# Predicting the 2006 Elections for the U.S. House of Representatives

Carl E. Klarner
Indiana State University
Room 301 Holmstedt Hall
Terre Haute, IN 47809
812-514-9060
cklarner@isugw.indstate.edu

Stan Buchanan
Indiana State University
Room 306 Holmstedt Hall
Terre Haute, IN 47809
812-237-2437
sbuchanan1@isugw.indstate.edu

Abstract

Utilizing information on House elections from 1974-2004, predictions for the outcome of the 2006 House elections are made. A novel approach to predicting Congressional elections is utilized, using candidate, district, and national level variables in a pooled design. The probability that the Democratic candidate will win is computed for each 2006 House race, and the 30 most competitive races are identified. National partisan tides favor the Democrats this year, in the form of pro-Democratic vote intentions, low presidential approval, and moderately low growth in real disposable income. These factors result in an estimated 94.9 percent chance that the Democrats will take control of the U.S. House of Representatives, and are predicted to have 224 seats after the election, a 22 seat gain. This prediction of substantial Democratic gains run counter to many claims made in the popular press in late April.

# Predicting the 2006 Elections for the U.S. House of Representatives

Carl E. Klarner and Stan Buchanan

Indiana State University

Methods of predicting Congressional election outcomes fall into two categories. In the first method, experts use district or state level information to try to call the election. The second method uses a more aggregate approach. In this approach, past elections are examined in order to compare the number of seats that change party hands with various measures of a national partisan tide. Forecasts are made for the upcoming election on the basis of these past patterns. The strength of the district level approach is that it recognizes the importance of district level factors. Its weakness is that it may not accurately assess the size of a national partisan tide. The strength of the aggregate approach is that it is better at assessing such a tide. Its weakness is that it does not assess how such a national partisan tide may interact with district level factors. Combining both approaches by using a panel model, that is, a model that combines many different time series into one pooled dataset overcomes both of these problems. National and district conditions, as well as how they influence election outcomes in combination, can be examined. The rest of this article explains this modeling strategy and its predictions for the 2006 midterm House elections.

## Methodology

The general strategy utilized in this paper is to assess relationships between independent variables and election outcomes in past elections and then use the values of these independent variables in 2006 to predict the election at hand. The dependent variable of the analysis is the percent of the two-party vote that goes to Democrats in

contested U.S. House of Representatives elections between 1974 and 2004. As mentioned above, a pooled design is utilized, as others have done for House (McGhee 2004) and Senate (Highton 2000) elections. Cases from 1988 and 1992 had to be excluded from the analysis for reasons explained below. Three general types of factors are used: district partisan composition, candidate attributes, and national partisan tides. All data used in this model save the national partisan tide variables are from data generously provided by Gary Jacobson. Variable names are set off by parentheses below.

The partisan composition of districts is measured in two ways. The first is the two-party vote for Democrats from the most recent regularly scheduled House election in the district ("Past House Vote"). When no Democratic (Republican) candidate contested the last election, a "0" ("100") was put into this variable. "Last Election Uncontested" was included to relax the assumption that an election with no Democratic candidate is equivalent to a "0" and vice versa. It is coded "-1" when the last election was not contested by a Democrat, "1" when the election was not contested by a Republican, and "0" otherwise.

The second measure of the partisan composition of districts is the result from the most recent presidential election in the district, but with the percent of the Democratic two-party vote in the nation as a whole subtracted from it ("Past Presidential Vote"). This subtraction is done so that past national partisan tides favoring a certain party in a presidential election are subtracted out of presidential vote returns. (See Highton 2000 for a full explanation.) The inclusion of this variable necessitates the exclusion of 1992 from the analysis.

