# Forecasting the 2008 U.S. House, Senate and Presidential Elections at the District and State Level

by
**Carl Klarner,**
Indiana State University

This paper applies the forecasting models of Klarner and Buchanan (2006a) for the U.S. Senate and Klarner and Buchanan (2006b) for the U.S. House of Representatives to the upcoming 2008 elections. Forecasts are also conducted for the 2008 presidential race at the state level.[1] The forecasts presented in this article, made July 28, 2008 (99 days before the election), predicted an 11-seat gain for the Democrats in the House of Representatives, a three-seat gain for the Democrats in the Senate, and that Barack Obama would obtain 53.0% of the popular vote and 346 electoral votes. Furthermore, Obama was forecast to have an 83.6% chance of winning the White House and an 85.9% chance of winning the popular vote.

Efforts to call elections before they occur, using statistical models, have been conducted since at least the 1970s. Tufte (1975) and Lewis-Beck and Rice (1984) pioneered some of the first attempts at predicting elections. More recent examples of forecasts for Congress were Abramowitz (2002; 2006) as well as a host of other models presented on the Political Forecasting group's web site for the 2006 elections.[2] Most forecasting models of House and Senate elections have not made predictions at the state or district level (see Kastellec, Gelman, and Chandler 2008). As one example, none of the seven presidential forecasting models appearing in the 2004 edition of *PS* made forecasts at the state level, although such models have appeared elsewhere (Campbell 1992). However, how national factors influence election outcomes is contingent on the distribution of votes across districts or states. Imagine this simple example: a national wave gives Republicans three percentage points more votes across all House districts in comparison to the last election. Whether this will result in more seats for Republicans depends on the number of districts in which Democrats obtained less than 53% of the two-party vote last time. If there are many such districts, the Republicans will pick up many such seats. If there are few such districts, they will not. This illustrates one reason to conduct forecasts at the district or state level.

Another advantage of state-level predictions for presidential elections is that the popular-vote winner in a presidential election may also be the Electoral College loser, as occurred in 2000. Although the popular-vote winner wins the vast majority of the time, a marginal increase in predictive power can be achieved by making forecasts for each state. Furthermore, state-level factors may influence how the election as a whole plays out. Accordingly, the relative economic health of states and candidate home-state advantages were shown to have noteworthy effects in the analysis of presidential elections conducted here.

## Methodology

Separate analyses were conducted for the House, the Senate, and the presidency. The dependent variable in each analysis was the percent of the major-party vote that the Democratic candidate received in a state or district. I analyzed House and Senate elections from 1954 to 2006. I excluded elections with nonmajor party incumbents, with more than three major-party candidates or that were uncontested by one of the major parties. All states from 1948 to 2004 were used in the analysis of presidential elections, except two states that had no Democratic candidate on the ballot.[3]

A number of methodological changes to the Klarner and Buchanan House and Senate models were implemented for 2008. First, the dynamics of the time series have been more adequately assessed in light of recent work on the subject (De Boef and Keele 2008). De Boef and Keele advised using lagged independent variables to correctly capture the dynamics of time-series data. Second, the nestedness of the data were taken into account with the use of random-effects maximum-likelihood regression.[4] National waves are not adequately captured by national-level predictor variables. Prediction errors that apply across the board to one election year can best be captured with random-effects models. Third, I utilized Gary King's Amelia II program for all three analyses to impute missing data (King et al. 2001; Honaker et al. 2001). Last, both the House and Senate datasets have been pushed back to 1954 as the 2006 forecasts only went back to 1974 and 1972, respectively. More information on the technical aspects of these modeling decisions is available in an unpublished supplement available on request.

