FORECASTING THE 2006 DEMOCRATIC PARTY TAKEOVER OF THE U.S. HOUSE OF REPRESENTATIVES

by Carl E. Klarner and Stan Buchanan

PREVIEW

Carl Klarner and Stan Buchanan (K&B) have applied an innovative approach to predicting congressional elections by pooling candidate, district-level, and national variables. National tides favored the Democrats in 2006, in the form of pro-Democratic voting intentions, low presidential approval ratings, and moderately slow growth in real disposable income. More than 6 months prior to the election, the K&B model predicted a 22-seat gain for the Democrats with a 95% chance that the party would take control of the House. The final tally gave the Democrats a gain of 30 seats, far above the 15 seats required for the party to recapture the House.

KEY POINTS

- Congressional election forecasts should consider not only national-level variables such as the president's popularity but district-level voting patterns as well. There is a synergy in the combination of national and district information.
- Pooling of election results across the country with time series of district election results permits analysis of thousands of races with a single model.
- The model was fairly accurate in its forecasts, made six months before the election. It predicted a 22-seat gain for the Democrats, while they actually picked up 30 seats. Its ability to forecast the outcomes of close races was less impressive, correctly predicting 61% of the 59 races that were won by less than 10 percentage points.

INTRODUCTION: THE TWIN BLADES OF THE SCISSORS

Methods of predicting results of congressional elections fall into two categories. In the first category, analysts use district-level information to predict election winners. In the second category, past elections are examined to relate the number of seats that change party hands to various measures of a national partisan tide. While the first approach recognizes the importance of district-level factors, it fails to fully assess the national partisan tide. The aggregate approach assesses the national mood but does not analyze how the national mood interacts with district-level factors.

To capture both viewpoints, we combined national and district conditions, as McGhee (2004) had done for House elections, and Highton (2000) had done for the Senate. Our model pools many time series – one for each district – into one dataset. Our pooled dataset covered 5899 elections over the period 1974-2004.

Carl Klarner is an Assistant Professor of Political Science at Indiana State University. He holds a PhD from Texas A&M University. His research interests include campaigns and elections, state politics, and welfare policy, especially as these three subjects pertain to political inequality and reform.

Stan Buchanan teaches graduate courses in public administration in the MPA and public administration certificate programs in the Department of Political Science, Indiana State University. He also serves as the Director of the department's program in public administration. He holds the PhD in political science from The University of Tennessee.

THE VARIABLES IN THE MODEL

The strategy behind the model was (1) to assess relationships between *past* election outcomes and the influencing district and national variables and (2) to use the *current* values of these influencing variables to predict the 2006 House elections. The relationships in past elections were estimated through multiple regression analysis. Since many of the variables in our model were proposed in previous election studies, we consider the model to be an extension of ideas gleaned from prior research.

The dependent variable was the percentage of the two-party vote for the Democrat in each of the *contested* House elections between 1974 and 2004. The elections in 1988 and 1992 had to be excluded from the analysis, for reasons explained below. All data used in this model (except the national partisan tide variables) were generously provided by Gary Jacobson (2004).

Our influencing factors – the independent variables – were of three types. They measured voter preferences within districts, candidate attributes, and national partisan tides. Below we explain how we measured each variable.

Measures of District Voting Preferences

Past House Vote: This was the two-party vote for the Democrat from the most recent, regularly scheduled House election in the district. If an election was uncontested, the Past House Vote was set equal to zero if there was no Democrat in the race. It was set equal to 100 if there was no Republican in the race.