Various attributes of House candidates are included in the model. Whether a candidate is an incumbent or not is measured ("Incumbent"). Whether a candidate has held a past elective office is measured by the two following variables. "Quality Candidate, Closed" measures such candidates when they face an incumbent. "Quality Candidate, Open" measures such candidates in open seats. Whether a candidate has held a seat in the House of Representatives is also considered, also taking into account whether they are now running against an incumbent or not ("Past House Member, Closed," and "Past House Member, Open"). These variables are coded "1" when only the Democratic candidate has these attributes, "-1" when only the Republican candidate does, and "0" for all other instances.

Four aspects of national partisan tides are measured. The first is the percent of respondents who express an intention to vote for a Democratic House candidate in the upcoming election, recorded by the Gallup poll conducted closest to March 10th of even numbered years ("Democratic Vote Intention").[1] Unfortunately, this question was not asked by Gallup in 1988, making it necessary to exclude that year from analysis.

The second national partisan tide factor to be examined is presidential approval ("Presidential Approval"),[2] (Jacobson 2004). Again, this is presidential approval expressed in the Gallup poll asked closest to March 10th of even numbered years. March 10th is used to make it compatible with the vote intention question. When a Democrat is president, it is the percent of respondents who approve of the president (of respondents either approving or disapproving of the president). When a Republican is president, it is the percent of respondents disapproving of the president.

The third national partisan tide variable is the relative advantage to each party the performance of the economy brings. The strength of the economy is measured as the percent change that occurs in per capita real disposable income between the February of the year before the election year, and the February of the election year ("Change in Real Disposable Income").[3] To reflect the fact that Democratic candidates are expected to be hurt by a good economy when a Republican is in the White House, this variable is multiplied by "-1" when a Republican is president, as Jacbobson (2004, 167) does.

The fourth national partisan tide variable captures the tendency of the party of the President to lose votes in a midterm election ("Midterm Penalty") (Jacobson 2004). This variable is coded "-1" for midterm elections in which a Democrat is in the White House, "1" for midterm elections in which a Republican is in the White House, and a "0" for presidential election years.

**Predictions**

The variables in the model behave much as expected. How they relate to the Democratic percent of the vote is displayed in Table 1. All variables have the expected signs, and all are statistically significant at conventional levels ($p<.05$) or higher with two minor exceptions identified in the table. From the coefficients generated from these past elections, predicted values are computed for 2006.

In order to make predictions about the 2006 House elections, the attributes of candidates running in 2006 must be determined. Lists of candidates and their incumbency status or prior experience were obtained from Ron Gunzburger's "Politics1.com" website.[4] Some ambiguity about which candidates would run was caused by the fact that many states have not had their primaries yet. To summarize very

concisely, three sets of codings for challenger attributes were coded: a set coded as favorably for Democrats as possible, a set coded as favorably for Republicans as possible, and a set with what we viewed as the most probable candidate characteristics of the primary winner. We judged that primary candidates with prior office holding experience were the most likely to win their primaries. Comparing the results from the first two coding decisions provides an absolute range between which the probability of Democratic success lies. Surprisingly, this range is not terribly large. The average probability of a Democratic win is 50.50 percent for the "pro-Republican" codings, while it is 51.96 percent for the "pro-Democratic" codings. We believe there is virtually no chance anything approaching either of these extreme scenarios will occur. We argue instead that this range constitutes the absolute bounds this type of uncertainty causes. Not surprisingly, our official prediction is based on the third set of candidate characteristics we identified: the most probable characteristics.

The predictions of the model for 2006 are discussed below. Unless stated otherwise, predictions are made on the basis of picking the "best" candidate for both parties whenever uncertainty about what type of candidate will win a primary exists (the third method identified above). The 30 most competitive races are listed in Table 2. Many of these races have also been highlighted by the media as "races to watch." For example, twelve of the races listed in Table 2 were listed among the top 20 races to watch by the Washington Post.[5] Fifteen of the races listed in Table 2 were listed among the top 32 most competitive races by Congressional Quarterly.[6]

Simulations of 2000 different "elections" were conducted to generate a distribution of possible outcomes for the election as a whole. This was done by

generating normally distributed random variables, multiplying them by the standard error of the estimate for the model as a whole, and adding them to the estimated percent of the vote that Democrats would get in any particular district. Districts in each simulation were then classified as either won by the Democrats, or won by the Republicans. All districts so coded were then added together for an overall prediction for how many seats the Democrats would have after the election.