For this study, I measured the partisan disposition of states and districts in three ways. First, Lagged Vote$_{t-1}$ measured the percent of the major-party vote that Democrats obtained in a state or a district in the last election for all three models. When a major party did not

 doi:10.1017/S1049096508081213 

contest a previous election, they were given 0% of the vote. To relax the former assumption, the variable Uncontested$_{t-1}$ was included, coded 1 when the Republicans did not contest an election last time, −1 when Democrats did not, and 0 when the election was contested. Senate elections presented a challenge to model because of their staggered nature. Accordingly, a variable measuring the vote six years ago was included and labeled Lagged Vote$_{t-2}$. An interaction between whether there was an incumbent running and the vote six years ago was also included called Lagged Vote$_{t-2}$ * Incumbency Dummy. This variable was included because vote share was expected to be more highly correlated between two elections when the same candidate was running in both, in comparison to when this was not the case. The variable Incumbency Dummy was coded 1 when an incumbent ran, and 0 if otherwise. Incumbency Dummy was included for technical reasons because of the interaction. The sign or magnitude of its coefficient has no substantive importance. Next, the variable Presidential Vote Centered$_{t-1}$, representing the percent of the vote attained by the Democratic presidential candidate in the district or the state in the last election, was included in the two congressional models. It was expressed as a deviation from the yearly mean. Normalizing the vote in this manner has been standard in the study of congressional elections. (See Highton 2000.) A third measure of the partisan disposition of a state for models for both senatorial and presidential elections was the percent of legislators in the state House who were Democrats one year prior to an election. This followed Campbell (1992). This third variable measuring partisan disposition was Percent State House Democratic. It came from Klarner (2003), as well as from the Council of State Governments (various years) for years prior to 1960. This variable was centered at its yearly mean in the same manner as Presidential Vote Centered$_{t-1}$ was.

A variety of candidate characteristics were included in the models. Whether House or Senate candidates were incumbents was captured by the variable Incumbency: coded 1 when the incumbent was a Democrat, −1 when the incumbent was a Republican, with other cases coded 0.[5] An additional variable Unelected Incumbent for senators was coded 1 when an unelected incumbent was a Democrat, −1 when an unelected incumbent was a Republican, and 0 if otherwise. It was hypothesized to be negative. For analysis of House elections, a variable called Freshman Incumbent was coded 1 when a House freshman incumbent was a Democrat, −1 when a House freshmen incumbent was a Republican, and 0 if otherwise. Freshman Incumbent was also hypothesized to be negative. Prior research has also indicated that the incumbency effect has increased over time (see Kastellec, Gelman, and Chandler 2008 for the House, and Highton 2000 for the Senate). Preliminary analyses described in the unpublished supplement indicated that the value of Senate incumbency was worth 37% less in 1954 than 1984 and after. House incumbency was worth 40% less in 1954 than 1980 or after. A variable was created that represented the proportion the incumbency effect was estimated to be reduced by in earlier years, by the preliminary analysis. This was then multiplied by the variable Incumbency, which yielded the variable Incumbency Adjuster, which was included in both the House and Senate analyses. This variable was hypothesized to have the opposite sign as Incumbency and also be of similar magnitude when comparing their coefficients.

Keith Poole's DW-Nominate scores from the immediately previous session of Congress for House and Senate incumbents were included as the variable DW-Nominate to take into account the propensity for more moderate candidates to do better in elections.[6] Such scores for the first half of the current session of Congress have recently become available on Poole's web site. A focus on incumbents was necessary because of limited information on challengers' ideology. Furthermore, previous research (Erikson and Wright 2005) has found that when both candidates' ideologies were measured, incumbent ideology was much more consequential for election outcomes in comparison to the challenger's ideology. Higher scores meant candidates were more conservative, and so this variable was hypothesized to be positive. More conservative Democratic incumbents would be more moderate, and would therefore attain a higher proportion of the vote. More conservative Republicans would be more extreme, and therefore the Democratic share of the vote would again be higher.

Whether challengers have held previous office has been found to be important for election outcomes (Jacobson 2004). A variable measuring whether non-incumbent House candidates had held previous office was included, designated Previous Office Holder. It was coded 1 when the Democratic candidate held an elective position in the past, −1 when the Republican candidate did so, and 0 if otherwise. Analogous variables for House elections were measured for both former state legislators and former House members. These two variables were called Former State Legislator and Former U.S. House Member. They captured the additional boost that these higher offices had.

