Article: "Presidential Vote Models: A Recount"
Author: Larry M. Bartels, John Zaller
Issue: Mar. 2001
Journal: ***PS: Political Science & Politics***

Presidential Vote Models: A Recount*

by
Larry M. Bartels,
Princeton University
John Zaller,
University of California, Los Angeles

Was it Al Gore's election to lose? Most political scientists, including us, believed that peace and a booming economy would give Gore a significant advantage in the 2000 presidential race. The election outcome—a virtual dead heat in the popular vote—has prompted two reactions that seem to us to be quite wrong-headed.

On one hand, many journalists and some political scientists have interpreted the 2000 result as casting doubt on the basic premise of presidential vote models, that economic and political fundamentals play a systematic and largely predictable role in shaping presidential election outcomes. If Gore was supposed to win easily but didn't, doesn't that just go to show that the so-called "fundamentals" are less important than they seem and that every election is a unique political event, beyond the reach of simple-minded historical regression analyses?

On the other hand, some observers, including some of the most prominent election forecasters, have concluded that the problem was not with the forecasting models but with Al Gore as a candidate. "Gore didn't run a campaign consistent with the model," according to Helmut Norpoth (quoted by Miller 2000). He didn't stress the economy. He didn't embrace the popular incumbent president. He was too uncharismatic; too ideological; not ideological enough. In short, as one post-mortem had it, Gore was "a flawed candidate who squandered a prime opportunity to capture the White House" (Berke 2000).

Here, we entertain the radical notion that Al Gore was a solid candidate—about as solid, at any rate, as George W. Bush—and that the 2000 election turned out (give or take a few hundred votes in Florida) almost exactly as should have been expected. Using data from the 13 previous post-war presidential elections (1948–96), we examined a total of 48 different regression models employing a variety of economic and political variables. The balance of evidence from these analyses suggests that Gore's advantage with respect to the fundamentals was modest at best and that the election outcome was well within the range one would expect if both candidates ran more or less equally competent campaigns. As we show below, there have been some elections that make presidential vote models look bad, but 2000 is not one of them.

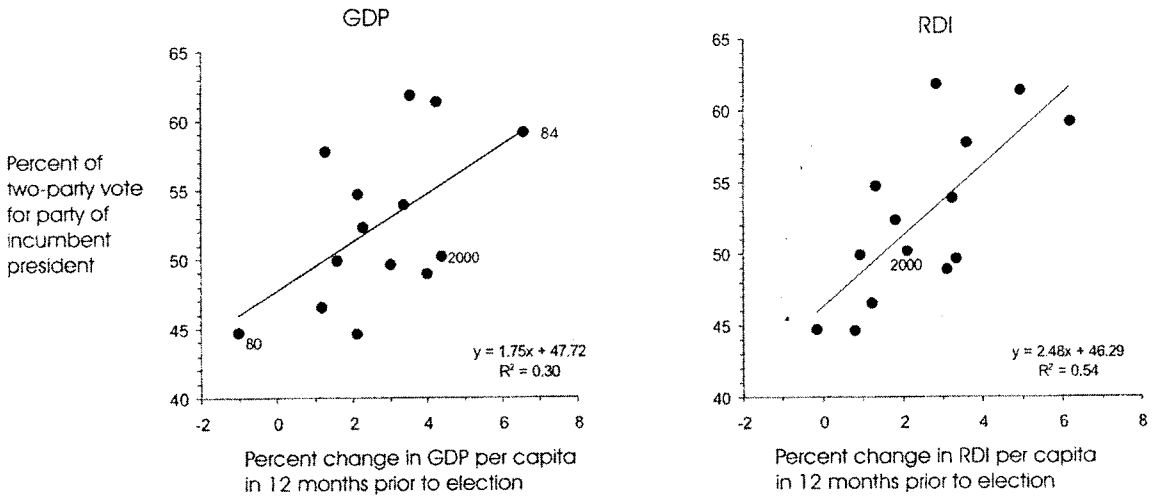
Figure 1 provides both a summary of our substantive argument and an introduction to some of the analytical issues we address below. The left-hand panel of the figure shows the relationship between one familiar measure of election-year economic conditions, the annual growth rate of per capita GDP in the four quarters prior to the election, and the incumbent party's share of the two-party vote in each presidential election from 1948 to 2000.[1] Because some of the economic data used to create this figure were not released until late November, and because the 2000 election is one of the points contributing to the regression line, the relationship cannot be construed as providing a forecast of the 2000 election outcome. Nevertheless, the implications of the relationship will be quite familiar to consumers of election forecasts publicized before and after the 2000 vote. Because GDP grew at a robust rate in the year before the 2000 election (among post-war election economies, only 1984's was stronger, as the figure shows), Gore might have been expected to win fairly easily (by more than six percentage points). Gore's actual popular vote margin was less than one-half of one percentage point, representing a significant underperformance for the incumbent party.

The right-hand panel of Figure 1 tells a very different story. Here, the incumbent party's vote share in each post-war election is related to the annual growth rate of real disposable income (RDI) per capita in the four quarters prior to the election. The RDI figure for 2000 is a

little worse than the historical average (that is, slightly to the left of the mean of points on the horizontal axis), indicating that income growth was not particularly strong in 2000. Given that fact, Gore did about as well as should have been expected; the actual 2000 election outcome is only about one percentage point below the regression line.

Which of these two stories is right? In the following pages, we bring some technical machinery to bear in constructing an answer to that question. But the basic logic underlying the technical argument can be gleaned from a comparison of the two graphs in Figure 1.

Figure 1
Effect of GDP and RDI on vote for president, 1948–2000



Clearly, income growth is a much better predictor of presidential election outcomes than GDP growth, as evidenced by the steeper slope, tighter clustering of points, and higher R^2 value in the right-hand panel of Figure 1. The story told by the right-hand panel—that Gore did about as well as should have been expected given the mediocre condition of the economy—is therefore the more credible of the two.

Yet, these two graphs embody the results of only two of many possible election models, and two unrealistically simple models at that. There are other plausible ways to measure economic conditions besides the two on display in Figure 1 and other variables that deserve to be taken into account besides the state of the economy. We explored some of these complexities by examining the deterministic effects of six different measures of economic performance and three noneconomic factors: peace, ideological moderation, and incumbent-party fatigue.

As will quickly become clear, expanding the range of potential explanatory factors generated a daunting variety of plausible model specifications. With so many variables to analyze, it is easy to find models that "predict" that Gore should have done about as well as he did, but harder to demonstrate, as we attempt to do here, that those models should be taken more seriously than the many models that showed Gore winning comfortably. Having settled upon what seems to us to be a reasonable way of synthesizing the available evidence, we examined the substantive impact of economic and political factors in post-war presidential elections generally and in the 2000 election more specifically. We detail our findings below and conclude with some observations regarding the "lessons" of the 2000 election, such as they are.

Because the models we examined all use data released by the U.S. Bureau of Economic Analysis (BEA) since the election, none can strictly be considered a forecast model. Nor, because we examined only a subset of variables used by other scholars, does our analysis fully engage the voluminous literature on forecasting presidential elections.[2] We do nonetheless cast some light on two issues of general interest: the extent to which the outcome of the 2000 election can be explained by the play of a plausible set of "political fundamentals," and how best to make inferences from small data sets like those analyzed with presidential vote models.

