

National Polls, District Information, and House Seats: Forecasting the 2014 Midterm Election

Joseph Bafumi, *Dartmouth College*

Robert S. Erikson, *Columbia University*

Christopher Wlezien, *University of Texas at Austin*

This article presents a forecast of the 2014 midterm House election based on information available four to six months in advance. The approach builds on our forecasts of the 2006 (Bafumi, Erikson, and Wlezien 2008) and 2010 (Bafumi, Erikson, and Wlezien 2010a,b) midterm elections.¹ We incorporate information about the national forces at work in the election, which are evident early in the election year from generic congressional polls plus the party of the president. We also incorporate information about the districts themselves, which is reflected in their partisan predispositions and in other ways, most notably, whether the incumbent seeks reelection. To forecast the 2014 election, we simulate the national vote and district outcomes using the past as our laboratory, details about which we provide in the text that follows.

Our forecast, based on information gathered 121 to 180 days in advance of the election (essentially May and June), is a near-certain Republican hold of the House. In terms of the national vote, the most likely outcome is a Republican plurality of about 52.5% of the two-party vote. Of course, seats are what matter, and by our reckoning, the most likely scenario is a Republican majority in the neighborhood of 248 seats versus 187 for the Democrats. Taking into account the uncertainty in our model, the Republicans have about a 1% chance of losing the House. As circumstances can change during the election year, we provide guidance for updating the forecast based on new information that will become available leading up to Election Day.

The expectation from our model is that the Republicans will win more seats in 2014 than they did even in their record showing in 2010, although with a lesser share of the vote than in 2010. If the forecast holds, why would a lesser voter share than 2010 yield more seats than 2010? First, unlike in 2010, it is the Republicans who hold the incumbency advantage. Seats they barely won in 2010 or 2012 are safer now with a Republican incumbent. A second factor is the reinforced Republican gerrymander following their capture of state legislatures in 2010. Still a third factor is the redistricting following the 2010 census. Because of voter migration, most newly created districts in 2012 are in Republican areas.

THE MODEL

As mentioned, our prediction model has two steps. The first step predicts the national vote division from two variables, the

generic poll result and the party of the president. With this estimate of the partisan tide, the second step forecasts the winners of 435 House races using separate models for open seats and for races with incumbent candidates. At both steps, the forecast takes into account uncertainty about the inputs and their effects. The final product of our simulations is a prediction of the partisan division of House seats, as well as a probabilistic statement regarding the likelihood of the Democratic Party regaining control of the chamber.

STEP 1: PREDICTING THE VOTE

We first predict the national division of the two-party vote using two independent variables. One is the current 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 is a dummy variable for the party holding the presidency. It is well-known that voters tend to punish the incumbent president's party during midterm elections, and we have shown, based on past congressional campaigns, that generic polls persistently underestimate the ultimate support for the nonpresidential (“out”) party (see Bafumi, Erikson, and Wlezien 2010a). The underestimate is greatest early in the election year and recedes as the campaign progresses, whereby poll respondents increasingly take into account the party of the president when reporting their generic vote. As the campaign progresses, voter preferences tilt toward the out-party. Our interpretation 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 during the election year the poll results are taken. In short, the midterm vote tends to be determined early, by the beginning of the year, 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 yields no improvement in the forecast equation (Bafumi, Erikson, and Wlezien 2010a).

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 registered voter polls with likely voter polls because the former yields estimates that are more favorable to the Democrats. Based on our statistical

analysis of past generic polls, we subtract 1.42 percentage points for registered-voter polls. This adjusts registered-voter polls to reflect likely-voter samples.⁴ With this adjustment and the generic polls measured 121 to 180 days in advance of the election, the vote forecasting equation for the 17 midterm elections between 1946 and 2010 is:

$$\text{National Vote\%} = -0.14 + 0.53 \text{ Generic Poll\%} - 2.04 \text{ Presidential Party} + e_j, \quad (1)$$

where (to aid interpretation) the vote and poll variables are measured as deviations from 50% of the Democratic vote, and Presidential Party = 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 0.75, and the root mean squared error (RMSE) is 1.94.

The pooled generic polls conducted 121 to 180 days in advance of the 2014 election show a very close division of 49.3% Democratic and 50.7% Republican (in terms of likely voters). But a slight Republican lead in the polls at this point projects to a significant vote plurality for the Republicans as the out-party. The predicted outcome is not because of any bias in the polls, but rather stems from the electorate's tendency to gravitate further toward the "out" party during the midterm year—ultimately gaining about two extra points beyond what the June polls show.

Our specific forecast is that the Democrats will win 47.5% of the two-party vote and the Republicans, the remaining 52.5%. Of course, we are not absolutely sure that this will be the vote; what we really are forecasting is a distribution of likely results.

Our specific forecast is that the Democrats will win 47.5% of the two-party vote and the Republicans, the remaining 52.5%. Of course, we are not absolutely sure that this will be the vote; what we really are forecasting is a distribution of likely results.

