Forecasting the 2012 Presidential Election With State-Level Economic Indicators

Michael J. Berry, *University of Colorado, Denver*

Michael.Berry@ucdenver.edu

Kenneth N. Bickers, *University of Colorado, Boulder*

Kenneth.Bickers@colorado.edu

Nearly all forecast models of US presidential elections provide estimates of the national two-party vote (Campbell 2008). Each of the nine forecasts published in the 2008 forecasting issue of *PS: Political Science & Politics* made national popular vote total predictions for the major party candidates, while only one provided an expected result in the Electoral College (Klarner 2008). These national vote models are assumed to be reliable forecasts of who is likely to win the general election. In most cases, this assumption is reasonable. It becomes problematic, however, at precisely the point that forecasts are most interesting: when elections are close. In tight elections, national forecasts can and have produced a "winner" different from the actual winner. Consider the forecasts and ultimate outcome of the 2000 election. Each of the 2000 presidential election forecasts predicted vice president Al Gore to win a majority of the two-party popular vote, which he did, but none correctly predicted governor George W. Bush to assume the presidency (Campbell 2001). Never in US history have White House residents been determined through a national popular vote. Presidential elections are decided through contests in the states and the District of Columbia. The forecast model we developed explicitly models the presidential contest based on factors inherent to these 51 jurisdictions. This modeling approach allows us to make a projection of the Electoral College result, which popular vote estimates cannot.

In its theoretical approach, our State-by-State Model is similar to national two-party forecasts that primarily focus on economic conditions (Abramowitz 2008; Cuzán and Bundrick 2008; Erikson and Wlezien 2008; Holbrook 2008; Lewis-Beck and Tien 2008). Methodologically, our model stands as an alternative to those of Campbell (1992), Cohen (1998), DeSart and Holbrook (2003), Holbrook and DeSart (1999), and Klarner (2008), which, like ours, predict outcomes on a state-by-state basis. Most of these models are based on horse-race public opinion polls in each state in the one to two months prior to the general election. When well

done, these polls provide likely voter responses to many factors influencing support for candidates in a state, including economic conditions, as well as factors that we are unable to observe. However, this strategy is dependent on the timeliness and quality of publicly available polls asking horse-race questions in each jurisdiction. Alternatively, Campbell (1992) and Klarner (2008) use a battery of polling, political, and contextual variables to forecast the presidential vote, but each model includes just a single state-level measure of economic conditions.

In contrast to these other Electoral College models, our model includes measures of change in real per capita income, as well as national and state unemployment figures. Accounting for both changes in personal income and unemployment provides a more robust approximation of state economic well-being and, thus, serves to model the impact of retrospective evaluations of the incumbent party's stewardship of the economy (Fiorina 1981). The data incorporated in our model are regularly released by the Bureau of Economic Analysis (BEA) in the US Department of Commerce and the Bureau of Labor Statistics in the US Department of Labor. This gives us high-quality, predictably available data to use as the feedstock for our models.

Data

The state-level economic indicators we use are available as far back as the 1980 election. Consequently, our estimations incorporate eight elections in 51 jurisdictions—a data set substantially wider spatially, but shorter temporally, than those used by most forecast models. With the resulting number of observations, we use a larger number of variables without loss of an excessive number of degrees of freedom. Our dependent variable is the incumbent party candidate's share of the two-party vote in the state.[1] In other words, our forecast predicts the

Democratic vote share when the office holder at the time of the election is a Democrat. Likewise it predicts the Republican share, when the office holder at the time of the election is a Republican. Based on a model estimating the incumbent party share of the two-party vote in each state, we use contemporary data to forecast the 2012 presidential election. Then, we use these state-specific point estimates to predict the number of Electoral College votes the major party candidates will obtain.[2] Independent variables incorporated in the model fall into four categories.