Forty-eight Models are Better than One

Given the vast array of more- or less-plausible alternative models described in the literature on aggregate election outcomes, how should one settle on a single "best" model as a basis for scientific inference and forecasting? In our view, one shouldn't. Rather, we believe one should (as we did) consider a wide variety of plausible models and attempt to synthesize the implications of all of them.

We examined the full set of 48 regression models generated by employing each of six different indicators of economic performance alone and in various combinations with three additional explanatory variables: a dummy variable for war years (1952 and 1968), a measure of the relative ideological moderation of the two presidential nominees, and a measure of the number of consecutive terms the incumbent party has held the White House. Regression results for these 48 models are presented in Table 1. The data on which the results are based are publicly available.[3]

But having compiled an array of models and results, we immediately faced the problem of how to make sense of so much data. We solved that problem by applying a relatively new and somewhat controversial statistical technique known as Bayesian model averaging.

Bartels (1997a) has summarized the theoretical rationale and technical aspects of model averaging elsewhere, so we will not rehearse them here. To understand our argument, it suffices for the nontechnical reader to understand two general principles. First, when plausible alternative models produce different results, it is important to recognize those differences—and the differences in the models that produced them—as a significant source of uncertainty in our statistical inferences, including out-of-sample forecasts. Rather than trusting (and touting) the results of any one model as if they were the final word, analysts should base their conclusions (whether formally or informally) on the range of evidence provided by plausible alternative models.

The second general principle of Bayesian model averaging is that the results of alternative models should figure more or less heavily in this synthesis depending, at least in part, on how well they fit the data. If, by some appropriate criterion, one model works better than another, then the results it generates should be given correspondingly more (though never total) credence. All reasonable models, even those that perform relatively poorly, deserve at least some weight.

It is not unusual for analysts to explore dozens of alternative regression models, as we did when performing the analyses reported in Table 1. The Bayesian model-averaging approach provides some rationale for this common practice. What is distinctive is that the model-averaging approach allows analysts to employ all of the resulting regression results, and to provide an honest assessment of uncertainty in light of this experimentation. Thus, "data-dredging" is disciplined by systematic attention to the evidence provided by a portfolio of plausible models, including the implications of inconsistencies among them.

Bayesian model averaging is actually weighted averaging, where the appropriate weight for each model (referred to as the "Bayes factor") is calculated from three simple numbers: the number of parameters in the model, the number of observations used to estimate those parameters, and the R^2 statistic for the regression analysis.[4] As one might expect, the weighted results give the most credence to models that fit the data well while employing relatively few parameters. Parsimony is as important as goodness of fit, because a parsimonious model is less likely than a complex one to have overfit the data.[5]

In Table 1, we report "posterior model probabilities" reflecting the relative plausibility of each of our 48 models given its performance from 1948 through 1996.[6] For example, the single best-performing model among the 48 we considered is a version of Douglas Hibb's (2000) "Bread and Peace" model (model C3 in Table 1), which includes a weighted average of RDI growth over each presidential term plus a dummy variable for wars beginning on the incumbent party's watch.[7] The posterior model probability for this model is .1932, which means that our version of Hibbs' model receives about 19% of the weight in our overall results.

While the "Bread and Peace" model is by far the "best" of the 48 models considered in Table 1, it is by no means the only plausible model in light of our data. Several other models (B4, B7, C5, A6, A7, C7, and A4) also received significant weight in our overall results, with posterior model probabilities ranging from 7 to 9% of the total. On the other hand, almost half of the 48 models in Table 1 seem quite implausible in light of our data, and their results receive very little weight—less than one-fifth of one percent each—in our subsequent analysis.

The range of posterior model probabilities reported in Table 1 illustrates the extent to which the model-averaging procedure is sensitive to the implications of the data: the best-fitting model receives more than a