A scale was utilized for Senate elections to reflect the previous office holding experience of challengers, also labeled Previous Office Holder. For non-incumbent candidates, points were awarded in the following manner. Candidates who were governors and U.S. senators received scores of 5; former members of the U.S. House received scores of 3. Additionally, the proportion of the state House members represented was added to their score. Statewide officers who were not governors received scores of 2. State legislators received scores of 1. All other candidates received scores of 0. All such scores were multiplied by two-thirds if the candidate did not hold the office immediately prior to the election. For Republican candidates, the scores were multiplied by −1. These were then added to the scores of Democratic candidates.[7] Like incumbency, the impact of challengers' previous office holding experiences varied over time for both the House and Senate. It was captured by the variable Previous Office Holder Adjuster, coded as Incumbency Adjuster described above was.

Attributes of presidential candidates were also measured, following Campbell (1992). Home State Advantage was coded 1 when a Democratic presidential candidate came from the observed state, while −1 was coded for the home state of Republican candidates. All other cases were coded 0. VP Home State Advantage was coded in the same way for the home states of the vice presidential candidates.

The economic health of a state was measured in all three analyses by State Per Capita Personal Income Growth. This variable was measured as the percent change in state per capita personal income from the third quarter of the year prior to an election to the first quarter of the election year.[8] This variable captured whether voters held the incumbent party of the president responsible for their state's economic growth, not only in presidential elections, but also in congressional elections as well. These data are mean centered, as they measured the relative economic health of states compared to each other. Scores were multiplied by −1 when Republicans controlled the presidency.

National-level factors were also utilized in all three models to assess the size of the national wave flowing in favor of one party or the other. National Per Capita Income Growth was measured as the percent change in real disposable, seasonally adjusted per capita income between July of the year preceding the election to May of the election year, using monthly data. In years when Republicans controlled the presidency, this variable was multiplied by −1. Vote Intention represented the percent of

**Table 1**
**Determinants of Democratic Vote: House and Senate Elections, 1954–2006**

| Independent Variable | Model One: House | | Model Two: Senate | |
|---|---|---|---|---|
| Lagged Vote$_{t-1}$ | .500* | .011 | .049 | .032 |
| Lagged Vote$_{t-2}$ | — | — | .061* | .037 |
| Incumbency Dummy | — | — | –6.478* | 2.416 |
| Lagged Vote$_{t-2}$ * Incumbency Dummy | — | — | .137* | .044 |
| Uncontested$_{t-1}$ | –12.767* | .449 | –1.785 | 1.818 |
| Presidential Vote Centered$_{t-1}$ | .305* | .010 | .193* | .041 |
| Percent State House Democratic | — | — | .050* | .018 |
| Incumbency | 12.344* | .306 | 12.499* | .798 |
| Incumbency$_{t-1}$ | –4.114* | .361 | –1.938* | .668 |
| Incumbency Adjuster | –13.768* | .970 | –14.473* | 2.125 |
| Incumbency Adjuster$_{t-1}$ | 7.098* | .979 | — | — |
| Freshman Incumbent | –.324 | .298 | — | — |
| Freshman Incumbent$_{t-1}$ | .526* | .191 | — | — |
| Unelected Incumbent | — | — | –5.280* | 2.076 |
| DW-Nominate | 9.248* | .738 | 6.352* | 1.675 |
| DW-Nominate$_{t-1}$ | –4.262* | .843 | –5.196* | 1.660 |
| Previous Office Holder | 3.001* | .279 | 2.154* | .204 |
| Previous Office Holder$_{t-1}$ | –.917* | .287 | — | — |
| Previous Office Holder Adjuster | –2.076* | .657 | –2.726* | .587 |
| Previous Office Holder Adjuster$_{t-1}$ | 1.558* | .739 | — | — |
| Former State Legislator | .975* | .243 | — | — |
| Former U.S. House Member | 1.802* | .494 | — | — |
| State Per Capita Personal Income Growth (%) | .031 | .024 | .131 | .123 |
| National Per Capita Personal Income Growth (%) | .117* | .061 | .046 | .092 |
| Vote Intention | .055 | .052 | .133 | .088 |
| Presidential Approval | .095* | .017 | .048* | .028 |
| Presidential Approval$_{t-1}$ | –.044* | .016 | — | — |
| Midterm Penalty | –3.947* | .446 | –3.149* | .716 |
| Constant | 20.015* | 2.977 | 34.991* | 5.364 |
| $\Sigma_u$ | 1.228* | .182 | 1.576* | .373 |
| $\Sigma_e$ | 6.359* | .046 | 7.290* | .181 |
| P | .036* | .010 | .045* | .020 |
| Log Likelihood | –32643.379 | | –2876.080 | |
| N | 9975 | | 841 | |
| Number of Groups (Years) | 27 | | 27 | |

