SYMPOSIUM

Forecasting House Seats from Generic Congressional Polls: The 2010 Midterm Election

Joseph Bafumi, *Dartmouth College*
Robert S. Erikson, *Columbia University*
Christopher Wlezien, *Temple University*

In this article, we present a forecast of the 2010 midterm House election based on information available in early July 2010. We combine this forecast with a note of caution, explaining why electoral circumstances might lead our forecast to err. Finally, we present guidance regarding how to update the electoral forecast for 2010 based on new information that will become available leading up to Election Day.

Our forecast, based on information gathered in early July 2010, is a likely Republican takeover of the House. In terms of the national vote, the most likely outcome is a solid Republican plurality of close to 53% of the two-party vote. Damage to the Democrats is minimized somewhat in the translation of votes into seats. By our reckoning, the most likely scenario is a Republican majority in the neighborhood of 229 seats versus 206 for the Democrats. Assuming our model is correct but taking into account its uncertainty, the Republicans have a 79% chance of winning the House.

THE MODEL

Our multi-level prediction builds on our successful model predicting the House seat division in 2006 (Bafumi et al. 2008).[1] The model has two steps. Step 1 predicts the midterm vote division from only two variables, the generic poll result and the party of the president. With this measure of the partisan tide in place, step 2 forecasts the winners of 435 House races using separate statistical models for open seats and races with incumbent candidates. At each step, the forecast takes into account uncertainty about the inputs. The final product is a prediction of the partisan division of House seats, as well as a probabilistic statement regarding the likelihood of the Republican Party regaining control of the chamber.

STEP 1: PREDICTING THE VOTE

Our prediction of the national division of the two-party vote is based on two independent variables. One variable is the recent reading of the generic polls—the frequently asked poll question regarding preferences on a generic (no candidate names) partisan ballot for Congress.[2] The second variable is a dummy variable for the party holding the presidency. We have modeled this process in detail elsewhere (Bafumi et al. 2010). In past congressional campaigns, generic polls have persistently underestimated the ultimate support for the non-presidential ("out") party. This tendency is greatest early in the election year and recedes as the campaign progresses. Poll respondents increasingly take into account the party of the president when reporting their generic vote. As the campaign progresses, they tilt toward the out-party. Our interpretation (Bafumi et al. 2010) is that voters seek more ideological balance between the president and Congress.

Remarkably, forecasts from our two variables—the generic poll results and the party of the president—are about equally predictive regardless of when the poll results are taken. In short, the midterm vote tends to be determined early, with the campaign serving mainly to draw the voters toward the out-party. Taking into account election-year changes in the president's approval rating or economic conditions has not offered any improvements to the forecast equation in the past (Bafumi et al. 2010). Our two variables predict the November vote approximately as well as final presidential trial-heat polls predict the presidential vote (see Wlezien and Erikson 2002).

For our forecast, we measure the Democratic percent of the two-party vote (minus 50%) in reported generic ballot polls conducted by personal interview (no robotic or Internet polls).[3] A slight adjustment is necessary to combine polls from different target populations—specifically, adults, registered-voter, and likely-voter polls. Reflecting the historical tendency for adult and registered-voter polls to favor Democrats more than likely-voter polls, we subtract 1.85 percentage points from the Democratic margin for polls of adults and 1.04 percentage points for registered-voter polls, thereby norming the measure, as with likely-voter samples.[4] With this adjustment and the generic polls measured 121 to 180 days in advance of the election (centered in early June), the vote forecasting equation for the 16 midterm elections between 1946 and 2006 is[5]:

$$\text{Nat'l Dem Vote\%} = -0.50 + 0.50\ \text{DemGenericPoll\%} - 1.99\ \text{PresParty} + e_j \quad (1)$$

where (to aid interpretation) the vote and poll variables are measured as deviations from 50%, and PresParty = 1 if a Democratic president and –1 if a Republican president. Both independent variables are statistically significant at the .001 level; the intercept is not significant. The adjusted R^2 for the equation is 0.73, and the root mean squared error (RMSE) is 1.94.

The pooled generic polls conducted 121 to 180 days in advance of the 2010 election show a very close division of

doi:10.1017/S104909651000106X

https://doi.org/10.1017/S104909651000106X Published online by Cambridge University Press

49.1% Democratic and 50.9% Republican.[6] But a virtual tie in the polls at this point projects a significant vote plurality for the Republicans as the out-party. This predicted outcome is not due to any bias in the polls, but rather stems from the electorate's tendency to gravitate further toward the "out" party over the midterm year—ultimately gaining about two extra points beyond what the June polls would otherwise show.