Figure 2
Distribution of 2000 forecast errors from alternative models

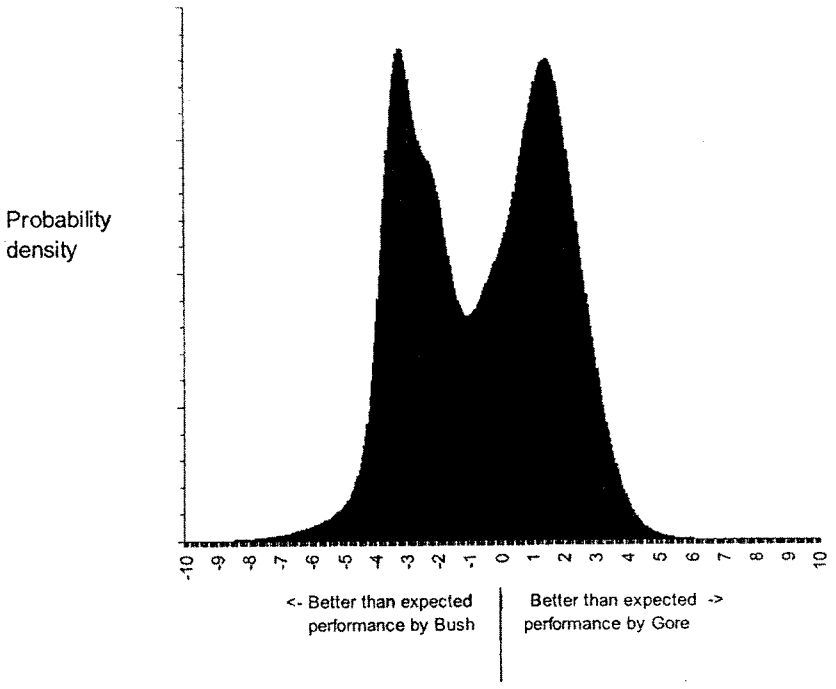


TABLE 1
Regression Results, 1948–1996

	Election Year RDI (A)	Quarters 12–15 RDI (B)	Weighted Average RDI (C)	Election Year GDP (D)	Quarters 12–15 GDP (E)	Weighted Average GDP (F)
Economic Growth	2.25	2.47	3.00	1.53	1.97	1.84
	(.67)	(.68)	(1.00)	(.76)	(.78)	(1.02)
Temporal Discount	—	—	.78	—	—	.71
			(.11)			(.22)
Intercept	46.27	46.42	44.91	48.68	47.53	48.83
	(2.27)	(2.10)	(2.57)	(2.49)	(2.48)	(2.63)
(1)	R^2 = .507	R^2 = .541	R^2 = .561	R^2 = .269	R^2 = .365	R^2 = .273
	ser = 4.39	ser = 4.23	ser = 4.34	ser = 5.34	ser = 4.98	ser = 5.59
	pr = .0024	pr = .0038	pr = .0014	pr = .0002	pr = .0005	pr = .0001
	err = −.54	err = −1.47	err = −.96	err = −4.63	err = −5.94	err = −4.54
Economic Growth	1.93	1.99	2.28	1.39	1.70	1.59
	(.56)	(.70)	(.71)	(.61)	(.65)	(.82)
Temporal Discount	—	—	.68	—	—	.65
			(.15)			(.29)
Incumbent Terms	−2.09	−1.58	−2.42	−2.54	−2.33	−2.58
	(.82)	(.97)	(.83)	(.97)	(.93)	(1.09)
Intercept	51.37	50.80	51.13	54.12	52.89	54.50
	(2.73)	(3.32)	(2.81)	(2.88)	(2.95)	(3.37)
(2)	R^2 = .701	R^2 = .638	R^2 = .787	R^2 = .569	R^2 = .611	R^2 = .570
	ser = 3.59	ser = 3.94	ser = 3.19	ser = 4.31	ser = 4.09	ser = 4.53
	pr = .0169	pr = .0016	pr = .0422	pr = .0016	pr = .0031	pr = .0004
	err = −.82	err = −1.68	err = −.38	err = −4.45	err = −5.49	err = −4.03
Economic Growth	2.15	2.33	3.71	1.61	1.94	2.49
	(.61)	(.63)	(.69)	(.68)	(.71)	(.87)
Temporal Discount	—	—	.86	—	—	.76
			(.05)			(.10)
War Dummy	−5.67	−5.29	−9.11	−7.18	−6.48	−9.62
	(3.07)	(3.00)	(2.49)	(3.67)	(3.46)	(3.74)
Intercept	47.43	47.58	45.17	49.56	48.59	49.00
	(2.15)	(2.04)	(1.68)	(2.27)	(2.31)	(2.05)
(3)	R^2 = .632	R^2 = .650	R^2 = .831	R^2 = .472	R^2 = .530	R^2 = .593
	ser = 3.98	ser = 3.88	ser = 2.84	ser = 4.76	ser = 4.49	ser = 4.41
	pr = .0044	pr = .0061	pr = .1932	pr = .0004	pr = .0009	pr = .0006
	err = −1.50	err = −2.35	err = −3.33	err = −5.85	err = −6.90	err = −7.12
Economic Growth	2.56	2.73	2.84	1.91	2.42	1.92
	(.50)	(.52)	(.95)	(.67)	(.64)	(.95)
Temporal Discount	—	—	.77	—	—	.69
			(.11)			(.22)
Ideological Moderation	4.05	3.83	2.31	4.11	4.31	3.16
	(1.26)	(1.22)	(1.60)	(1.80)	(1.56)	(2.02)
Intercept	43.77	44.24	44.34	46.02	44.62	47.39
	(1.84)	(1.72)	(2.46)	(2.41)	(2.22)	(2.61)
(4)	R^2 = .757	R^2 = .769	R^2 = .644	R^2 = .521	R^2 = .641	R^2 = .430
	ser = 3.23	ser = 3.15	ser = 4.12	ser = 4.54	ser = 3.93	ser = 5.22
	pr = .0661	pr = .0903	pr = .0015	pr = .0008	pr = .0051	pr = .0001
	err = 2.72	err = 1.46	err = .83	err = −2.14	err = −3.56	err = −2.23
Economic Growth	1.95	2.08	3.09	1.46	1.75	2.02
	(.57)	(.70)	(1.29)	(.62)	(.66)	(.97)
Temporal Discount	—	—	.81	—	—	.71
			(.14)			(.20)
War Dummy	−2.80	−3.69	−6.27	−3.60	−3.24	−6.00
	(3.26)	(3.55)	(5.27)	(3.92)	(3.73)	(4.91)
Incumbent Terms	−1.67	−.98	−.97	−2.00	−1.84	−1.55
	(.97)	(1.12)	(1.58)	(1.14)	(1.10)	(1.42)
Intercept	50.91	49.94	47.76	53.39	52.30	52.47
	(2.82)	(3.40)	(4.59)	(3.01)	(3.07)	(3.89)
(5)	R^2 = .723	R^2 = .677	R^2 = .841	R^2 = .605	R^2 = .641	R^2 = .654
	ser = 3.63	ser = 3.93	ser = 2.92	ser = 4.34	ser = 4.14	ser = 4.31
	pr = .0078	pr = .0029	pr = .0796	pr = .0008	pr = .0014	pr = .0005
	err = −1.24	err = −2.21	err = −2.20	err = −5.10	err = −6.06	err = −5.64

TABLE 1— continued
Regression Results, 1948–1996

	Election Year RDI (A)	Quarters 12–15 RDI (B)	Weighted Average RDI (C)	Election Year GDP (D)	Quarters 12–15 GDP (E)	Weighted Average GDP (F)
Economic Growth	2.30	2.61	2.32	1.67	2.12	1.68
	(.51)	(.65)	(.82)	(.62)	(.63)	(.86)
Temporal Discount	—	—	.69	—	—	.66
			(.18)			(.27)
Ideological Moderation	3.02	3.55	.72	2.55	3.03	1.51
	(1.38)	(1.55)	(1.60)	(1.86)	(1.70)	(2.02)
Incumbent Terms	−1.20	−.32	−2.19	−1.85	−1.46	−2.17
	(.81)	(.98)	(1.09)	(1.05)	(.97)	(1.26)
Intercept	47.32	45.27	50.32	50.99	48.84	52.90
	(2.97)	(3.68)	(3.77)	(3.58)	(3.51)	(4.15)
(6)	R^2 = .805	R^2 = .771	R^2 = .793	R^2 = .643	R^2 = .712	R^2 = .598
	ser = 3.05	ser = 3.30	ser = 3.34	ser = 4.13	ser = 3.71	ser = 4.65
	pr = .0748	pr = .0270	pr = .0141	pr = .0015	pr = .0060	pr = .0002
	err = 1.73	err = 1.20	err = .05	ser = −2.95	err = −3.98	err = −3.03
Economic Growth	2.46	2.61	3.53	1.91	2.34	2.40
	(.48)	(.50)	(.79)	(.62)	(.60)	(.85)
Temporal Discount	—	—	.84	—	—	.74
			(.06)			(.13)
War Dummy	−3.63	−3.32	−8.25	−5.46	−4.47	−8.19
	(2.47)	(2.45)	(2.94)	(3.36)	(2.97)	(3.87)
Ideological Moderation	3.51	3.34	.84	3.38	3.66	2.03
	(1.25)	(1.23)	(1.36)	(1.72)	(1.53)	(1.83)
Intercept	44.84	45.24	45.01	47.17	45.79	48.12
	(1.89)	(1.81)	(1.76)	(2.35)	(2.23)	(2.16)
(7)	R^2 = .804	R^2 = .808	R^2 = .839	R^2 = .629	R^2 = .713	R^2 = .649
	ser = 3.06	ser = 3.03	ser = 2.94	ser = 4.21	ser = 3.70	ser = 4.34
	pr = .0742	pr = .0832	pr = .0733	pr = .0012	pr = .0061	pr = .0005
	err = 1.68	err = .54	err = −2.32	err = −3.51	err = −4.58	err = −5.09
Economic Growth	2.31	2.70	3.01	1.74	2.16	2.09
	(.51)	(.63)	(1.35)	(.63)	(.64)	(.99)
Temporal Discount	—	—	.80	—	—	.70
			(.15)			(.20)
War Dummy	−2.52	−3.65	−5.89	−3.51	−3.08	−5.93
	(2.77)	(2.89)	(5.51)	(3.76)	(3.37)	(5.01)
Ideological Moderation	2.96	3.54	.78	2.52	2.98	1.49
	(1.40)	(1.50)	(1.32)	(1.87)	(1.72)	(1.90)
Incumbent Terms	−.83	.27	−.87	−1.32	−1.00	−1.17
	(.91)	(1.06)	(1.65)	(1.20)	(1.10)	(1.53)
Intercept	46.98	44.45	47.33	50.31	48.32	50.95
	(3.02)	(3.62)	(4.82)	(3.68)	(3.59)	(4.40)
(8)	R^2 = .823	R^2 = .809	R^2 = .849	R^2 = .678	R^2 = .739	R^2 = .682
	ser = 3.08	ser = 3.20	ser = 3.05	ser = 4.16	ser = 3.74	ser = 4.42
	pr = .0392	pr = .0244	pr = .0304	pr = .0008	pr = .0032	pr = .0002
	err = 1.31	err = .67	err = −1.47	err = −3.61	err = −4.55	err = −4.60