Note: The dependent variable is the percentage of the major-party vote obtained by Democrats in a district year or state year, respectively. Cell entries are the unstandardized random-effects ML regression coefficient, followed by its standard error in the third and fifth columns. Random effects are grouped by year. * = p < .05, one-tailed test.

respondents—those who said they would vote for the candidate of a major party—in Gallup surveys who also said they would vote for a Democratic candidate. One series of survey questions asked about vote intention for Congress, while another series specifically asked about presidential vote intention. The congressional vote intention data comes from surveys asked as close to June 15 of election years as possible, while the presidential vote intention question comes as close to July 4 as possible. These were the latest dates these questions were asked in 2008 before I performed this data analysis. Presidential Approval is the percent of respondents approving of the job of the president when a Democrat was in office (of those approving or disapproving), or the percent disapproving when a Republican is in office. These data are from Gallup polls, which commenced as close to June 9 of election years as possible. Midterm Penalty was included in the Senate and House models to capture the well-known tendency of the party of the president to lose seats during midterm elections. It was coded 1 during midterm election years when the Democrats controlled the presidency, −1 when the Republicans did so, and 0 if otherwise. Midterm Penalty was hypothesized to be negatively related to Democratic vote share. For the presidency, a variable was included labeled Two Term Penalty. It was coded 1 when the Democrats had inhabited the White House for two or more terms, −1 when the Republicans had done so, and 0 if otherwise. This variable was also hypothesized to be negative (Bartels and Zaller 2001).

All of the above independent variables were also lagged one period and included in initial models (unrestricted models). But lagged variables were dropped if they did not attain statistical significance, in a restricted model, consistent with the suggestions of De Boef and Keele (2008). I only report the results of the restricted models here, while those of the unrestricted models are available in the unpublished supplement.

## Findings

Space constraints necessitate that the findings of Tables 1 and 2 be summarized succinctly. Suffice it to say for the vast

**Table 2**
**Determinants of Democratic Vote: Presidential Elections, 1948–2004**

| Independent Variable | Model One | |
|---|---|---|
| Lagged Vote$_{t-1}$ | .544* | .030 |
| Percent State House Democratic | .157* | .031 |
| Percent State House Democratic$_{t-1}$ | –.078* | .029 |
| Home State Advantage | 3.937* | 1.198 |
| Home State Advantage$_{t-1}$ | –2.057* | 1.241 |
| VP Home State Advantage | 2.554* | 1.194 |
| State Per Capita Personal Income Growth (%) | .296* | .120 |
| State Per Capita Personal Income Growth$_{t-1}$ (%) | –.462* | .111 |
| National Per Capita Personal Income Growth (%) | .337* | .173 |
| Vote Intention | .121 | .121 |
| Vote Intention$_{t-1}$ | –.222* | .081 |
| Presidential Approval | .176* | .081 |
| Two-Term Penalty | –4.524* | 1.664 |
| Constant | 18.236* | 5.424 |
| $\Sigma_u$ | 2.604* | .540 |
| $\Sigma_e$ | 6.272* | .167 |
| P | .147* | .053 |
| Log Likelihood | –2468.061 | |
| N | 753 | |
| Number of Groups (Years) | 15 | |

Note: The dependent variable is the percentage of the major-party vote obtained by Democrats in a state year. Cell entries are the unstandardized random-effects ML regression coefficient, followed by its standard error in the last column. Random effects are grouped by year. * = $p < .05$, one-tailed test.

majority of district- or state-level variables that they behaved as expected and achieved statistical significance at conventional levels ($p < .05$). Variables that measured the partisan disposition of the district or state always achieved statistical significance, with the single exception of Lagged Vote$_{t-1}$ for the Senate. However, the novel modeling of lagged Senate voting does require a more detailed comment. As just stated, the coefficient for Lagged Vote$_{t-1}$ did not achieve statistical significance in the Senate model, although Lagged Vote$_{t-2}$ did so. Furthermore, when an incumbent senator was on the ballot, the impact of the vote—lagged six years—was three times greater than when an incumbent did not run. This was implied by the coefficient associated with Lagged Vote$_{t-2}$ * Incumbency Dummy, which is 0.137 ($p < .05$). This statistic indicated that previous vote share had a 0.198% impact when an incumbent was running (i.e., 0.061 + 0.137) in comparison to 0.061% when an incumbent was not running.