One, we control for the normal two-party vote in the state with the inclusion of the two-party vote in the prior presidential election in that same state. This variable captures much of the underlying partisan distribution in each state. Two, we include variables to identify the incumbency conditions that exist in a given election. Because of the strong possibility that one party "owns" certain economic issues more than the other party (see, for example, Petrocik, Benoit, and Hansen 2003/2004), we include an incumbent party binary variable that is coded 1 when Democrats hold the White House at the time of an election and 0 when Republicans do. This incumbent party variable is interacted with each of our variables that tap economic conditions, as discussed in the next paragraph. We also control for the term that the incumbent party is seeking. We consider two possibilities: incumbent parties seeking a second presidential term and incumbent parties seeking a third or higher term in the White House. This occurred most recently in 2008 when senator John McCain ran for a third consecutive presidential term for the Republican party. In our modeling approach, only one binary variable is required to capture these alternatives. If a candidate is seeking a second consecutive term for the candidate's party, we code as 1 a second-term contest, 0 otherwise. Notice that if a candidate is of the same party as the incumbent in the past two (or more) presidential terms, the second-term contest variable is coded as 0. This was the case for president George H.W. Bush, vice president Al

4

Gore, and senator John McCain. This variable continues to be coded as 0 if candidates seek a fourth or even fifth consecutive term for their party.

The heart of our forecast centers on the third set of independent variables. We use two basic measures of economic conditions: unemployment levels and change in real income per capita. Unemployment is measured in two capacities. First is the national unemployment rate. The second is the corresponding unemployment rate in each state. Operationally, we use the U3 measure of unemployment, which is the "headline" unemployment figure most often reported by the media. It is typically released by the Bureau of Labor Statistics for the nation as a whole on the first Friday of each month for the prior month and at the state level on the third Friday of each month. Both the national unemployment rate and the state-level unemployment rate are interacted with the incumbent party binary variable. In our preliminary model, which is reported here, we use May 2012 unemployment figures.[3] While these data provide measures of nationwide and state-specific economic well-being a full five months in advance of the general election, American National Election Studies (ANES) survey data from the past several election cycles report that nearly two-thirds of voters had determined their presidential vote choice before Labor Day, the traditional kick off of the general election campaigns. Thus, although changing economic conditions during the run-up to the election may weigh more heavily on voters' minds, these survey data suggest that the proportion of undecided voters is substantially reduced as the election approaches.

Beyond the state and national unemployment figures, our third measure of economic well-being taps the extent to which people have more or less real disposable income at their discretion during the current incumbent's presidential term. The measure included in our model is the percentage change in each state in real per capita non-farm income from the fourth quarter

of the prior presidential election year to the first quarter of the current election year. These quarterly data are released by the BEA of the US Department of Commerce typically during the last week of the following quarter. The BEA releases the data in current dollars. To put these in constant dollar terms, we used the national GDP implicit price deflator for each quarter (2005 dollars = 1.00). To standardize the figures, income is divided by the population of the state using US Census estimates of population in each year. One caveat is that the Census, as of this writing, has not yet published estimates of population by state for 2012. As a proxy, we use 2011 population counts for the calculation of per capita income change for 2012. We also include the interaction of this variable measuring change in real per capita income with the incumbent party binary variable. Again, our logic is that one incumbent party may be harmed more by negative economic performance on this measure than will the other party.

The fourth category of independent variables is included to capture state-to-state idiosyncrasies of a given presidential contest. In doing this, we include binary variables identifying the home state of the Democratic and Republican presidential candidates in each contest. As shown later in the text, we find evidence that some candidates perform better, on average, in their home states. Therefore, we also include binary variables to identify the home states of the candidates in the last election. The inclusion of these variables is necessary because, absent such controls, the lagged two-party vote percentage will over predict the current vote for that party's candidate in any state in which the prior election featured a major party candidate who hailed from that state. In essence, we "turn off" the prior home state effect in predicting current support for the incumbent party's nominee. In earlier iterations of the model, we also included binary variables to identify vice presidential candidate home states, the states in which the nominating conventions were held, and governor partisanship (see also Powell 2004).

Despite frequent media speculation that such things play a role in the final outcome, no statistically significant effect of any of these binary variables in the models that incorporated these variables or subsets of these variables is found.