Ordinary least squares parameter estimates with standard errors in parentheses. Dependent variable is incumbent party's share of two-party presidential vote (%).

N = 13. "ser" = standard error of regression. "pr" = model posterior probability. "err" = 2000 forecast error.

(A) Election Year RDI: Percentage change in real disposable income per capita, quarters 13–16.

(B) Quarters 12–15 RDI: Percentage change in real disposable income per capita, quarters 12–15.

(C) Weighted Average RDI: Average percentage change in real disposable income per capita, quarters 2–15 (with temporal discounting).

(D) Election Year GDP: Percentage change in gross domestic product per capita, quarters 13–16.

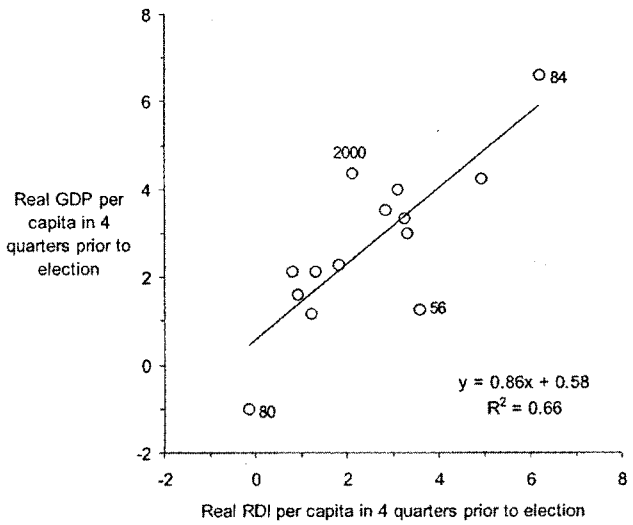
(E) Quarters 12–15 GDP: Percentage change in gross domestic product per capita, quarters 12–15.

(F) Weighted Average GDP: Average percentage change in gross domestic product per capita, quarters 2–15 (with temporal discounting).

thousand times as much weight as the worst-fitting models in the model-averaging calculations. At the same time, the fact that the best-fitting model receives less than one-fifth of the total weight emphasizes the point that no single specification is likely to capture adequately the inferential implications of the available data in situations like the one considered here, where theory and evidence are both relatively weak.

One potential advantage of basing our conclusions on a broad portfolio of plausible models rather than any single model is illustrated by the performance of these models in "forecasting" the outcome of the 2000 election.[8] Although it is easily the single best-performing model among the 48 presented in Table 1, the prediction yielded by our "Bread and Peace" model happens to miss Gore's popular vote share by 3.3 percentage points, a fairly substantial miss given that the standard error of the regression is only 2.8 percentage points. By comparison, a simple average of all 48 forecasts is off the mark by only 2.5 percentage points, and a performance-weighted average forecast (that is, with each of the 48 models weighted by its posterior model probability) is only off the mark by a little more than half a percentage point. Thus, at least in this case, relying on a portfolio of credible models produces a notably better forecast than relying on any one model, even one selected on the basis of superior past performance.

Figure 3
GDP vs. RDI in years prior to elections, 1948–2000



The probability distribution shown in Figure 2 nicely conveys the sense in which the portfolio of models presented in Table 1 produces a better out-of-sample forecast of the 2000 election outcome than even the best of those models considered in isolation. The figure displays the distribution of "forecast" errors for 2000 for each of our 48 models, each weighted by the posterior model probability.[9] The distribution displays a prominent mode around – 3, mostly reflecting the forecast error for the "Bread and Peace" model, which suggests that Gore should have won by three or four percentage points. But there is a second prominent mode around +2, which reflects that other plausible models (especially B4, A6, A7, and A4) suggeted Gore would lose, given the mediocre rate of income growth in 2000. Considering either of the two modes in Figure 2 by itself would make the 2000 election outcome seem at least somewhat surprising, but the distribution as a whole does not.

Of course, there is no guarantee that a portfolio of plausible models will outperform a single well-chosen model in every instance, but the argument for model averaging is that it often will. Why? Because when data are scarce, as they are here (only 14 presidential elections have been held in the half-century since reliable, detailed economic data became available), and plausible models are numerous, as they are here (we identified 48 different model specifications in a very selective canvassing), any single model—especially one selected through extensive empirical experimentation, with the attendant dangers of data dredging and overfitting—is likely to obscure much of the real uncertainty that ought to be reflected in statistical inferences and forecasts. Model averaging better represents that uncertainty, and thus reduces the risk that new observations will bring unhappy surprises.

Our analysis of the substantive implications of the results presented in Table 1 begins with a discussion of findings for the most important variable in nearly all presidential vote models, economic performance. Next we examine how war, ideological positioning of the candidates, and the incumbent party's tenure in office affected the outcome of the 2000 vote. Finally, we discuss how incorporating information from the 2000 election into our historical regression analyses should modify our assessment of the impact of our various explanatory factors and the relative plausibility of our various alternative models.

The Real Economy: Less than Great and Getting Worse

The most important basis for optimism about Gore's prospects among political scientists—and Democrats—was the perception that the nation was in the midst of an unprecedented period of economic prosperity. But as we have already suggested, the most relevant economic indicators were never as favorable for the Democrats as many observers (including us) believed in the months leading up to the 2000 election. The economic indicators most commonly employed in election forecasting models are based either on growth in economic output (GDP per capita) or on growth in income (RDI per capita). On theoretical grounds, income growth seems to us to be more relevant than output growth, since it reflects more directly the economic experiences of

prospective voters. However, rather than relying solely on our own theoretical intuitions, we included among our 48 regression models 24 in which economic growth is measured by GDP in one form or another and 24 in which economic growth is measured by RDI in one form or another.

The results presented in Table 1 strongly indicate RDI is a more electorally relevant economic indicator than GDP. Holding other aspects of the regression specification constant, models using RDI outperform models using GDP in every one of the 24 comparisons, often by a substantial margin. The sum of the posterior model probabilities for the RDI models is .964; less than 4% of the posterior probability is assigned to models employing GDP.