For the House and Senate, I found that incumbents, more moderate incumbents, and previous office holders received more of the vote than candidates who did not share those characteristics ($p < .05$). The variables in Table 1 that took into account the variation in the incumbency and previous office-holding effects over time also performed as expected. They had negative signs and magnitudes that approximated that of the variables they adjusted. For example, Incumbency in model one of Table 1 for the House had a coefficient of 12.3, while Incumbency Adjuster had a coefficient of −13.8. Consistent with prior research, presidential candidates received about 4% more of the vote in their home states, as the coefficient for Home State Advantage in Table 2 indicates, which was statistically significant. Vice presidential candidates were estimated to bring a 2.5% boost to their parties in their home states (also statistically significant).

The attainment of statistical significance for most of the district-level lagged variables for the House was especially noteworthy. Models that have not taken the dynamic nature of this phenomenon into account have been underspecified. Lagged variables in the Senate model of Table 1 were not as apt to attain statistical significance (although Incumbency$_{t-1}$ does). This was partly a function of the small coefficient associated with Lagged Vote$_{t-1}$ in the Senate model. The intuition behind why the lagged independent variables would attain statistical significance and have coefficients of the opposite (usually negative) sign can be illustrated with lagged incumbency. Since a Democratic incumbent obtained more votes for Democrats in a previous election, lagged Democratic vote was therefore a weaker sign of a strongly Democratic district. The lagged incumbency variable therefore cleaned districts of this previous boost. Imagine two districts with the same vote percentage in the last election. In one of those districts, a Republican incumbent was running. In the next election, we would have expected the Republicans to do better in the district (all else being equal) where the incumbent was not running last time because it was a more strongly Republican district. Consistent with this explanation, when the lagged dependent variable was dropped from the House analysis, the variables Incumbency$_{t-1}$ and Previous Office Holder$_{t-1}$ switched their signs from negative to positive, and were statistically significant. Kastellec, Gelman, and Chandler (2008) cleaned previous vote share of incumbency deterministically by subtracting out a fixed percentage from the lagged vote, but here I favored a strategy that estimated the effect empirically.

The evidence presented in Table 1 indicated that the health of a state's economy did not have an impact on votes for the House or Senate, as the coefficient associated with State Per Capita Personal Income Growth did not attain statistical significance in either model. However, this variable was statistically significant in the model for the presidency displayed in Table 2 as was its lagged component. The better a state's economy relative to the rest of the nation, the more voters seemed to reward the party of the president. Not many studies have examined the impact of state-level economies on election outcomes, so these findings are noteworthy.

The national-level variables did not always attain statistical significance, and generally were more important for presidential elections. The only variable that attained statistical significance in all three models was Presidential Approval. Not surprisingly, it had the largest impact in presidential elections, but also had about twice the impact in House elections than in Senate elections. National per Capita Personal Income Growth attained statistical significance for both House and presidential models, but not for the Senate model. Vote Intention failed to attain statistical significance in either the House model or Senate model, and only its lagged component attained statistical significance in the presidential model. These results are displayed in Table 2. The national wave was estimated to be 4% more against the party of the president during midterm elections in the House model of Table 1, as the coefficient associated with Midterm Penalty indicated. It attained statistical significance at conventional levels. This variable was also statistically significant for the Senate, with a magnitude of about 3%. Last, consistent with previous research, parties that were in the White House for eight years or

more received about a 4.5% penalty, as evidenced by the coefficient for Two-Term Penalty in Table 2 ($p < .05$).

Another noteworthy result was the performance of national-level variables when random-effects regression was used, compared to when ordinary least squares (OLS) regression was used. For the House and presidency models, OLS yielded statistically significant findings for all national-level variables and all their lagged components. However, the Senate results were fairly similar for the two methodologies. It has been common for higher level variables to attain statistical significance when OLS is used. And so the comparison between OLS and random-effects regression underscores the utility of the latter method in such tests.