The Model

Table 1 presents the regression model estimates for the full model.[4] Statistically, the model ably estimates the vote in the states. One indicator of this is the model's R-squared, which equals 88.7%. As table 1 indicates, nearly all of the variables play a statistically significant role in determining the two-party vote. The lagged presidential vote in each state is highly determinative when controlling for other variables. The coefficient on this variable indicates that for each point received in a state by the current incumbent's party in the prior election, the incumbent party's candidate will garner 0.994 points of the two-party vote in that state. The second-term incumbency variable is likewise very strong, indicating that incumbents seeking a second term enjoy a 9.5 point advantage over candidates seeking a third or higher term of office for their party.

[Table 1 about here]

Also, it initially appears that Democratic candidates running when the Democrats are in the White House enjoy a huge advantage over all other candidates. The coefficient on that binary variable is almost 18.5 points on the two-party vote. Notice, however, that this variable is interacted with all three of the economic variables. When considered in that light, the apparent Democratic advantage turns into a contingent advantage, specifically an advantage when the national unemployment rate is below its long-term average. The interaction of the national unemployment rate with the incumbent party variable (coded 1 when Democrats are the in-party candidates) is both significant and large in magnitude. It shows that Democrats lose 3.3 points on

7

the vote for every percentage point that unemployment rises above zero. Or, in other words, the apparent advantage of being a Democratic candidate and holding the White House disappears when the national unemployment rate hits 5.6%. Beyond that, support for Democratic in-party candidates continues to drop. For Republicans running as the in-party candidates, the impact of the national unemployment rate is not significant.

State-level unemployment also factors into the equation, but at the margins of the national unemployment rate. Specifically, when Republicans are running as in-party candidates, the impact on the vote of state-level unemployment for the Republican is actually positive, although at a modest level even if statistically significant. Where Republicans running as in-party candidates are helped or harmed is through changes in real per capita income in each of the states. For every percentage point increase in real per capita income, the vote for Republican in-party candidates increases by nearly a quarter point. The reverse is also true; falling real income cuts into Republican support. Furthermore, Democrats running as the in-party candidates do not benefit from rising real per capita income. Their opponents do. Republican out-party candidates do better in states where real per capita income has increased. This effect is less than the benefit received by Republicans running as the in-party candidates, although still significant. Finally, a statistically significant home-state advantage exists for the out-party challengers, but not for candidates from the incumbent party. Yet this may be slightly misleading. Removing home-state effect from the prior election cycle provides a corresponding drop in support in the states that both the challenger and the in-party candidate called home in that election.

Putting these pieces together, clearly the President Obama is in electoral trouble. To be sure, he enjoys some advantages. First, Obama's successful campaign in 2008 gives him a substantial leg up. He can lose some states that he carried four years ago without losing the

election. Second, a prominent second-term incumbency advantage should prove advantageous. Still, the big issue is the fragile economy. With an unemployment rate in excess of 8%, Obama is about two-and- a-half points beyond the break-even point for a Democrat running as the in-party candidate. This situation translates into slightly more than an eight-point reduction in his two-party vote, wiping out virtually the entire bump accruing to an incumbent seeking a second term. Moreover, as the country continues to rebound from the largest recession in generations, whether voters will ultimately judge the economy in relative or absolute terms is unclear. Beyond economic considerations, benefits of home-state advantages basically are lost in the 2012 election. Illinois is predicted to go for Obama, regardless of any home-state advantage. Romney may do better in Massachusetts, but our prediction is that his vote share in that state will be so low that any home-state bump remains inconsequential. Likewise, the drop in support for the Republican candidate in Arizona is unlikely to offset the margin that the GOP enjoys in that state even when the candidate does not call the state home.

Reapportionment following the 2010 Census provides another interesting aspect to the 2012 election. In the 2008 election, Obama won eight of the 10 states that lost House seats, and, therefore, Electoral College votes. If Obama carried the exact same coalition of states he won four years ago, he would receive six fewer electoral votes. As recent history suggests, a handful of electoral votes can be the difference between winning and losing.