The distinction between income growth and output growth is of modest importance in most election years, since the two indicators tend to move in tandem. However, in 2000 the distinction was more consequential, as a comparison of the two panels in Figure 1 suggests, because income growth was significantly less robust than output growth. The rate of growth in GDP per capita in the four quarters before the 2000 election was 4.37 percent, which ranks second-best among the 14 election years in the post-war period, while the rate of growth in RDI per capita over the same one-year period was only 2.12 percent, which ranks eighth-best among the 14 post-war election years. Thus, 2000 appears (along with 1956) as a notable outlier among election years in terms of the historical relationship between income growth and output growth, which is displayed in Figure 3.

Why the unusually large discrepancy between (robust) output growth and (mediocre) income growth in 2000? That question is perhaps best addressed to economists. We note, however, that the federal budget surplus for fiscal year 2000 was $217 billion, more than $750 for each man, woman, and child in America. If half that wealth had been added to disposable income (say, in the form of a middle-class tax cut) it would have increased election-year income growth by about 1.6 percent, which would have made Clinton's second-term economic performance notably strong in terms of income growth as well as output growth. Clinton may have displayed more fiscal discipline than political sense in spurning Republican proposals for a tax cut. Not only might his administration have put additional income in voters' pockets by Election Day; it could also have denied Republicans what was perhaps their most potent issue in the 2000 campaign.

As weak as the 2000 economy was, at least as measured by income growth, it became noticeably weaker over the course of the election year.[10] Preliminary government data indicate that per capita RDI grew at an annual rate of only 1.4% in the third quarter of the election year (as compared with 2.0% in the first half of the year), and actually declined by 0.7% in October (U.S. Bureau of Economic Analysis 2000).[11]

To what extent did this economic slowdown contribute to Gore's defeat? It is somewhat hard to tell, since scholars disagree about the precise time horizon over which voters weigh economic conditions. In running the regression models presented in Table 1 we examined three distinct possible economic time horizons. In one set of estimates, we tested economic effects for the calendar year of the election, including the nearly two months that follow Election Day. (Since economic data for the fourth quarter of 2000 have not been released at the time of our writing, we assume real per capita annualized growth rates of 1% in this period.) In a second set of estimates, we tested economic effects for the four quarters immediately prior to the election, that is, data from quarters 12 through 15 of each administration (and omitting the 16th quarter, which begins October 1). Both of these models embody the assumption that voters give total weight to recent quarters and ignore the rest. For a final set of estimates, we used economic data from quarters 2 through 15 plus one-third of quarter 16 of each administration, and included an extra parameter to estimate the precise extent to which voters give more or less weight to recent quarters. This approach has been developed by Hibbs (1987, 2000).

The regression results tend to favor the models employing the Hibbs weighting scheme, giving them 44% of the total posterior probablility. This is an impressive performance in light of the fact that the Bayesian model-averaging procedure heavily penalizes Hibbs's scheme for using an extra parameter to estimate how voters discount past economic performance. Yet, the Hibbs models are scarcely dominant. The models employing economic performance in the calendar year of the election get 29% of the total posterior probability, while the results from models employing performance in quarters 12 through 15 of each administration get the remaining 27%.

Since all three economic time horizons seem fairly plausible in light of our data, all three of the corresponding RDI effects should contribute to our account of the 2000 election outcome. How much they should

TABLE 2
Model-Averaged Parameter Estimates, 1948–1996

Election Year RDI	.68 (1.11)
Quarters 12–15 RDI	.64 (1.18)
Weighted Average RDI	1.45 (1.74)
(Temporal Discount)	.82 (.12)
Election Year GDP	.01 (.15)
Quarters 12–15 GDP	.06 (.36)
Weighted Average GDP	.01 (.12)
(Temporal Discount)	.71 (.20)
War Dummy	−3.94 (4.45)
Ideological Moderation	1.86 (1.99)
Incumbent Terms	−.45 (1.02)

Model-averaged parameter estimates with standard errors in parentheses. Dependent variable is incumbent party's share of two-party presidential vote (%). Derived from parameter estimates in Table 1.

contribute is not at all obvious from the scattered individual coefficients in Table 1; but the model-averaging approach provides a coherent basis for synthesizing various parameter estimates in the form of summary coefficients for each explanatory variable. Each summary coefficient is simply a weighted average of the corresponding coefficients in each of our 48 models, with the results from each model weighted by its posterior model probability.[12] These summary coefficients and their standard errors are presented in Table 2.

The model-averaged regression coefficients in Table 2 represent our best estimates of the distinct effects of each of our explanatory variables. Since six distinct economic variables figure in our analysis, any calculation of the total impact of economic conditions on presidential election outcomes must include all six distinct effects. However, it is clear from the magnitudes of the summary coefficients that the three GDP effects are all trivially small, reflecting the predominance of RDI over GDP as a politically relevant economic indicator in our analysis.

On the other hand, the summary coefficients for the three distinct RDI variables suggest that the time horizons they represent must all be given considerable weight in our account of election outcomes. The relatively large standard errors of the estimates reflect uncertainty, not about the general electoral significance of RDI growth, but about the precise magnitudes of the distinct contributions of each specific variable. This uncertainty limits our ability to say exactly how the declining economy affected Gore's vote share in 2000. Our best guess, however, is that the slowdown was a major factor in Gore's defeat. Had real income continued to grow through Election Day at even the moderate annual rate observed through the first half of the year (about 2%), our estimates suggest that Gore would have won an additional half percent of the popular vote. Thus, the long economic boom that arrived just a little too late to get George Bush reelected in 1992 seems to have ended just in time to elect his son in 2000.

Peace, Moderation, and Incumbent "Fatigue"

Our regression models include three explanatory variables in addition to economic performance: a dummy variable for war years, a variable measuring the relative ideological moderation of the two parties' presidential candidates, and a variable indicating how long the incumbent party has held the White House. The parameter estimates for these three variables, and their standard errors, appear in Table 1, with weighted averages appearing in Table 2.

Of our three political variables, the one that probably requires the least explication is war years. We count Korea in 1952 and Vietnam in 1968 as politically costly wars for the incumbent party. We do not count Vietnam in 1972 (since Nixon had inherited the war and had gone a considerable way toward winding it down by the time of the 1972 election); nor do we count the Gulf War in 1992 or any of the many other military events of the post-war era that did not generate significant American casualties.

The war dummy variable has the expected negative parameter estimate in each of the 24 models in Table 1 in which it appears. The (weighted) average parameter estimate, based upon all of the models we examined, is 3.94 percentage points—an impact too small to have swung the 1952 election to Eisenhower, but certainly big enough to have been decisive in 1968.

Our second political variable, ideological moderation, is especially hard to pin down, given the notorious complexity and abstraction of political ideology and the careful efforts of politicians to obscure potentially disadvantageous ideological extremism. Notwithstanding these difficulties, ideological positioning is so central to the dynamics of American politics that it seems well worth doing the best one can to measure its effects.

We began by rating each major-party presidential candidate on a left-right scale whose zero-point corresponds to the ideological position of the median voter in a given election year. We then subtracted the absolute value of the incumbent candidate's score from the absolute value of the challenger's score to produce a relative moderation variable with positive scores indicating that the candidate of the incumbent party is more moderate than his opponent. Thus, we expected the coefficient on the relative moderation variable to be positive.