Our specific forecast is that the Democrats will win 47.1% of the two-party vote and the Republicans the remaining 52.9%. We take into account the uncertainty about this forecast by simulating 1,000 cases based on the forecast error associated with our prediction for 2010.[7] This model yields a probability density as a distribution around the forecast. By this estimate, the 95% confidence band is a range from 43.1% to 51.1% Democratic. The Republicans win the popular vote 91% of the time.

STEP 2: PREDICTING SEATS

The national vote only tells us part of the story, and we still need to determine how it would translate into seats. For each simulated value of the national vote, we need to simulate the outcome in 435 congressional districts. The district vote (D_{jk}) is a function of the stochastic simulations of the national vote (simulation j) and the local conditions (simulation k):

$$D_{jk} = N_j + L_k + u_k$$

where L_k = the expected local component of the vote in the kth district and u_k = the simulation of the district k error. The national vote in the jth simulation (N_j) consists of the specific prediction (P) from equation 1 plus the error (e_j) around that prediction. Substituting in these components yields:

$$D_{jk} = P + L_k + e_j + u_k$$

The formula for L_k differs for open seats and incumbent races. For open seats (no incumbent running), the district vote simulation comes from equation 2:

$$D_{jk} = -3.92 + (0.82)\ \text{Obama\%}_k + P + e_j + u_{k_\text{open}} \qquad (2)$$

where Obama%$_k$ = the Obama percentage of the two-party vote in district k in 2008. D_{jk} and Obama% are calibrated as deviations from 50%. For seats of districts where the incumbent seeks reelection, the district vote simulation comes from equation 3:

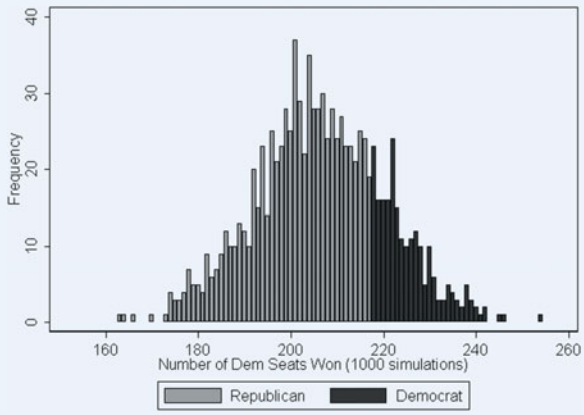
$$D_{jk} = -5.79 + 0.98\ \text{DemVote}(2008)_k + 3.85\ \text{Frosh}_k + P + e_j + u_{k_\text{incumbent}} \qquad (3)$$

where Dem Vote (2008)$_k$ = the Democratic vote percentage for the House in district k in 2008 (minus 50%) and Frosh$_k$ = 1 if a freshman Democrat in 2010, –1 if a 2010 freshman Republican, and 0 otherwise.

The numerical parameters of equations 2 and 3 are derived from regression equations predicting (1) the 2008 district vote for 2008 open seats with the presidential vote in that district, and (2) 2008 incumbent races with the 2006 district-level vote for the incumbent plus the freshman indicator.[8] Recalling that all vote variables are calibrated as deviations from 50% Democratic, the two constant terms represent the expected vote when the partisan trend plus the Obama district vote or the lagged incumbent district vote equal 50% Democratic. For equation 2, the constant term equals the intercept for the 2008 open-seat equation minus the 2008 national vote for Democrats. For equation 3, the constant term is set to the value at which, given any value of the national vote, the projected average change from 2008 to 2010 in the mean district vote equals the projected 2008–10 change in the national vote.

Figure 1

One Thousand Simulations of the 2010 Election



Importantly, the RMSEs from the two 2008 regression equations provide the standard deviations for u_{k_open} and $u_{k_\text{incumbent}}$. For open seats, this value is 6.86. For incumbent races, it is 4.91. These values represent the uncertainty around the district vote after taking into account the projected district vote from the 2008 presidential vote (open seats) or 2008 congressional vote (incumbent seats). For vote estimates drawn from $P + e_j$, we simulate the winner in each district. We conducted 1,000 simulations to take into account the variability characterized by the associated error terms, e_j and u_{k_open} or $u_{k_\text{incumbent}}$.

