] Some forecasters provide two separate predictions. In these cases I include both predictions.

—— 2002. Who Will Win in November: Using the Generic Ballot Question to Forecast the Outcome of the 2002 Midterm Election. Symposium on midterm elections, <www.apsanet.org/~elections/archives.html>.

—— 2004. When Good Forecasts Go Bad: The Time-for-Change Model and the 2004 Presidential Election. *PS: Political Science & Politics*, 37: 745–6.

Abramowitz, A. I. 2006. National Conditions, Strategic Politicians, and U.S. Congressional Elections: Using the Generic Vote to Forecast the 2006 House and Senate Elections. *PS: Political Science & Politics*, 39:863–9.

Armstrong, J. S. 2001. Combining Forecasts. In *Principles of Forecasting: A Handbook for Researchers and Practitioners*, ed. J. S. Armstrong. Norwell, Mass.: Kluwer Academic Publishers.

Bardwell, K., and Lewis-Beck, M. S. 2004. State-level Forecasts of U.S. Senate Elections. *PS: Political Science & Politics*, 37: 821–6.

Beck, N. 1992. Forecasting the 1992 Presidential Election: The Message is in the confidence Interval. *The Public Perspective*, 3(6, September/October).

Berg, J. E., Nelson, F. D., and Rietz, T. A. 2008. Prediction Market Accuracy in the Long Run. *International Journal of Forecasting*, 24: 285–300.

Berke, R. L. 2001. Gore and Bush Strategists Analyze Their Campaigns. *New York Times*, 12 February 2001. Accessed online at <http://query.nytimes.com/gst/fullpage.html?res=9B03E0DA1131F931A25751C0A9679C8B63&sec=&spon=>.

Campbell, J. E. 1992. Forecasting the Presidential Vote in the States. *American Journal of Political Science*, 36: 386–407.

—— 1993. Weather Forecasters Should Be So Accurate: A Response to "Forewarned Before Forecast." *PS: Political Science & Politics*, 28: 165–6.

—— 2001. The Referendum that Didn't Happen: The Forecasts of the 2000 Presidential Election. *PS: Political Science & Politics*, 34: 33–8.

—— 2004. Forecasting the Presidential Vote in 2004: Placing Preference Polls in Context. *PS: Political Science & Politics*, 37:763–7.

—— 2008a. Editor's Introduction: Forecasting the 2008 National Elections *PS: Political Science & Politics*, 41/4 (October 2008): 679–81.

—— 2008b. Evaluating U.S. Presidential Election Forecasts and Forecasting Equations. *International Journal of Forecasting*, 24: 259–71.

—— Ali, S., and Jalalzai, F. 2006. Forecasting the Presidential Vote in the States, 1945–2004: An Update, Revision, and Extension of a State-level Presidential Forecasting Model. *Journal of Political Marketing*, 5: 33–57.

—— and Lewis-Beck, M. S. 2008. U.S. Presidential Election Forecasting: An Introduction. *International Journal of Forecasting*, 24: 189–92.

Colomer, J. M. 2007. What Other Sciences Look Like. *European Political Science*, 6: 134–42.

Cuzán, A. G., Armstrong, J. S., and Jones, R. 2005. Combining Methods to Forecast the 2004 Presidential Election. Presented at the Southern Political Science Association, New Orleans, January 6, 2005.

—— and Bundrick, C. M. 2005. Deconstructing the 2004 Presidential Election Forecasts: The Fiscal Model and the Campbell Collection Compared. *PS: Political Science and Politics*, 38: 255–62.

Desart, J. A. 2005. Forecasting the 2004 Presidential Election with State Trial-Heat Polls. Paper presented at the annual meeting of the Midwest Political Science Association, Chicago, Ill., April 7–10.

Desart, J. A. and Holbrook, T. M. 2003. Statewide Trial-Heat Polls and the 2000 Presidential Election: A Forecast Model. *Social Science Quarterly*, 84: 561–73.

Erikson, R. S., and Bafumi, J. 2002. Generic Polls, Late Balancing Acts, and Midterm Outcomes: Lessons from History for 2002. Symposium on midterm elections, <www.apsanet.org/~elections/archives.html>.