We based our candidate ratings upon a combination of expert ratings of candidates in elections between 1948 and 1980 (Rosenstone 1983) and survey data on candidates between 1984 and 1996 (from better-informed respondents in American National Election Studies surveys in those years). Because the relevant survey data for 2000 are not yet available, we were forced to rely upon our own wholly subjective (and correspondingly tentative) ideological ratings of Bush and Gore. (These ratings do not enter into our main regression analysis, but do affect our point estimates of the outcome of the 2000 election.) We rated Bush, who emphasized tax cuts and education reform, as somewhat closer to the center of American politics (at least, on the issues that proved to be most salient in the 2000 campaign) than Gore, who launched populist attacks on "big business" and expressed a preference for government spending over tax cuts.[13] Neither candidate, however, seems to us to have been very far from the center of the ideological spectrum, and the difference between them was also relatively modest, about half as large as the average difference for the previous 13 presidential elections.

As with the war dummy variable, the ideological moderation variable has the expected (positive) coefficient in each of the 24 distinct models in which it appears in Table 1, and in the summary (model-averaged) results presented in Table 2. The magnitude of the weighted-averaged parameter estimate suggests that a candidate who was as extreme, relative to his opponent, as Goldwater in 1964, McGovern in 1972, or Reagan in 1980 would stand to lose almost three percentage points of the popular vote. More typical relative moderation effects are on the order of one

percentage point, and our provisional ideological ratings for the 2000 race suggest that Bush gained a bit more than half a percentage point by being modestly closer to the median voter than Gore was.

Despite its relatively modest impact on the 2000 election outcome, moderation has an important impact on our overall results. Many of the parameter estimates in Table 1 are affected by the inclusion of moderation, and most of the right mode in Figure 2 is contributed by results from models including moderation; the weighted 2000 residual for the 24 models omitting moderation is –2.55, compared to –.56 for that of all 48 models.[14]

Our third political variable, the number of consecutive terms the incumbent party has held the White House, is much easier to measure than ideological moderation, but represents a more complex combination of political forces. We offer the term "incumbent-party fatigue" as a shorthand summary of those forces. The longer an incumbent party has been in power, the more likely it is that political innovators will give way to less skillful successors (for example, Eisenhower to Nixon, Reagan to Bush, and Clinton to Gore). Seasoned advisers are likely to burn out and be replaced by second-stringers. Scandals tend to accumulate, as with the "corruption" of the Truman era, Watergate, Iran-Contra, and the various scandals of the Clinton years culminating, eventually, in the president's impeachment. Meanwhile, "easy" issues are likely to be dealt with and disappear from the political agenda, leaving increasingly intractable problems and increasingly disaffected constituency groups— the electoral equivalent of Mueller's (1973) "coalition-of-minorities" effect.

On the other hand, the longer the out party has been cut off from power, the more eager its core supporters are likely to be able to regain the White House—even if doing so requires them to modify regid ideologies and replace familiar party leaders with more appealing figures. Thus, for example, the Democrats' long lock on the presidency during the New Deal era prompted the Republicans of the 1940s and 1950s to pass over "Mr. Republican" Robert Taft, in favor of popular moderates Thomas Dewey and Dwight Eisenhower. The same Republicans who nominated a veteran ideologue, Robert Dole, in 1996 flocked to a moderate new face, George W. Bush, in 1999. Bush's primary selling point was that he could win back the White House for the Republicans. The impact of that consideration is likely to vary in direct proportion to the length of a party's spell in the political wilderness. In that sense, part of the apparent effect of ideological moderation in our regression analyses should probably be thought of as an indirect effect of incumbent-party fatigue.

The estimated effect of incumbent terms is negative, as expected, in all but one of the 24 regression models in which it appears in Table 1, and the summary (model-averaged) coefficient in Table 2 suggests that each additional term in office costs the incumbent party about half a percentage point of the popular vote. Thus, the Democrats' 20-year run from 1932 to 1952 probably cost Stevenson about two percentage points in 1952, and the Republicans' 12-year run from 1980 to 1992 probably cost the senior Bush about a point in 1992— not enough to swing either election, but certainly enough to care about. In the close elections of 1960, 1976, and 2000, Nixon, Ford, and Gore probably each did about half a point worse than they would have if their respective parties had only been in power for four years rather than eight, which may have been enough to have made a decisive difference in 1960 and 2000.

The (weighted) average parameter estimates presented in Table 2 indicate that all three of our political variables have impacts that are fairly small. Moreover, the standard errors of those parameter estimates reflect considerable uncertainty about the true magnitudes of the relevant political effects. What should we make of this?

First, we think it is worth noting an important structural explanation for the statistical uncertainty of our results: Politicians work diligently to make it hard for political scientists to identify the political effects they are interested in. Imagine, for example, what would happen if parties frequently nominated ideological extremists like Barry Goldwater or George McGovern. They would lose with impressive regularity, whether or not times were prosperous. However, we only rarely see extremist candidates running on major-party tickets, because rational vote-seeking parties can and do refrain from nominating them. That fact makes the impact of ideological extremism harder to discern, and only rarely decisive, but no less important in a scientific sense.

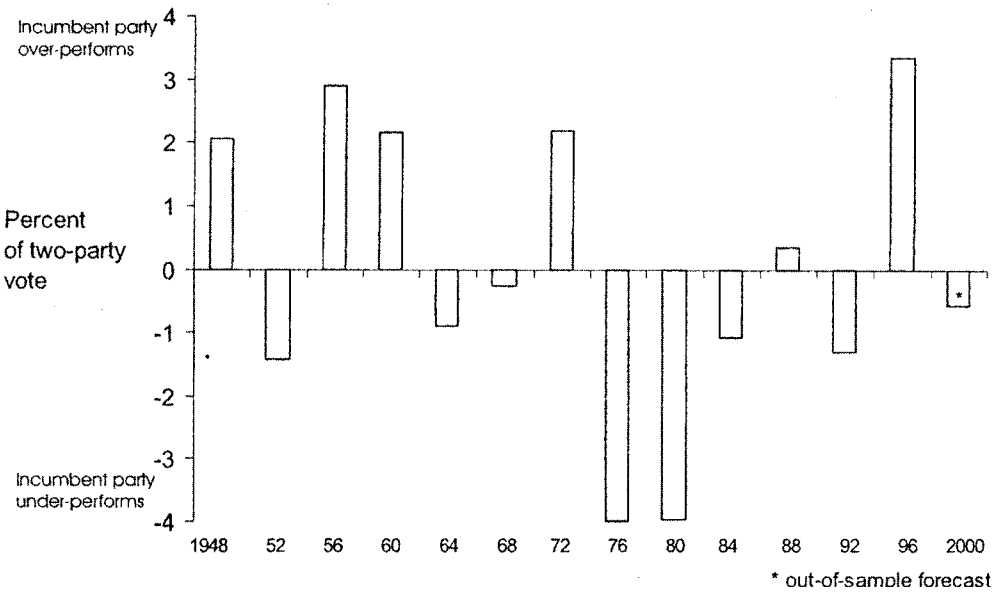
The same point could be made about each of the other political variables in our analysis. Only two costly wars have occurred in the period we considered, at least in part because election-minded politicians try to avoid potentially unpopular foreign entanglements. (And one of the two politically costly wars that did occur, in Vietnam, was probably delayed by the incumbent president's ambition to be reelected.[15]) Similarly, the very play of incumbent-party fatigue tends to limit our opportunity to observe the electoral effects of parties holding the White House for long stretches and, hence, also our ability to detect convincing evidence of the electoral importance of the fatigue factor.[16]

From the point of view of pure election forecasting—simply trying to anticipate what will happen next—variables that don't vary much are unimportant. But for anyone hoping to understand American electoral politics, variables that don't vary much can offer vital insights, provided that they vary to some extent. Economic performance may be the dominant explanatory variable in most presidential vote models, not because the economy is in any deep sense the most important determinant of election outcomes, but, paradoxically, because it is the determinant least susceptible to effective political control.