Model Diagnostics

Table 2 presents model diagnostics regarding the accuracy of the model during the eight elections from 1980 through 2008. The most important point to emphasize from this table is the final column: the model successfully predicts every Electoral College victor. The 2000 election is of particular interest because no forecasting model published in advance of that election

correctly predicted George W. Bush as the winner. Because our model is predicated on the notion that during close elections, the Electoral College winner may not win the popular vote, it is critical that our forecast classify this election accurately. In 2000, the model correctly classifies 47 states, most notably Florida, which Bush was estimated to win with 51.2% of the two-party vote. The state's final certified result awarded Bush a razor thin majority of the two-party vote at 50.004%. The only states incorrectly predicted in 2000 are Pennsylvania, West Virginia, Arkansas, and Louisiana. Despite these inaccuracies, the model predicted Bush to win 274 electoral votes to Gore's 264, an error of a mere two votes.

[Table 2 about here]

In 2008, the model preformed similarly well, predicting 48 states correctly and missing Obama's actual Electoral College vote total of 365 by just five votes. For comparison, Klarner's (2008) median estimate of Obama's electoral vote total was 346. The average error rate during the past four election cycles is about four states. Including all eight election cycles, the successful prediction rate is nearly 90%, with an average deviation from the actual Electoral College result of 21.3.

The state-by-state forecasts for the elections of 1980 and 1992 have the most errors, with nine and 10, respectively. Of course, in these two elections independent candidates performed well. John Anderson received about 6.5% of the popular vote in 1980 and Ross Perot received nearly one out of every five votes cast for the presidency in the 1992 election. Although our model is constructed to be insulated by the presence of independent or minor party candidates by forecasting the two-party vote share for the Democratic and Republican candidates, strong showings by insurgent candidates can affect the winner of statewide elections. Many of the prediction errors in these two elections occurred in extremely competitive states. In the 1980

election, for example, our model predicts that President Carter would win Tennessee, Massachusetts, Alabama, and South Carolina with an average two-party vote share of 50.4%. Carter ultimately lost these states receiving vote shares of 49.9, 49.9, 49.3, and 49.2%, respectively. The 1992 election offers similar examples. In each case, changes in a small fraction of the independent candidate could easily have swapped states from one candidate's column to the other.

The Forecast

Our prediction, based on the model analyzing returns from the prior eight presidential elections is that the president will win 17 states, plus the District of Columbia. The point predictions for every state are listed in table 3. Figure 1 provides a graphic depiction of these predictions along with 90% confidence interval bands, which illustrate the degree of uncertainty around each expected result.

[Figure 1 and Table 3 about here]

As figure 1 shows, the states we predict President Obama will carry include a substantially reduced set than those he carried in 2008.[5] This is supported by the fact that no states won by McCain are predicted to flip to Obama. What is striking about our state-level economic indicator forecast is the expectation that Obama will lose almost all of the states currently considered as swing states, including North Carolina, Virginia, New Hampshire, Colorado, Wisconsin, Minnesota, Pennsylvania, Ohio, and Florida. Three other states that might be viewed as swing states— Michigan, New Mexico, and Nevada—are predicted to stay in Obama's column. Our forecast is that the President will receive 218 Electoral College votes, putting him well short of the 270 needed to win reelection.

11

Finally, although not our primary objective, we can use state-by-state vote projections to generate a forecast of the national popular vote as well. The model generates vote share percent predictions for each candidate, but we lack preelection data on the number of voters likely to cast ballots in each state necessary to tabulate the aggregate number of votes received by each candidate. Without such contemporaneous data, we use state-level turnout data from the prior presidential election as a proxy for the expected turnout in the upcoming election. Turnout at time t–1 provides a reliable estimate for turnout at time t on account of the average in-sample correlation of 0.94 between the two factors across all states. To briefly demonstrate this process, about 2.4 million voters cast a presidential ballot in the state of Colorado in 2008. As seen in table 3, our 2012 forecast expects Obama to receive 48.19% of the state's two-party vote. Using 2.4 million voters as the expected turnout for 2012, a 48.19% vote share corresponds to an estimated 1.15 million votes for Obama in Colorado. Aggregating these values provides a predicted vote total for each candidate, which can be expressed as a percentage.