—— and Wlezien, C. 2008 Are Political Markets Really Superior to Polls as Election Predictors? *Public Opinion Quarterly*, 72: 190–215.

Fair, R. 2004. Presidential Vote Equation – October 20, 2004. <http://fairmodel.econ.yale.edu/vote2004/vot1004.htm.>

Fiorina, M. P. 1981. *Retrospective Voting in American National Elections*. New Haven, Conn.: Yale University Press.

Greene, J. P. 1993. Forewarned Before Forecast: Presidential Election Forecasting Models and the 1992 Election. *PS: Political Science & Politics*, 26: 17–21.

Holbrook, T. M. 2001. Forecasting with Mixed Economic Signals: A Cautionary Tale. *PS: Political Science & Politics*, 34: 39–44.

—— 2004. Good News for Bush? Economic News, Personal Finances, and the 2004 Presidential Election. *PS: Political Science & Politics*, 37: 759–61.

—— and Desart, J. A. 1999. Using State Polls to Forecast Presidential Election Outcomes in the American States. *International Journal of Forecasting*, 15: 137–42.

Klarner, C. 2008. Forecasting the 2008 U.S. House, Senate and Presidential Elections at the District and State Level. *PS: Political Science & Politics*, 41: 723–8.

—— and Buchanan, S. 2006a. Forecasting the 2006 elections for the U. S. House of Representatives. *PS: Political Science & Politics*, 39: 857–61.

—— —— 2006b. Forecasting the 2006 Elections for the United States Senate. *PS: Political Science & Politics*, 39: 849–855.

Lewis-ᵇeck, M. S. 1985. Election Forecasts in 1984: How accurate were they? *PS: Political Science & Politics*, 18: 53–62.

—— 2005. Election Forecasting: Principles and Practice. *British Journal of Politics & International Relations*, 7: 145–64.

—— and Rice, T. W. 1984. Forecasting U.S. House Elections. *Legislative Studies Quarterly*, 9: 475–86.

—— —— 1992. *Forecasting Elections*. WashinGton, D.C.: CQ Press.

—— and Skalaban, A. 1989. Citizen Forecasting: Can Voters See into the Future? *British Journal of Political Science*, 19: 146–53.

—— and Tien, C. 1999. Voters as Forecasters: A Micromodel of Election Prediction. *International Journal of Forecasting*, 15: 175–84.

—— —— 2001. Modeling the Future: Lessons from the Gore Forecast. *PS: Political Science & Politics*, 34: 21–3.

Lewis-Beck, M. S., and Tien, S. 2004. Jobs and the Job of President: A Forecast for 2004. *PS: Political Science & Politics*, 37: 753–8.

Lockerbie, B. 2004. A Look to the Future: Forecasting the 2004 Presidential Election. *PS: Political Science & Politics*, 37: 741–3.

Nadeau, R., and Lewis-Beck, M. S. 2001. National Economic Voting in U.S. Presidential Elections. *Journal of Politics*, 63: 159–81.

Norpoth, H. 2004. From Primary to General Election: A Forecast of the Presidential Vote. *PS: Political Science & Politics*, 37: 737–40.

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

Soumbatiants, S., Chappell, H. W., Jr., and Johnson, E. 2006. Using State Polls to Forecast U.S. Presidential Election Outcomes. *Public Choice*, 127: 207–23.

Strumpf, K. F., and Rhode, P. 2004 Historical Presidential Betting Markets. *Journal of Economic Perspectives*, 18: 127–42.

Tetlock, P. E. 2005. *Expert Political Judgment: How Good Is It? How Can We Know It?* Princeton, N.J.: Princeton University Press.

Wlezien, C. 2001. On Forecasting the Presidential Vote. *PS: Political Science and Politics*, 34: 24–31.

—— and Erikson, R. S. 2002. The Timeline of Presidential Election Campaigns. *Journal of Politics*, 64: 969–93.

—— —— 2004. The Fundamentals, the Polls, and the Presidential Vote. *PS: Political Science & Politics*, 37: 747–51.

Wolfers, J., and Zitzewitz, E. 2004. Prediction Markets. *Journal of Economic Perspectives*, 18: 107–26.