Figure 1 displays the probability that the Democrats will have a given number of seats after the 2006 election. The median estimate is that Democrats will end up with 224, six seats more than needed for control of the House. This constitutes a gain of 22 seats for the Democrats, from their current 202 seats. Overall, the model predicts there is a 94.9 percent chance the Democrats will take control of the House. Ignoring the most extreme 2.5 percent of scenarios at the low and high end of estimates indicates that the Democrats will control between 216 and 233 seats after the election.

When values for 2006 non-incumbents are coded to be as favorable for Republicans as possible, slightly different outcomes are predicted. The median estimate is that the Democrats will take 220 seats, and they are only estimated to have a 69.2 percent chance of taking control of the House. Ninety-five percent of the time the Democrats will obtain between 212 and 228 seats, under these (extremely unrealistic) assumptions. If results are coded as favorably as possible for Democrats, there is a 97.4 percent chance they will take control of the House, the median estimate of seats is 226, and there is a ninety-five percent chance they will get between 217 and 234 seats. These findings indicate that uncertainty about who will win primaries is not problematic for the predictions this model makes.

## Conclusion

A novel method of predicting House elections is showcased in this article. Additionally, this study builds on the foundation of election studies to date in order to identify the most important factors for predicting the outcome of Congressional elections. Making use of variables that have long been studied as correlates of success in Congressional elections, it is predicted that the Democrats will pick up 22 seats in the upcoming House of Representatives election, leaving them with 224. It also predicts that there is a 94.9 percent chance that the Democrats will take control of the House.

This election forecast contrasts with current seat-by-seat analyses of the 2006 House election. *Congressional Quarterly* has called 224 seats for the Republicans. Even if the Democrats won all of the nine seats *CQ* labeled "no clear favorite" this would leave them with only 211 seats.[7] Examinations of a bivariate aggregate seat swings examining Democratic vote intention yield predictions that Democrats will have 250 seats when the election is over (analysis not shown). The results of the model showcased here lie between these two estimates, although our estimates are closer to *CQ*'s than to the simple national partisan tide model, which is a testament to the insulation of most House districts from national partisan tides (McGhee 2004; Campbell 2003). Some evidence about which of the three methods (district by district, aggregate, or a "combined approach" as is used here) makes the best forecasts will be obtained on election night.

**Bibliography**

Abramowitz, Alan. 2002. "Using the Generic Vote Question to Forecast the Outcome of the 2002 Midterm Election." Elections, Public Opinion and Voting Behavior section of the American Political Science Association Symposium on Mid-Term Elections.

Abramowitz, Alan I., and Jeffrey Segal. 1986. "Determinants of the Outcomes of U.S. Senate Elections." *Journal of Politics* 48(2):433-39.

Abramowitz, Alan I., and Jeffrey A. Segal. 1992. *Senate Elections*. Ann Arbor: The University of Michigan Press.

Abramson, Paul R., John H. Aldrich, and David W. Rohde. 2006. *Change and Continuity in the 2004 Elections*. Washington, DC: CQ Press.

Campbell, James E. 2003. "The 2002 Midterm Election: A Typical or an Atypical Midterm?" *PS: Political Science and Politics* 36(April):203-07.

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

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

Jacobson, Gary C. 2005. "The Congress: The Structural Basis of Republican Success." In *The Elections of 2004*, ed. Michael Nelson. Washington, DC: CQ Press.

McGhee, Eric. 2004. "National Tides and Local Results in U.S. House Elections." Presented at the Annual Meeting of the American Political Science Association, Chicago.