To take into account the uncertainty, we simulate the vote in 1,000 "elections" based on the forecast error associated with our prediction for 2014.⁵ This yields a probability density as a distribution around the forecast of the national vote. By this estimate, the 95% confidence band is a range from 44.7% to 50.3% Democratic. In other words, the Republicans are almost certain to win the popular vote.

STEP 2: PREDICTING SEATS

Next, we need to determine how the swing in national vote will influence the number of actual seats the parties win. For each simulated value of the national vote, we need to simulate the outcome in the 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 k th district and u_k = the simulation of the district k error, the

latter of which reflects our uncertainty about the prediction. Likewise, the national vote in the j th 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:

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

If there is a major party candidate running unopposed in 2012, we assign the seat to the solo candidate's party, even if contested in 2014. For the 360 districts contested in 2012 and presumed to be contested in 2014, we estimate the change in the mean district vote. This is identical to the projected change from 2012 to 2014 in the national vote division, but with an adjustment because of the expected differential in turnout rates from 2012 to 2014. Based on our forecast of the national vote, the expected swing of the 2012–2014 national vote is -3.21. For reasons explained in the technical appendix, we expect the swing of the mean district vote to be slightly less, an average drop of 2.54 points in the Democratic vote.

Based on our estimates of the mean district vote swing (-2.54), we simulate the open-seat and incumbent-contested elections. For open seats, our template is the equation predicting the 2010 district vote from the Obama vote in 2008. This equation is:

$$\text{DemVote\%2010}_k = -6.01 + 0.95 \text{ Obama2008\%}_k + u_k. \quad (2)$$

In our simulations, we substitute the 2012 Obama vote for the 2008 Obama vote in equation 2 and adjust the intercept so that the mean 2014 vote in open seats is 2.54 points less Democratic

than in 2012, and with a variance based on the forecast error of the national vote forecast. We include error variances based on the root mean squared error of the 2010 open seat equation.

For incumbent-contested seats, our template is an equation predicting the district vote from the 2008 Obama vote plus the 2008 Democratic vote for the House plus a dummy for freshmen in 2010 (-1 = Republican freshman, +1 = Democratic freshman, 0 = veteran).⁶ This equation is:

$$\text{DemVote\%2010}_k = -7.54 + 0.63 \text{ DemVote\%2008}_k + 0.46 \text{ Obama\%2008}_k + 2.03 \text{ Frosh}_k + u_k. \quad (3)$$

In our simulations of the incumbent-contested seats in 2014, we substitute the 2012 Obama vote for the 2008 Obama vote and the 2012 congressional vote for the 2008 congressional vote in equation 3.⁷ We adjust the intercept so that the mean 2014 vote in incumbent-contested seats is 2.54 points less than the 2012 mean, with a variance based on the forecast error of the national vote forecast. We include error variances based on the root mean squared error of equation 3, the 2010 incumbent-contested equation.⁸

FORECASTING 2014

First, we generated 1,000 simulations of the national vote based on equation 1. Then, taking each of these simulated national outcomes, we simulated the vote in each congressional district using the formulas shown in equations 2 and 3 for all 2014 open seats as well as seats with incumbents in 2014 that were contested in 2012. 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) outcomes. As can be seen from the predominance of solid bars, the Republicans win the majority of seats in nearly all of the trials. On average, the Republicans win 248 seats, increasing their margin substantially, by 14 more than what they won in the 2012 election. However, the simulations yield considerable variation, with a 95% confidence interval of 228 to 268.⁹

FORECASTING FROM OCTOBER POLLS

When this article reaches print in October of 2014, forecasts based on current polls will be more valuable than those using information from the spring, our publication deadline. As an aid to predict the outcome late in the campaign, we present the vote forecasting equation using generic polls from the final 30 days of past campaigns in the 17 midterm elections between 1946 and 2010:

$$\text{National Vote\%} = 0.58 + 0.62 \text{ Generic Poll\%} - 0.74 \text{ Presidential Party} + e_j, \quad (4)$$

where the vote and poll variables are measured as deviations from 50% of the Democratic vote, and Presidential Party = 1 if a Democratic president and -1 if a Republican president.¹⁰

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 more than two points to less than one point. Not surprisingly, the coefficient for the polls is slightly larger. Should the generic polls stay close to 50-50, as they were through the spring, prospects will look a little better for the Democrats. At 121 to 180 days in advance of the election, the Democratic share in the generic polls was about 49%. A continuation at this value into October would project the Democrats winning 49.2% of the actual national vote with a 5% chance of regaining 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 Republicans are favored to maintain