The popular vote predictions generated for each of the eight most recent election cycles are incredibly accurate with an average deviation from the actual result of a mere 0.6%. In every election, except for 2008, the popular vote forecast is within one percentage point of the true result. In 1980, 2000, and 2004, the forecast error rate is less than 0.3%. The greatest deviation from the actual popular vote occurred in 2008 when our forecast overestimated the Obama vote by 1.3%. Thus, while this study's intention was to construct a state-level model to generate accurate Electoral College result predictions, the model does a good job estimating the national popular vote percentage over the 1980 through 2008 period as well.

Moving to 2012, our forecast is that Obama will receive 47.14% of the two-party popular vote. Using confidence intervals around each individual state forecast and aggregating to a

national popular vote as earlier described, our model projects a likelihood of 77% that Romney will receive a majority of the ballots cast for the two major parties. Of course, this does not mean we possess the same level of certainty that Romney will ultimately win the election because minor changes in the distribution of votes across a handful of battleground states can affect the outcome. This justifies using a state-level model to predict Electoral College results. Transforming our state-by-state forecasts into a national popular vote prediction does, however, allow for a comparison of our model to other models that focus on popular vote predictions.

What caveats should be noted? And how do we arrive at this forecast? The first caveat is that we have slightly less confidence in this forecast than one that uses economic data measured somewhat closer to Labor Day. The second caveat, which ties back to the first, is that a substantial number of cases depicted in figure 1 where the 90% confidence band around the state's prediction includes the 50% mark. This indicates that the two-party vote could plausibly flip to the other side of the 50-50 line on which some of these states are currently predicted to land. Finally, our model performs well in estimating election outcomes from 1980 through 2008. However, as scholars and pundits know, each election has unique elements that, while they may randomize over time, could lead one or more states to behave in ways in a particular election that the model is unable correctly to predict. As the great Yankee catcher Yogi Berra famously quipped, "It's tough to make predictions, especially about the future."

References

Abramowitz, Alan I. 2008. "Forecasting the 2008 Presidential Election with the Time-for-Change Model." *PS: Political Science & Politics* 41(4): 691–96.

Campbell, James E. 1992. "Forecasting the Presidential Vote in the States." *American Journal of Political Science* 36(2): 386–407.

Campbell, James W. 2001. "Taking Stock of the Forecasts of the 2000 Presidential Election." *American Politics Research* 29: 275–78.

Campbell, James W. 2008. "Editor's Introduction: Forecasting the 2008 National Elections." *PS: Political Science & Politics* 41(4): 679–82.

Cohen Jeffrey E. 1998. "State-Level Public Opinion Polls as Predictors of Presidential Election Results." *American Politics Quarterly* 26(2): 139–59.

Cuzán, Alfred G., and Charles M. Bundrick. 2008. "Forecasting the 2008 Presidential Election: A Challenge for the Fiscal Model." *PS: Political Science & Politics* 41(4): 717–22.

DeSart, Jay A., and Thomas M. Holbrook. 2003. "Statewide Trial-Heat Polls and the 2000 Presidential Election: A Forecast Model." *Social Science Quarterly* 84(3): 561–73.

Erikson, Robert S., and Christopher Wlezien. 2008. "Leading Economic Indicators, the Polls, and the Presidential Vote." *PS: Political Science & Politics* 41(4): 703–08.

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

Holbrook, Thomas M. 2008. "Incumbency, National Conditions, and the 2008 Presidential Election." *PS: Political Science & Politics* 41(4): 709–12.

Holbrook, Thomas M. and Jay A. DeSart. 1999. "Using State Polls to Forecast Presidential

    Election Outcomes in the American States." *International Journal of Forecasting* 15: 137–

    42.

Klarner, Carl. 2008. "Forecasting the 2008 U.S. House, Senate and Presidential Elections at the

    District and State Level." *PS: Political Science and Politics* 41(4):723–28.

Lewis-Beck, Michael S., and Charles Tien. 2008. "The Job of President and the Jobs Model

    Forecast: Obama for '08?" *PS: Political Science & Politics* 41(4): 687–90.

Petrocik, John R., William L. Benoit and Glenn J. Hansen. 2003/2004. "Issue Ownership and

    Presidential Campaigning, 1952–2000."*Political Science Quarterly* 118(4): 599–626.

Powell, Richard J. 2004. "The Strategic Importance of State-Level Factors in Presidential

    Elections. *Publius* 34(3): 115–30.