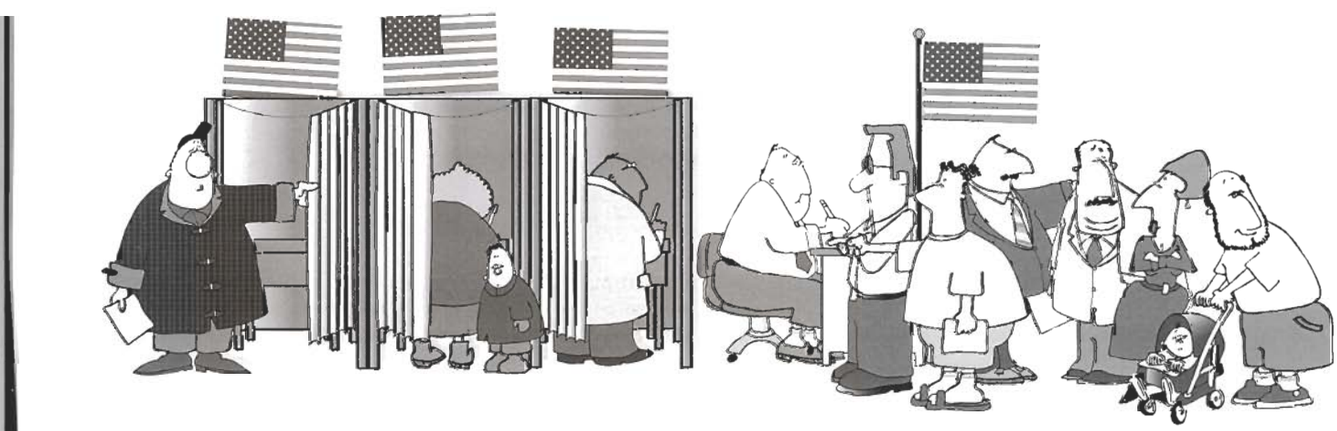
Past Presidential Vote: This variable was the percentage of the district vote for the Democratic presidential candidate minus the percentage of the national vote for the Democratic candidate. We subtracted the national votes so that past national partisan tides favoring a certain party in the last presidential election were eliminated from presidential vote returns. This variable was meant to capture only district partisanship. Subtracting the percentage of the national vote in a given election achieved this by normalizing the vote. The usage of this variable required the exclusion of the 1992 election results from the analysis because redistricting prevented measurement of the 1988 election's returns.

Last Election Uncontested: This variable was designed to relax the assumptions that (1) an election with no Democratic candidate was equivalent to a score of 0 for the Past House Vote and (2) that an election with no Republican candidate was equivalent to a score of 100 for Past House Vote. This variable was coded -1 when the last House election was not contested by a Democrat. It was coded as 1 when the election was not contested by a Republican, and it was coded as zero if otherwise.

Candidate Attributes

Incumbent: Incumbent was equal to 1 if the Democratic candidate was the incumbent, -1 if the Republican candidate was an incumbent, and zero if no incumbent ran in the district.

Quality Candidate: This variable represented a candidate who previously held elective office. We allowed the benefit of past elective experience to vary depending on whether or not the candidate faced an incumbent. If there was an incumbent opponent for such a candidate, we called it a race for a Closed Seat. If there was no incumbent, we called it a race for an Open Seat.



Quality Candidate, Closed: This variable was coded 1 if the Democratic candidate held prior elective office and faced an incumbent opponent. It was coded -1 if the Republican candidate held an elective office and faced an incumbent opponent. It was equal to zero in all other instances.

Quality Candidate, Open: This variable was equal to 1 if the Democratic candidate held prior elective office but did not face an incumbent. It was coded as -1 for similar Republicans. It was coded as zero in all other instances.

Past House Member: Past House Member was equal to 1 if the Democratic candidate previously held a seat in the House of Representatives but was not an incumbent at the time of the election. Cases in which Republicans had previously held such offices were equal to -1. All other situations were coded zero for this variable.

Quality Candidate had the same values as *Past House Member* when Past House Member was non-zero – i.e either -1 or +1. This means that the coefficient for *Past House Member* represents the additional positive effect that having served in the US House had on the challenger's chances of winning the race, over and above that of the benefit of having held any other elective office.

The effect of prior service in the House depended upon whether or not the candidate was running against an incumbent. The variables *Past House Member, Closed* and *Past House Member, Open* were coded as 1 when only the Democratic candidate had one of these attributes and as -1 when only the Republican candidate had one of them. These variables were coded as zero in all other instances.