Finally, the parameter $\Sigma_u$ indicated the standard deviation of the year-specific residuals across repeated samples. This parameter described the uncertainty that existed about the size of a national wave, over and above the national level factors that forecast such a wave. For the House, this parameter was 1.2. This means that about two-thirds of the time the national wave will be within 1.2% of the model's estimates, and about 95% of the time it will be within 2.4% of those estimates. This parameter was greater for the Senate, at 1.6%, and greater still for presidential elections, at 2.6%. Errors in prediction at the district or state level were greater in magnitude, with the parameter $\Sigma_e$ in Tables 1 and 2 representing that quantity. This parameter was 6.4 for the House, 7.3 for the Senate, and 6.3 for the presidency. The uncertainty about any particular national wave was substantial, but not so large as to make forecasts meaningless. Uncertainty about district-level results were larger in comparison to national-level uncertainty, but were not quite as damaging for national-level forecasts. This was because such uncertainty "cancels itself out" to an extent. However, these results indicated that district- and state-level forecasts should be made with caution.

## Forecasts

Values of all independent variables as well as their estimated impacts were utilized to yield predicted values for each state and district for the House, Senate, and presidential elections. I then conducted 2,000 simulations to compute the probabilities of different national outcomes for the three types of elections in the following manner. The predicted values were added to the following two quantities 2,000 times, to capture the uncertainty the models had in their predictions. One was a normally distributed random variable that varied across districts (or states) with mean 0 and standard deviation $\Sigma_e$, taken from Table 1 or 2 as appropriate. The second was a random variable that varied across simulations but not across districts or states with mean 0 and standard deviation $\Sigma_u$. The resulting variables were then recoded to declare a winner in each state or district and these were totaled for an overall national result for each of the 2,000 simulated national elections. For the presidency, I multiplied the variables representing winners by their electoral votes before addition. One problem with the presidential forecast was that the vice presidential running mates had not been chosen before the July 28, 2008, forecast, meaning that information could not be utilized in the predictions. The 81 House districts that were uncontested by the Republicans were assigned to the Democrats, and the 20 districts uncontested by the Democrats were assigned to the Republicans. The state of Arkansas was assigned to the Democrats for the Senate, because no Republican appeared on the ballot there.

National-level forces had a mixed impact on the upcoming election. First, percent growth in real personal disposable income per capita was surprisingly robust between July of 2007 and May of 2008, at 8.72%, the third highest growth rate during a presidential election year in the post-World War II period. This high level of growth is clearly in the Republicans' advantage. The high score was caused by the tax rebates that arrived in May. An alternative measure can be constructed that assumes that the boost in income from these checks was spread across the prior months of 2008, because the high score is an artifact of the one-time mailing. This adjustment would bring income growth down to 4.8%. For the "official" forecast here, the unaltered measure of personal income growth was favored so as not to arbitrarily alter the input variables. However, the substitution of the altered growth rate does make a substantial difference in the forecast for the presidential election (but not the House or Senate elections), and an alternative, "unofficial" forecast for the presidency is reported below. In contrast to income growth, presidential approval was low for the Republicans, at 31.9% of those approving or disapproving. Only in 1964 was presidential approval so favorable for the Democrats in a presidential election, and only in 1962, 1964, and 1974 was it so favorable in House and Senate elections. As noted above, presidential approval was found to influence House, Senate, and presidential elections. Last, the two-term penalty was operative for the presidential election, which was estimated to reduce Republican vote share by 4.5%.

The House currently sees 236 Democrats and 199 Republicans going into the election, meaning that 54.3% of the House is currently Democratic. The median forecast from the simulations was that the Democrats would win 247 seats in the 2008 election (56.8%), an 11-seat gain (mean forecast of 247.7). The 95% confidence interval for the number of seats the Democrats will have after the election was between 233 and 266. The 67% interval was between 240 and 255 seats. There was also a 93.5% chance that the Democrats would pick up at least one seat. The 95% confidence interval reported in Klarner and Buchanan (2006b) for 2006 was 18-seats wide, in contrast to the 34-seat confidence interval for 2008. Again, the necessity of models allowing for a year-level random component was illustrated by this contrast. Last, the simulations predict a 0% probability of the Democrats losing the House.