Figure 1. 2012 State-by-State Forecasts of Obama's Two-Party Vote Percent
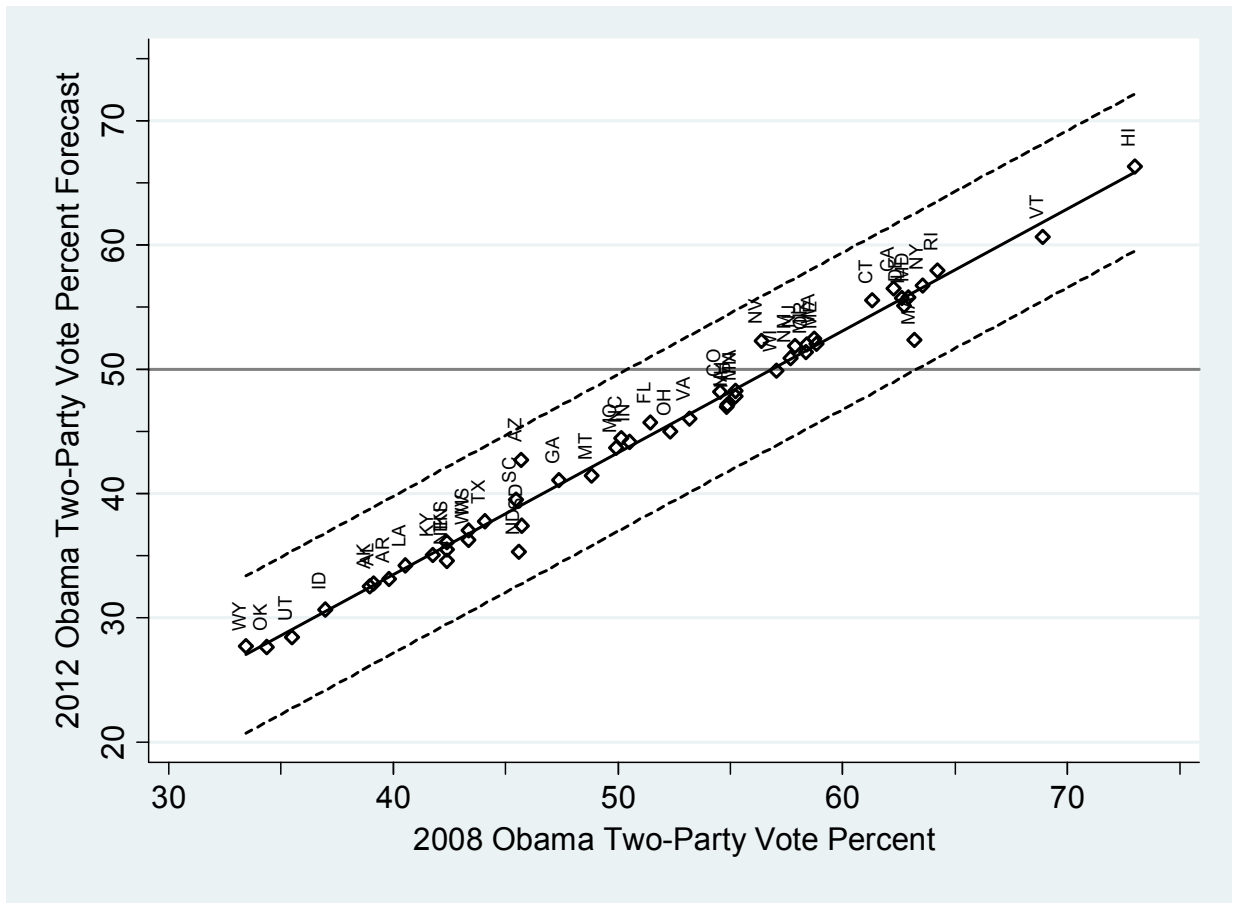
Table 1. State-Level Economic Forecasting Model Regression Analysis: 1980–2008