In any case, it seems clear that the political forces we considered are important. Note, for example, that the eight models in Table 1 that include economic variables but no political variables receive less than one percent of the total posterior model probability in our model-averaging calculations, despite their significant advantage with respect to parsimony. Clearly, our results provide little support for simple-minded economic determinism. Yet the precise impacts of the political

variables remain uncertain and they are far from being statistically significant by any conventional definition. If political scientists expect more than this from small data sets, it is probably because they are accustomed to reading articles by authors more interested in inspiring faith in "the one best model" than in clearly conveying the uncertainty of their inferences.

Figure 4
Weighted incumbent residuals from models of post-war election outcomes



Gore, Bush, and the Lessons of 2000

So how did Gore lose and Bush win? Obviously, in an election as close as the presidential election of 2000, almost any potentially significant factor can be considered decisive. That is certainly true of the factors included in our regression models. Had Gore been as moderate as Bush, our analysis suggests that he would have done more than half a percentage point better and almost certainly would have won, all else being equal. Had economic conditions been as favorable for the incumbent party as in the average post-war election year (that is, significantly better with respect to income growth), our analysis suggests that Gore would have done about one percentage point better, almost certainly enough to win. Had the Democrats been in office for only one term rather than two, our analysis suggests that Gore would have done about half a percentage point better, almost certainly enough to win.

Most commentators on the 2000 election seem to us to have erred in overlooking these fundamental factors and focusing instead upon a variety of more idiosyncratic factors, mostly having to do with Gore's purported failings or strategic errors. Even among those who agree that Gore was "a flawed candidate" (Berke 2000), there is considerable disagreement regarding the precise nature of the flaws that sunk his candidacy. According to some, he foolishly downplayed the nation's economic prosperity. To others, Clinton's high approval rating was the Democrats' key wasted asset. Still others considered Gore too wooden, too pandering, too partisan, or not partisan enough. On the other hand, some observers have attributed the outcome less to any failing on Gore's part than to Bush's genial charisma, "compassionate conservatism," or political sure-footedness.

Detailed analysis of survey data may, in time, shed some useful light on these and other possibilities. For now, however, the very variety of the explanations offered for Gore's failure (or, conversely, for Bush's success) suggests that a good deal of caution is warranted in accepting any or all of these "explanations." It is easy to suggest factors that might have mattered, but notoriously difficult to pinpoint which ones did matter, and how much.

Moreover, it is crucial not to lose sight, in retrospect, of the fact that Bush had his own weaknesses and made his own mistakes. Had he managed to lose an election that seemed in the final week of the campaign to be all but won, there would no doubt have been considerable grumbling about his sometimes-lackadaisical campaign style, his well-publicized verbal miscues, and his rather bizarre decision to devote millions of dollars to television advertising in California in the final weeks of the campaign with no apparent effect. Bush's failings are much less salient than Gore's because Bush won and Gore lost; but it is not at all clear that an observer unaware of the election outcome could find compelling grounds to consider Bush the better candidate.

In any case, our primary point is that appeals to election-specific explanations of the 2000 result are quite superfluous. Given the systematic, predictable effects of the economic and political "fundamentals" embodied in our 48 regression models, there is simply nothing special about the 2000 election outcome to be explained. Indeed, to a greater extent than most elections, 2000's really did come out the way it "should" have, at least with respect to the national popular vote.

Figure 4 makes this point very clearly. The figure displays the model-averaged residuals for each of the 13 presidential elections from 1948 to 1996, plus the model-averaged out-of-sample forecast error for 2000. Most of the residuals are quite modest in magnitude; only those for 1976, 1980, and 1996 exceed three percentage points. Clearly, big surprises have been rare throughout this period.[17] But even by this historical standard, the model error for 2000, –.56, is conspicu-

ously small.[18] Notwithstanding Gore, Bush, Nader, and everything else that was unique about the 2000 election, the election outcome fit the historical pattern almost exactly. In this important sense, the 2000 election seems to have very little to teach election forecasters.

Conclusion

Contrary to most observers, we find little to be surprised by in the outcome of the 2000 presidential election. Of course, we do not pretend to be able to account for the razor-thin vote margins in several states, much less for the extraordinary postelection political and legal wrangling required to adjudicate the eventual outcome. But we do claim that the overall closeness of the fourteenth presidential election of the post-war era reflected the same fundamental economic and political factors at work in the first 13.[19] We see no need, and little warrant, to posit either unusual incompetence on Gore's part or unusual skill on Bush's part. Likewise, we see no need, and little warrant, to posit any significant change in established patterns of voting behavior. Our rather unremarkable conclusion is that voters in 2000 behaved much as they have in previous election years—and that political scientists' understanding of presidential elections requires fine-tuning, not wholesale rethinking.

Notes

* We thank Adam Berinsky, James Campbell, David Karol, Michael Rothschild, Lynn Vavreck, and Christopher Wlezien for helpful comments on an earlier draft.

1. We ignore votes for minor-party candidates, even when they are numerous or appear to be drawn disproportionately from one or the other of the major-party contenders.

2. These variables figure in our analysis primarily because we happen to have employed them in previous work (Bartels 1992; Bartels 1997b; Zaller 1998, 1999). A more comprehensive consideration of the wide range of explanatory variables employed in election forecasts would be desirable, but is beyond the scope of the present analysis.

3. All economic data are from the November 29, 2000 release of data on the web page of the Bureau of Economic Analysis (see <www.bea.doc.gov/dn1.htm>). The transformation of these raw data into the economic data we used in our analysis are documented on Zaller's web page in the spreadsheet titled "Dataprep2000." The construction of our ideological moderation variable is documented in the same place (see <www.sscnet.ucla.edu/polisci/faculty/zaller/>).

4. The formula for calculating the Bayes factors is

$$\exp\left(-\frac{n}{2}\ln[1-R^2]-\ln[N]\frac{K-1}{2}\right)$$

where N is the number of observations in the regression analysis, K is the number of parameters estimated, and R^2 is the coefficient of determination.

5. The Bayes factors are analogous in this sense to adjusted R^2 values, except that they impose an even greater penalty for estimating additional parameters.