FORECASTING 2010

First, we generated 1,000 simulations of the national vote based on equation 1. Then, we applied each of the 1,000 simulated national outcomes to each congressional district, using the formulas shown in equations 2 and 3, predicting the likelihood of a Democratic victory in each. For each of the 1,000 simulated vote outcomes, we arrived at a projected outcome in terms of the partisan division of 435 congressional districts. Figure 1 displays the resulting density of results. As can be seen from the predominance of gray bars, the Republicans win the majority of seats in 79% of the trials. On average, the Republicans win 229 seats, 23 more than the Democrats and 11 more than the 218 needed for a majority. However, the simulations yield considerable variation, with a 95% confidence interval of 176 to 236.[9]

Note that the Democrats' chance of holding the House (21%) is greater than the chance of the Democrats winning the popular vote (9%). This apparent discrepancy arises because the

https://doi.org/10.1017/S104909651000106X Published online by Cambridge University Press

critical value of the popular vote at which the identity of the majority party becomes a 50-50 proposition is 48.8%. The Democrats need not win the majority of the vote to win the most seats. The fact that popular Democratic incumbents currently hold many of the seats that would normally be competitive provides the Democrats with a partial barrier against a Republican deluge.

A NOTE OF CAUTION

Applying our model of the vote to 2010 assumes that the forces at work in 2010 are unchanged from past midterm elections. However, we should be wary of the possibility that the underlying model of the national vote works differently in 2010 or is influenced by variables we have not taken into account. In particular, we note that in past elections, the out-party usually has increased its support in the generic polls as the election year campaign has progressed. But this increase did not happen in the first six months of 2010. The generic polls have hovered around a 50-50 division rather than sliding further in the Republicans' favor, which may reflect a strong public engagement in the campaign far earlier than has occurred in the run-up to previous midterms. From the start, the 2010 campaign has had a sharper ideological tone than in the past. The tilting toward the out-party may thus have been realized earlier than in the past.[10] It also is possible that an unusually conservative Republican campaign could dampen the expected shift toward the Republicans. If either of these possibilities is true and the polls remain unchanged through November, the Democrats actually could have the edge come Election Day.

FORECASTING FROM OCTOBER POLLS

From the perspective of the reader in October of 2010, forecasts based on late polls from October are more valuable than from July, our publication deadline. As an aid to predict the outcome late in the campaign, we present the vote forecasting equation based on generic polls during the final 30 days of past campaigns in the 16 midterm elections between 1946 and 2006:

$$\text{Nat'l Dem Vote\%} = 0.21 + 0.59\ \text{DemGenericPoll\%} - 0.82\ \text{PresParty} + e_j \quad (4)$$

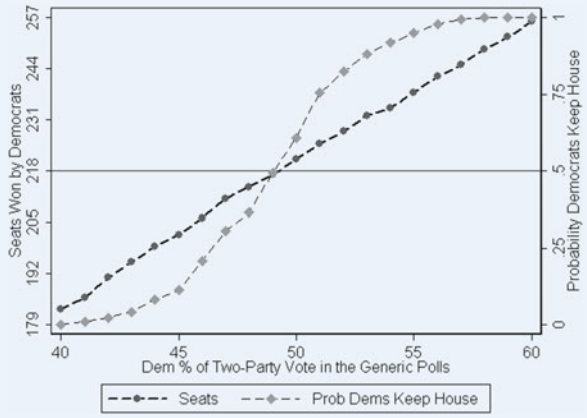
where the vote and poll variables are measured as deviations from 50%, and PresParty = 1 if a Democratic president and –1 if a Republican president.[11]

The key difference between equation 4 and equation 1 is that the penalty for belonging to the presidential party is much lower in October, declining from almost two points to less than one point. Should the generic polls stay close to 50-50, as they were through early July, prospects will look much better for the Democrats. At 121 to 180 days out, the Democratic share in the generic polls was 49.1%. A continuation at this value into October would project the Democrats winning 48.8% of the actual national vote and having an even (50%) chance of retaining the House.

As a guide, we have estimated both the expected seat outcome and the probability of Democratic control, given a continuum of scenarios regarding the late generic polls. Figure 2 shows the results. The Democrats are favored to maintain control of the House as long as they win at least 49.2% of the two-party vote in October's (likely voter) generic polls, which converts to an expectation of 48.9% or more of the popular vote. Should they hold a 50% poll share at the end of the campaign, the Democrats will have a 63% chance to retain the House.