Independent Variable	Estimate
Lagged State Two-Party Vote Percent	0.99*** (0.02)
First Term Incumbent	9.46*** (0.46)
State Unemployment Rate—May	0.30*** (0.15)
State Personal Income Change—1st Quarter	0.24*** (0.05)
National Unemployment Rate—May	-0.09 (0.27)
In-Party Candidate Home State	2.15 (1.67)
Lagged In-Party Candidate Home State	-3.22* (1.67)
Out-Party Candidate Home State	-3.38*** (1.32)
Lagged Out-Party Candidate Home State	3.83*** (1.34)
Democratic Incumbent	18.53*** (2.18)
Democratic Incumbent * State Personal Income Change	-0.45*** (0.09)
Democratic Incumbent * State Unemployment Rate	-0.04 (0.28)
Democratic Incumbent * National Unemployment Rate	-3.31*** (0.40)
Constant	-8.45*** (1.94)
N	408
Elections	8
R^2	0.8874
Root Mean Squared Error	3.6664

Note: Dependent variable is the two-party vote percent received by the incumbent party. Standard errors reported in parentheses. *$p < .10$; **$p < .05$; ***$p < .01$ (two-tailed).

Table 2. State-Level Economic Forecasting Model Diagnostics: 1980–2008

Year	States Correctly Predicted	Democrat States Predicted	Democrat States Actual	Democrat Electoral College Votes Predicted	Democrat Electoral College Votes Actual	Outcome Correctly Predicted
2008	48	30	29	370	365	Yes
2004	46	15	20	217	251	Yes
2000	47	23	21	264	266	Yes
1996	45	34	32	402	379	Yes
1992	41	29	33	347	370	Yes
1988	45	11	11	171	111	Yes
1984	50	3	2	17	13	Yes
1980	42	8	7	68	49	Yes

Correctly Predicted:	364 (89.2%)	Average Error:	21.3
Incorrectly Predicted:	44 (10.8%)		

Table 3. 2012 State-by-State Forecasts of Obama's Two-Party Vote Percent

State	Percent	State	Percent
Oklahoma	27.68	New Mexico	50.93
Wyoming	27.72	Michigan	51.38
Utah	28.42	New Jersey	51.91
Idaho	30.65	Oregon	52.00
Alaska	32.57	Maine	52.04
Alabama	32.75	Nevada	52.32
Arkansas	33.14	Massachusetts	52.39
Louisiana	34.24	Washington	52.49
Nebraska	34.67	Illinois	55.17
Kentucky	35.08	Connecticut	55.56
North Dakota	35.34	Delaware	55.73
Tennessee	35.47	Maryland	55.82
Kansas	36.09	California	56.52
West Virginia	36.30	New York	56.73
Mississippi	37.06	Rhode Island	57.94
South Dakota	37.41	Vermont	60.69
Texas	37.78	Hawaii	66.34
South Carolina	39.53	District of Columbia	86.36
Georgia	41.09		
Montana	41.46		
Arizona	42.75		
Missouri	43.69		
Indiana	44.17		
North Carolina	44.46		
Ohio	45.00		
Florida	45.75		
Virginia	46.05		
Iowa	46.98		
New Hampshire	47.20		
Minnesota	47.87		
Colorado	48.19		
Pennsylvania	48.24		
Wisconsin	49.92		

NOTES

[1] For expositional purposes, henceforth we refer to District of Columbia as a state.

[2] Two caveats should be noted. First, our model as currently configured cannot predict Electoral College votes for third-party candidates. Thus, were conditions once again to resemble those of 1948 or 1968 when substantial numbers of states cast their Electoral College votes for third party candidates, our model would be limited in its utility. Second, for simplicity, we treat the two states that do not use the unit rule in allocating Electoral College votes (Nebraska and Maine) as if they do. Our model thus miss-predicts one Electoral College vote in 2008 beyond any other estimation errors produced by the model.

[3] Our final preelection model uses unemployment figures closer to Election Day, and the percentage change from the fourth quarter of the prior election year to the second quarter of the current election year.

[4] This model was also estimated using both a random and fixed effects panel regression, each of which resulted in identical results in terms of the sign and significance of the coefficients. Accordingly, we present the OLS regression model results.

[5] The District of Columbia is omitted from figure 1 because it is a significant outlier from the 50 states exhibiting a predicted Obama vote in excess of 20% greater than any state.