6. To keep things simple, we treat each of our 48 models as equally credible *a priori*. Given equal prior probabilities, the posterior model probabilities simply normalize the Bayes factors defined in note 4 to sum to 1.0 across the entire set of models under consideration. Since the data strongly favor some of our models over others, assigning unequal prior probabilities to different models would not significantly affect most of our conclusions.

7. The model estimated by Hibbs is actually slightly more complicated than the version estimated here, since it includes a measure of cumulative combat deaths rather than the simple dummy variable for wars in 1952 and 1968 employed here. In addition, we use more recent economic data than Hibbs did, and include the 1948 election, which Hibbs ignored due to limitations in the available economic data. (See the "Dataprep2000" spreadsheet on Zaller's webpage for detailed information on how we finessed this data limitation.) These differences account for discrepancies between our results and those reported by Hibbs; we successfully replicated his results using his data and procedures.

8. In some cases, these calculations are based on data not available until after the 2000 election or revised by the BEA after the election. But they are honest out-of-sample forecasts in the limited sense that the 2000 election is not included in the set of elections used to estimate the regression coefficients reported in Table 1.

9. Each forecast error is represented in Figure 2 by a distribution rather than a point because the parameter estimates used to calculate the forecast error are themselves random variables.

10. As Samuelson and Thomas put it in a pithy summary published in *Newsweek* a month after the election, "The economy has made an abrupt shift in a new direction—south. A steady stream of negative data has been released in recent weeks, providing sobering evidence of widespread softness" (2000, 52). The *New York Times* was equally categorical, reporting that "The economy is clearly faltering, on all fronts" (Uchitelle 2000). By year's end, 2000 was being referred to as "the case of the good economy gone bad. . . . In the first half of 2000, the American economy was growing at more than 5 percent annually. This quarter, it will be lucky to expand at all" (Berenson 2000).

11. It is worth noting, however, that measures of economic perceptions such as the University of Michigan's survey of Consumer Sentiment (<http://athena.sca.isr.umich.edu>) remained highly positive through Election Day, which raises some doubt about the extent to which voters reacted to the slowdown in income growth evident in the BEA data. The relationship between objective economic conditions and subjective economic perceptions deserves more analysis, both in this case and more generally.

12. Many of the coefficients entering into each of these average values will be zeros, since each of our explanatory variables is excluded from at least half of our 48 models. These zeros need not be damaging to the average coefficient for a variable if the variable has an important impact, since models in which it does not appear will then get much smaller weights than models in which it does appear. Exclusions per se do not hurt average performance; exclusions that don't matter do. The Bayesian model-averaging scheme is tough on variables that can be safely omitted from models without significantly affecting their performance.

13. To see the numerical values of the moderation variable and how they were calculated, see the "Dataprep2000" spreadsheet at <www.sscnet.ucla.edu/polisci/faculty/zaller/>. We shall revise our relative moderation score for the 2000 election (and make the revised data available on Zaller's web page) when NES data on the perceived ideological positions of Bush and Gore become available.

14. Notwithstanding Rosenstone's (1983) demonstration of the importance of ideological positioning, the moderation variable is not routinely included in presidential vote models. This might prompt suspicion that we dragged it out of the closet simply to get the right result for 2000 or, more charitably, that we are just lucky it happens to be one of our favorites. The model-averaging methodology, by demonstrating the importance of moderation in a range of specifications, should mitigate this suspicion—and ought to do still more to establish the actual importance of ideological positioning once this variable is further refined and tested in a wider range of specifications.

15. It is arguable that American entry into World War I was delayed for the same reason. More generally, Gaubatz (1990) found that, for the period from 1838 to 1973, democracies were less likely to initiate wars in election years than at other times.

16. Our confidence regarding the electoral significance of war and incumbent-party fatigue is reinforced by a simple auxiliary regression

analysis covering the 25 presidential elections held between 1900 and 1996 and employing those two variables along with election-year changes in GNP per capita as explanatory variables. The resulting regression parameter estimates for war (4.46, with a standard error of 3.85) and especially incumbent terms (2.95, with a standard error of 1.29) are larger than those reported here, and larger in the first half of the century than in the second half.

17. One might turn to the history books for explanations of the residual variance that does appear in Figure 4. For example, the 1976 and 1980 results might be attributed to Watergate and the Iranian hostage crisis, respectively. Alternatively, one might be tempted to discern the impact of candidate quality, with Eisenhower, Nixon, and Clinton each doing better than expected every time they ran. Of course, such retrospective explanations are subject to the same problems of *post hoc* inference we have already noted in explanations of the 2000 result.

18. When we include the 2000 result in our regression analyses, the model-averaged residual is –.13, even smaller in magnitude than the model-averaged out-of-sample forecast displayed in Figure 4, and smaller than for any other post-war election.

19. To be sure, our analysis rests on economic data from the fourth quarter of 2000 and ideological rating data that are still just estimates. We anticipate posting a revised version of our results on Zaller's web page in the summer. However, it seems quite unlikely that revisions to our data will fundamentally upset our interpretation of the 2000 outcome.

References

Bartels, Larry M. 1992. "The Impact of Electioneering in the United States." In *Electioneering: A Comparative Study of Continuity and Change*, ed. David Butler and Austin Ranney. Oxford: Clarendon Press.

———. 1997a. "Specification Uncertainty and Model Averaging." *American Journal of Political Science* 41:641-74.

———. 1997b. "Econometrics and Elections." *Journal of Economic Perspectives* 11:195-97.

Berenson, Alex. 2000. "The Case of the Instant Recession." *The New York Times*, December 31, 4:10.

Berke, Richard L. 2000. "Many Seem Skeptical of Gore's Future." *The New York Times*, December 17, A1.

Gaubatz, Kurt T. 1990 "Election Cycles and War." *Journal of Conflict Resolution* 35: 212-44.

Hibbs, Douglas A. Jr. 1987. *The American Political Economy: Macroeconomics and Electoral Politics.* Cambridge, MA: Harvard University Press.

———. 2000. "Bread and Peace Voting in U.S. Presidential Elections." *Public Choice* 104: 149-80.

Miller, D.W. 2000. "Election Results Leave Political Scientists Red-Faced Over Their Forecasting Models" <http://chronicle.com/free/2000/11/2000110807n.htm>. *The Chronicle of Higher Education*, November 8. Accessed: January 21, 2001.

Mueller, John E. 1973. *War, Presidents and Public Opinion.* New York: John Wiley & Sons.

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

Samuelson, Robert J., and Rich Thomas. 2000. "Chugging Toward a Slowdown." *Newsweek*, December 18, 52-54.

Uchitelle, Louis. 2000. "Unfortunate Timing, No Matter Who Wins." *The New York Times*, December 3, 3:4.

U.S. Bureau of Economic Analysis. 2000. "Personal Income and Outlays: October 2000" <www.bea.doc.gov/newsrel/pi1100.htm>. Press release, November 30. Accessed: January 21, 2001.

Zaller, John. 1998. "Monica Lewinsky's Contribution to Political Science." *PS: Political Science and Politics* 31(June): 182-89.

———. 1999. "Know Nothing Voters in U.S. Presidential Elections, 1948-1996" <www.sscnet.ucla.edu/polisci/faculty/zaller/>. Manuscript.