Figure 2

Simulations of the 2010 Election Outcome, Conditional on the Generic Polls in the Final 30 Days before the Election



CONCLUSION

This article offers an advance assessment of the likely distribution of seats in the House following the 2010 election. We first forecasted the national vote based on the historical relationship between generic polls, the party of the president, and the vote in previous midterm elections. Taking into account the expected vote using polls from early July and the unique circumstances of 435 House districts, we then simulated the final Election Day outcome. The average result of our simulations is a 229 Republican to 206 Democratic split of seats, but with a wide dispersion of possible outcomes. The Republicans have a 79% chance of gaining the House. The uncertainty about party control largely reflects uncertainty in our forecast of the national vote. Seemingly minor variation in the national vote can have major consequences for the distribution of seats. ■

NOTES

1. Using an earlier version of the model presented here, on October 24, 2006, we forecasted a 32-seat gain by the Democrats. In November, the Democrats gained 30 seats. The model generated equally solid predictions when applied from the beginning of the 2006 election year.
2. Respondents typically are asked for which party they plan to vote (or whom they want to win) "if the election were being held today," although variation in question wording occurs. For instance, some organizations use the wording "looking ahead to the Congressional elections in November" while others use "thinking about the next election for U.S. Congress."
3. This restriction to interviews via personal contact (telephone or in-person) is not intended as a statement about the relative accuracy of different kinds of polls. Our historical analysis of past polls excludes

https://doi.org/10.1017/S104909651000106X Published online by Cambridge University Press

robotic and Internet polls, and the exclusion of these polls for our 2010 estimate ensures procedural continuity.

4. The adjustments are based on analysis of generic polls in all midterm election years between 1946 and 2006.
5. Equation 1 differs slightly from the equation reported in Bafumi et al. (2010). Subsequent to publication, a slight glitch was found in our algorithm for merging the marginals from adult, registered, and likely voter polls. This correction does not affect any substantive conclusions. For specifics, see http://www.dartmouth.edu/~jbafumi/.
6. The average of poll percentages unadjusted for polling universe is 50.1% Democrat and 49.9% Republican.
7. The forecast error (2.17) takes into account both the RMSE of the equation and the uncertainty about the coefficients.
8. We treat 2010 districts with bi-elections between 2008 and 2010 as incumbent races, but with an adjustment for the lagged vote from the interim period. First, we predict the 2008 congressional vote by applying the predictions from the 2008 open-seat equation and subtracting the mean deviation. We then treat these cases as 2010 incumbent races, using the adjusted measure of the lagged vote. In Georgia's ninth district, the special election resulted in a one-party run-off, so the vote was calculated based on the partisan division in the primary. In New York's special elections, the votes of minor parties who cross-endorsed the Republican or Democratic candidate were added into each party's total. In New York's twenty-third district, the votes for the Republican Party candidate (who dropped out) and the Conservative Party's candidate (who became the Republican Party's candidate in the race) were added together. In Hawaii's first district, the votes for both of the Democrats in the general election race were added together. Apart from special elections, Louisiana's second district required special attention. This district is very Democratic but with a Republican incumbent who won because of the Jefferson bribe scandal. We assigned this race to go Democratic in 2010. All incumbent-contested seats that were uncontested in 2008 were assigned to the incumbent. Districts that were uncontested in 2008 but open in 2010 were predicted with equation 2.
9. Most of the variation is due to uncertainty about the national vote, with a lesser amount due to seeming chance variation in the districts once the national vote is known. For instance, suppose we assign each district the predicted vote from equation 1 (ignoring the disturbance term), assuming the equation predicts the vote perfectly. This version of the simulations produces a 95% confidence interval of 202 to 210, a range of only eight seats.
10. In 2009, the generic polls already revealed a plunge in the Democratic share of the vote, from approximately 60% at the beginning of 2009 to about 50% at the end of the year.
11. The adjusted R^2 for the equation is .77 and the RMSE is 1.77.

REFERENCES

Bafumi, Joseph, Robert S. Erikson, and Christopher Wlezien. 2008. "Forecasting House Seats from Generic Congressional Polls." In *Elections and Exit Polling*, ed. Wendy Alvey and Fritz Scheuren, 118–23. New York: Wiley and Sons.

———. 2010. "Balancing, Generic Polls, and Midterm Congressional Elections." *Journal of Politics* 72: 705–19.

Wlezien, Christopher, and Robert S. Erikson. 2002. "The Timeline of Presidential Election Campaigns." *Journal of Politics* 64: 969–93.

https://doi.org/10.1017/S104909651000106X Published online by Cambridge University Press