National Variables

Democratic Vote Intention: In Gallup polls, this is the percentage of respondents who express the intention to vote for a Democratic House candidate in the upcoming election (of people saying they would vote for either a Democrat or a Republican). The poll used was the one conducted closest to March 10 of even-numbered years. Unfortunately, Gallup did not ask this question in 1988, so we had to exclude 1988 from our analysis.

Presidential Approval: This was the percentage approval rating for the President as expressed in the same Gallup poll (of those respondents who expressed approval or disapproval). Note that when a Republican was president, this variable showed the percentage of respondents disapproving of the president.

Change in Real Disposable Income: This measure was chosen as an indicator of the performance of the economy. It was the percentage change that occurred in per capita real disposable income between the February of the year before the election and the February of the election year. To reflect the expectation that Democratic candidates should have been hurt by a good economy when a Republican was in the White House, we multiplied this variable by -1 when a Republican was president.

Midterm Penalty: This variable captured the tendency of the party of the President to lose votes in a midterm election. This variable was coded as 1 for midterm elections in which a Democrat was in the White House, as -1 for midterm elections in which a Republican was in the White House, and as zero for presidential election years.

DISCUSSION OF REGRESSION RESULTS

In Table 1 we display our regression results. With the adjusted R-Square = .856, the model explains about 85% of the variance across elections in the Democratic percentage

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

INDEPENDENT VARIABLES	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)
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

of the two-party vote. The Standard Error of the Estimate indicates that on average the model comes within about 7 percentage points of predicting the Democrats' share of the House election. All influencing variables have the expected signs, all but two are statistically significant at conventional levels ($p<.05$), and most are highly significant ($p<.01$).

A few examples will illustrate what the coefficients in Table 1 mean. The coefficient associated with *Incumbent* indicates that, other things equal, incumbents received an average advantage over non-incumbents of 8.4 percentage points in the elections between 1974 and 2004. *Quality Candidates*, those with prior elective experience but not in the U.S. House, gained about 3 percentage points over those without prior elective office. The *Midterm Penalty* appears to have cost House members of the president's party about 3.6 percentage points.

PREDICTIONS FOR NOVEMBER 2006

To use the model to predict the 2006 House elections, we had to determine the attributes of the candidates running in 2006. We obtained lists of candidates and their incumbency status or prior experience from Ron Gunzburger's "Politics1.com" Web site (Politics1.com 2006).

However, as of April 2006, many states had not yet held their primaries; in these cases, we could not be certain who would receive the nomination. So we created three scenarios: one as favorable as possible to the Democrats, a second as favorable as possible to the Republicans, and a third that we considered the most probable primary winner. We used the most probable scenario to make our official model predictions. However, the results from the first two scenarios provided a range within which the probability of Democratic success lay. Surprisingly, this range turned out to be less than 2 percentage points.

Table 2. Predicted Outcomes of the 30 Most Competitive House Races

DISTRICT BY STATE AND NUMBER	PREDICTED PERCENTAGE CHANCE OF DEMOCRATIC VICTORY	PREDICTED DEMOCRATIC PERCENTAGE OF TWO-PARTY VOTE	ACTUAL DEMOCRATIC PERCENTAGE OF TWO-PARTY VOTE
CT 2	51.9	50.3	50.0*
AZ 8	47.7	49.6	56.2*
CT 4	52.7	50.5	48.4
IL 6	46.4	49.4	48.7
GA 3	54.4	50.7	32.4
NM 1	45.3	49.2	49.6
IN 9	43.0	48.8	52.3*
FL 13	42.5	48.7	49.9
PA 8	41.6	48.5	50.3*
MN 6	60.3	51.8	45.7
CA 50	62.1	52.1	44.9
PA 6	35.2	47.4	49.4
NH 2	33.3	47.1	53.6*
WA 8	31.6	46.7	48.5
OH 12	31.5	46.7	41.9
NY 24	31.5	46.7	54.3*
PA 15	30.6	46.5	45.5
CO 4	30.4	46.5	48.2
NJ 7	29.7	46.4	49.1
OH 1	29.4	46.3	47.2
OH 15	29.4	46.3	49.1
CT 5	28.2	46.1	56.2*
WI 8	27.0	45.8	51.2*
IA 1	73.5	54.3	56.0*
MN 2	26.3	45.7	41.6
FL 22	25.8	45.6	52.0*
TX 22	24.7	45.3	55.3*
IA 2	23.3	45.0	51.4*
IN 8	22.7	44.9	61.0*
WV 2	22.5	44.9	42.8