Endnotes

[1] These data are from www.gallup.com and were accessed in late March of 2006.
[2] These data are from www.gallup.com and were accessed in late March of 2006.
[3] These data come from the Bureau of Economic Analysis Web site, in the Department of Commerce. Table 2.6 at http://www.bea.gov/bea/dn/nipaweb/selecttable.asp?selected=n, and were accessed April 9th, 2006.
[4] This Web site is located at http://www.politics1.com/states.htm and was accessed April 22nd, 2006.
[5] See http://projects.washingtonpost.com/elections/keyraces/map/, accessed April 27, 2006.
[6] http://www.cqpolitics.com/risk_rating_house.html, accessed April 28, 2006.
[7] See http://www.cqpolitics.com/, accessed April 28, 2006.

Table 1: Determinants of Democratic Share of the Two-Party Vote in House Elections: 1974-2004

| Independent Variables | Unstandardized Regression Coefficient with Standard Error in Parentheses |
|---|---|
| Past House Vote | .405 (.011)*** |
| Last Election Uncontested | -9.515 (.525)*** |
| Past Presidential Vote | .314 (.011)*** |
| Incumbent | 8.403 (.195)*** |
| Quality Candidate, Closed Seat | 3.181 (.264)*** |
| Quality Candidate, Open Seat | 2.890 (.433)*** |
| Past House Member, Closed Seat | 1.368 (.941)* |
| Past House Member, Open Seat | .540 (2.312) |
| Democratic Vote Intention | .033 (.015)** |
| Presidential Approval | .103 (.008)*** |
| Change in Real Disposable Income | .139 (.040)*** |
| Midterm Penalty | -3.644 (.166)*** |
| Constant | 23.147 (.942) |
| R-Squared | .856 |
| Adjusted R-Squared | .856 |
| Standard Error of the Estimate | 6.815 |
| N | 5899 |

Note: The cell entries are, respectively, the unstandardized regression coefficient and the standard error in parentheses. All tests of statistical significance are one-tailed, save for the constant.
* = P<.10, ** = P<.05, *** = P<.01

Table 2: 2006 Predicted Outcomes of the 30 Most Competitive House Races

| District by State and Number | Predicted Democratic Percent of Two-Party Vote | Predicted Percent Chance of Democratic Victory |
|---|---|---|
| CT 2 | 50.3 | 51.9 |
| AZ 8 | 49.6 | 47.7 |
| CT 4 | 50.5 | 52.7 |
| IL 6 | 49.4 | 46.4 |
| GA 3 | 50.7 | 54.4 |
| NM 1 | 49.2 | 45.3 |
| IN 9 | 48.8 | 43.0 |
| FL 13 | 48.7 | 42.5 |
| PA 8 | 48.5 | 41.6 |
| MN 6 | 51.8 | 60.3 |
| CA 50 | 52.1 | 62.1 |
| PA 6 | 47.4 | 35.2 |
| NH 2 | 47.1 | 33.3 |
| WA 8 | 46.7 | 31.6 |
| OH 12 | 46.7 | 31.5 |
| NY 24 | 46.7 | 31.5 |
| PA 15 | 46.5 | 30.6 |
| CO 4 | 46.5 | 30.4 |
| NJ 7 | 46.4 | 29.7 |
| OH 1 | 46.3 | 29.4 |
| OH 15 | 46.3 | 29.4 |
| CT 5 | 46.1 | 28.2 |
| WI 8 | 45.8 | 27.0 |
| IA 1 | 54.3 | 73.5 |
| MN 2 | 45.7 | 26.3 |
| FL 22 | 45.6 | 25.8 |
| TX 22 | 45.3 | 24.7 |
| IA 2 | 45.0 | 23.3 |
| IN 8 | 44.9 | 22.7 |
| WV 2 | 44.9 | 22.5 |

**Figure 1: Chance for Number of Democratic Seats, 2006 Election**