Because there are two Senate seats up that were not regularly scheduled to have elections, there are 35 Senate seats in play in 2008. The Republicans held 23 of these seats before the election, making them vulnerable to Democratic gains. The median forecast for the simulations was that the Democrats will win 15 seats (mean = 15.4), meaning a three-seat gain. The 95% confidence interval for the forecast was between 12 and 19 seats (between 13 and 17 for the two-thirds interval), an interval of eight seats. There was also a 92.5% chance that the Democrats would pick up at least one seat. Klarner and Buchanan (2006a) also reported an eight-seat, 95% confidence interval, so OLS does not always yield overconfident standard error of the estimates. Of the seats not up, Democrats hold 37, not including Bernie Sanders or Joe Lieberman who caucused with the Democrats. Together with this information, the simulations indicate that there is a 2.4% chance that the Democrats will lose control of the Senate, and a mere 0.3% chance that they would obtain a filibuster-proof majority of 60. Both figures assume Sander's support, but not Lieberman's. The forecast predicted they would be most likely to have 52 total seats after the election.

The simulations for presidential elections indicated that the Democrats had an 83.6% chance at victory in November, with a 0.5% probability of an Electoral College tie. The median prediction was that the Democrats would obtain 346 electoral votes (mean = 340.6). The 95% confidence interval for the forecast was fairly wide: from the Democrats obtaining 194 electoral votes to 467 electoral votes (the 67% confidence is 269 to 410). If the altered measure of per capita income growth were inserted for 2008, the probability of a Democratic victory would rise to

90.5%. If the assumption were made that the ratio of major-party votes between the 50 states will be the same in 2008 as it was in 2004 (an assumption, it should be noted, that national-level models have also implicitly made), the model predicted that Obama will receive 53.0% of the popular vote (mean = 53.1), with a 95% confidence interval of 47.3 to 58.9% (50.3 to 56.0% for the two-thirds interval). Furthermore, there was an 85.9% chance that Obama would win the popular vote. The simulations also allowed the computation of the probability of an Electoral College winner that differs from the popular-vote winner. In the simulations, the Democrats won the popular vote 5.4% of the time while having lost the Electoral College vote, while the Republicans won the popular vote but lost the Electoral College vote 2.7% of the time. All told, there was an 8.1% chance of the popular vote winner losing the upcoming election. This was a small but significant amount, which again implied that examining state-level outcomes in forecasting models is worthwhile.

## Conclusion

Prospects for the Democrats appeared well in the forecasting models reported in this article, but not overwhelmingly so. Barack Obama is predicted to get 53% of the popular vote and have an 84% chance of winning the presidency: no landslide was forecast. Democrats were predicted to pick up a small but significant 11 seats in the House. This small level of seat change is consistent with the overall reduction in the number of seats that are likely to change hands over time, documented elsewhere (Campbell 2003). Furthermore, the Democrats were forecast to pick up three seats in the Senate. Both congressional models predicted that it was very unlikely that the Democrats would lose control of the House or Senate, but also predicted their gains would be minimal.

One problem with making election forecasts on the basis of past elections is that considerations that are unique to the election at hand are impossible to model. An obvious unique factor in the upcoming election is the race of Barack Obama, which may have a substantial impact on the outcome of the 2008 presidential election because of some voters' prejudicial racial attitudes. Swing states that possess enough blacks to make whites feel threatened, such as Florida, Missouri, and Ohio, may well see Obama receive a smaller portion of the vote than I forecasted. This may cause him to lose these crucial swing states, and as a result, the election. Because the presidential election is forecast to be somewhat close in the analysis presented here, racial considerations during the election could result in a Democratic defeat.

## Notes

1. I would like to thank Dan Kiehl, Marcel Oliviera, and Richard Stangle for data collection and advice; Gary Jacobson and Jim Campbell for sharing data; and Stan Buchanan for editing the manuscript. I would also like to thank Indiana State University, the Office of the Provost and Vice President of Academic Affairs, and the Lilly Endowment, Inc., for financial support for this project.

2. www.forecastingprinciples.com/Political/election_2006.html.