* = Democratic capture of a House seat

Our model supplied a prediction for each of the 435 races for seats in the House of Representatives. Of 379 contested races, the model correctly called the winner of 348 (91.8%), and predicted the overall Democratic share of the vote to within 0.1%. The median absolute value of the difference between actual and predicted amounts in these contested elections was 3.0 percentage points. Given that many House seats are uncompetitive, this is not a notably high level of accuracy. For example, of the 59 seats that were won by less than 10 percentage points we called only 36 correctly (61.0%).

Table 2 shows the model predictions and the actual outcomes in what we had called the 30 most competitive races. The media also had identified many of the contests noted in Table 2 as "races to watch." For example, the *Washington Post* considered 12 of the races from Table 2 to have been among the top 20 races to watch (Washingtonpost.com 2006). *Congressional Quarterly* (CQ) considered 15 of the races from Table 2 to have been among the top 32 most competitive races (CQpolitics.com).

The first column of Table 2 shows the probability of a Democratic victory in the district. In order to determine the probability that the Democrats would recapture the House (i.e. win at least 218 seats), we conducted a Monte Carlo Simulation. (Ed note: The article by Sam Sugiyama in this issue of *Foresight* provides a tutorial on Monte Carlo Simulation. See especially his section entitled *Simple Illustrative Example.*)

We generated normally distributed random variables, multiplied them by the standard error of the estimate for the model as a whole, and added them to the estimated percentage of the vote that Democrats would get in any particular district. Districts in each simulation were then classified as either Democratic or Republican victories. Finally we aggregated the results into an overall prediction of how many seats the Democrats would win.

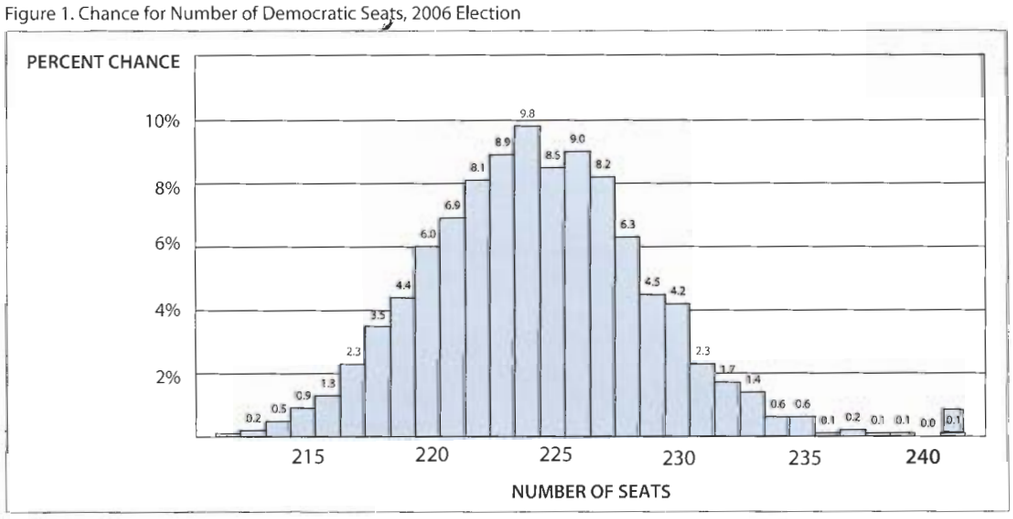
Figure 1 displays the probability distribution of the number of seats that the model predicted as wins for the Democrats in the 2006 election. The median estimate – the 50th percentile of this distribution – was 224 seats, 6 more than needed for control of the House. This prediction constituted a gain of 22 seats for the Democrats. Overall, the model simulation resulted in a 94.9% chance that the Democrats would win at least 15 additional seats, thus taking control of the House. If we ignore the most extreme 2.5% of results at the low and high ends of our estimates, we are left with a 95-percent probability that Democrats would control between 216 and 233 seats after the election. (The final tally was 233, the upper limit of this confidence interval.)