3. All district and candidate level data for House elections were generously provided by Gary Jacobson. Election returns for the U.S. Senate and presidential elections are from *Congressional Quarterly's* "Voting and Elections Collection" accessed through the Grinnell College library web site at www.lib.grinnell.edu/db.php.

4. This was done with STATA 10's xtreg command, with the i(year), mle subcommand.

5. The inclusion of the variable Incumbent Dummy in the same model as this variable, for the Senate, may strike some as potentially problematic. Evidence that it was not problematic is the fact that Incumbency performed almost identically whether Incumbent Dummy was included in the model or not.

6. I obtained DW-Nominate scores from Keith Poole's web site at www.voteview.com/DWNL.

7. I obtained data on whether candidates were former governors from Moore, Preimesberger, and Tarr (2001). Information on former U.S. Senate and House incumbency was from http://bioguide.congress.gov/biosearch/biosearch.asp. Information on statewide and state legislative experience was from Council of State Governments (various years).

8. These data are from www.bea.gov/regional/index.htm#state and www.bea.gov/bea/regional/spi/default.cfm?satable=summary.

## References

Abramowitz, Alan I. 2002. "Who Will Win in November: Using the Generic Vote Question to Forecast the Outcome of the 2002 Midterm Election." *Symposium on Mid-Term Elections*. APSA Elections, Public Opinion, and Voting Behavior Section. www.apsanet.org/~elections/archives.html.

———. 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 and Politics* 39 (October): 863–66.

Bartels, Larry M., and John Zaller. 2001. "Presidential Vote Models: A Recount." *PS: Political Science and Politics* 34 (March): 9–20.

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

———. 2003. "The 2002 Midterm Election: A Typical or an Atypical Midterm?" *PS: Political Science and Politics* 36 (April): 203–7.

Council of State Governments. Various Years. *The Book of the States, Supplement I: State Elected Officials and the Legislatures*. Chicago: Council of State Governments.

De Boef, Suzanna, and Luke Keele. 2008. "Taking Time Seriously." *American Journal of Political Science* 52 (January): 184–200.

Erikson, Robert S., and Gerald Wright. 2005. "Voters, Candidates, and Issues in Congressional Elections." In *Congress Reconsidered*, ed. Lawrence C. Dodd and Bruce L. Oppenheimer. Washington, D.C.: CQ Press.

Highton, Benjamin. 2000. "Senate Elections in the United States: 1920–94." *British Journal of Political Science* 30 (July): 483–506.

Honaker, James, Anne Joseph, Gary King, Kenneth Scheve, and Naunihal Singh. 2001. *Amelia: A Program for Missing Data*. (Gauss Version). Cambridge, MA: Harvard University.

Jacobson, Gary C. 2004. *The Politics of Congressional Elections*. 6th ed. New York: Longman.

Kastellec, Jonathan P., Andrew Gelman, and Jamie P. Chandler. 2008. "Predicting and Dissecting the Seats-Votes Curve in the 2006 U.S. House Election." *PS: Political Science and Politics* 37 (January): 139–45.

King, Gary, James Honaker, Anne Joseph, and Kenneth Scheve. 2001. "Analyzing Incomplete Political Science Data: An Alternative Algorithm for Multiple Imputation." *American Political Science Review* 95 (March): 49–69.

Klarner, Carl. 2003. "Measurement of Partisan Balance of State Government." *State Politics and Policy Quarterly* 3 (Fall): 309–19.

Klarner, Carl E., and Stan Buchanan. 2006a. "Forecasting the 2006 Elections for the United States Senate." *PS: Political Science and Politics* 39 (October): 849–55.

———. 2006b. "Forecasting the 2006 Elections for the U.S. House of Representatives." *PS: Political Science and Politics* 39 (October): 857–61.

Lewis-Beck, Michael S., and Tom W. Rice. 1984. "Forecasting U.S. House Elections." *Legislative Studies Quarterly* 9 (August): 475–86.

Moore, John L., Jon P. Preimesberger, and David R. Tarr. 2001. *Congressional Quarterly's Guide to U.S. Elections*. 4th edition. Washington, D.C.: CQ Press.

Tufte, Edward R. 1975. "Determinants of the Outcomes of Midterm Congressional Elections." *American Political Science Review* 69 (September): 812–26.