Our simulations further indicated that even when we used a scenario of primary winners most favorable to the Republicans, the Democrats were still predicted to win 220 seats with a 70% chance of taking majority control of the House.

OTHER FORECASTS

Our election forecast was in sharp contrast to the *seat-by-seat* analyses of the midterm House elections published earlier in 2006 by the Congressional Quarterly. *CQ* called 224 seats for Republicans. Even if Democrats had won all of the nine seats that *CQ* had labeled "no clear favorite,"

Figure 1. Chance for Number of Democratic Seats, 2006 Election



they would have been left with 211 seats in the House (CQpolitics.com 2006). At the other extreme, we had looked at the relationship in past elections between Democratic vote intention and the percent of seats Democrats won. Our bivariate analysis yielded predictions that Democrats would win 250 seats.

The results of this model lay between these extremes. However, our estimates were closer to *CQ*'s than to the simple national partisan tide model. We would argue that this was a testament to the insulation of most House districts from national partisan tides (McGhee, 2004; Campbell, 2003). The results of the election provided strong evidence that of the three methods (district by district, aggregate, or a "combined approach," as we used here), our combined approach yielded the most accurate forecasts.

REFERENCES

Campbell, J.E. (2003). The 2002 midterm election: A typical or an atypical midterm? *PS: Political Science and Politics*, 36, 203-207.

CQPolitics.com (2006). Balance of power scorecard: House, http://www.cqpolitics.com/risk_rating_house.html (accessed 28 April 2006).

Highton, B. (2000). Senate elections in the United States: 1920-94, *British Journal of Political Science*, 30, 483-506.

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

McGhee, E. (2004). National tides and local results in U.S. House elections. Presented at the Annual Meeting of the American Political Science Association, Chicago, Sept., 2004 http://www.allacademic.com/meta/p60265_index.html

Politics1.com (2006). The 50 state pages: State & congressional candidates, campaigns, elections, state parties, local news sources and more, http://politics1.com/states.htm, (accessed 28 April 2006).

Washingtonpost.com (2006). 2006 Election results, http://projects.washingtonpost.com/elections/keyraces/map/ (accessed 27 April 2006).

Contact Info:
Carl E. Klarner
Indiana State University
cklarner@isugw.indstate.edu

Stan Buchanan
Indiana State University
sbuchanan1@isugw.indstate.edu

UPCOMING
FORESIGHT 2007

PHARMAFORECASTING: A NEW FEATURE SECTION

The Changing Role of the Pharmaceutical Forecaster

How to Project Patient Persistence

How to Capture Uncertainty in Pharmaceutical Forecasts

FORECAST PROCESS IMPROVEMENT

S&OP as a Knowledge-Creating Process

Good and Bad Judgment in Forecasting: Lessons from Four Companies

FORECASTING PRINCIPLES & METHODS

A Primer on Structural Models for Users of Exponential Smoothing

Prediction-Realization Diagrams to Depict Forecasting Performance

Scenarios for Forecasting the Unforecastable

ASK THE EXPERTS

A new column addressing key concerns of business forecasters and their organizations

REGULAR FEATURES

Forecast Accuracy Measurement

Hot New Research Column

Forecasting Software Column

The World of Forecasting

Book Reviews

If you are interested in contributing an article or commentary on any forthcoming article, please contact Len Tashman, FORESIGHT Editor, lentashman@cs.com