Figure 1

One Thousand Simulations of the 2014 Election

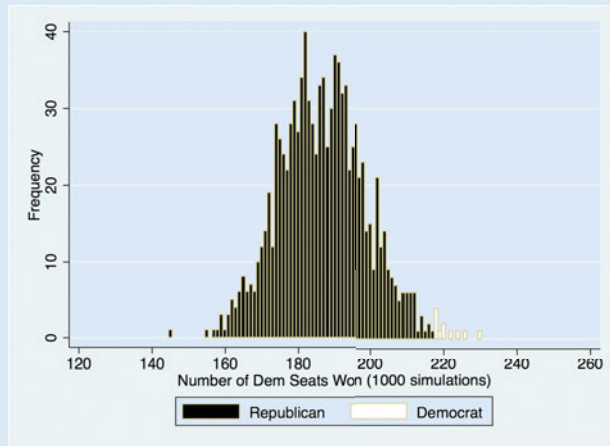
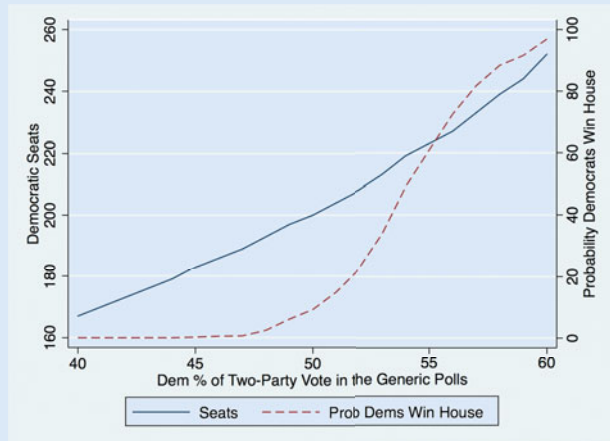


Figure 2

Simulations of the 2014 Election Outcome, Conditional on the Generic Polls during the 30 Days before Election Day



Should they hold a 50% poll share at the end of the campaign, the Republicans will have a 90% likelihood of retaining the House.

...when the Republican Party captured new seats in the 2010 Republican tide, many of the new Republican members of Congress inoculated themselves from defeat when the Republican tide receded in 2012 with their newfound incumbency advantage.

control of the House as long as they have at least 46.5% of the two-party share in October's (likely voter) generic polls, which converts to an expectation of 48% or more of the popular vote.

Clearly, the Republicans hold a structural advantage in 2014 in terms of translating votes to seats. To begin with, the Republicans enjoy a long-standing advantage in the partisan composition

of congressional districts from the geographic tendency of Democrats to cluster in large cities (Chen and Rodden 2013). Democratic votes in these urban districts go wasted while Republicans are slightly favored in many other districts. Beyond that, when the Republican Party captured new seats in the 2010 Republican tide, many of the new Republican members of Congress inoculated themselves from defeat when the Republican tide receded in 2012 with their newfound incumbency advantage. Their larger numbers and associated incumbency advantage will continue to aid them in 2014 and beyond. The Republican Party was further helped by the partisan gerrymandering of compliant Republican state legislatures elected in 2010. Finally, the redistricting induced by the 2010 Census corrected district population imbalances that had favored the Democrats following a decade with no redistricting.

CONCLUSION

This article offers an advance assessment of the likely distribution of seats in the House following the 2014 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 121 to 180 days in advance of the election and the unique circumstances of 435 House districts, we then simulated the final Election Day outcome. The average result of our simulations is a 248 Republican to 187 Democratic split of seats, but with a wide dispersion of possible outcomes. The Republicans have a 99% chance of holding the House. The near certainty of this result is partly clear from the fact that the Republicans were leading in the generic polls from early in the election year. It is reinforced by the fact that voters tend to punish the sitting president at midterm elections, and this becomes increasingly clear during the course of the campaign. This is critical because the national vote is the most important structuring factor—seemingly minor variation in the national vote can have major consequences for the distribution of seats. But, even if Democrats make inroads on the national scene, they face an uphill battle against a majority of Republican incumbents aided by recent redistricting. ■

NOTES

1. Using an earlier version of the model presented here, we forecasted a 32-seat gain by the Democrats in 2006, close to the 30 seats they actually gained (Bafumi, Erikson, and Wlezien 2008). We did less well in 2010,

forecasting a 50-seat gain by the Republicans, which was 13 short than what they realized on Election Day (Bafumi, Erikson, and Wlezien 2010b).

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 there is variation in question wording. 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.” We do not attempt to account for the effects of these differences.
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, and is done to ensure procedural continuity. We used polls listed in RealClearPolitics.com and Pollster.com.
4. The adjustments are based on analysis of generic polls in all midterm election years between 1946 and 2010. First, we weigh each poll by its sample size. Then we regress the vote (weighted) on indicators for sample universe (adults, registered, likely) and midterm election year. The sample universe parameters provide the adjustment.
5. The forecast error (2.10) takes into account both the root mean squared error of the equation and the uncertainty about the coefficients.
6. The presidential vote variable was not included in our previous (2006 and 2010) midterm forecasts. It helps to predict the incumbent vote above and beyond the previous incumbent vote itself, which implies that partisanship’s predictive power has grown relative to the “personal” vote in recent times.
7. We treat 2014 districts with bi-elections (also known as special elections) between 2012 and 2014 as incumbent races. We use their bi-election vote outcomes but adjust them to reflect the likely outcome had the election been held in 2012. First we regress the 2012 vote in these districts on the 2012 presidential vote and generate predicted values from the regression. The average difference between the predicted vote in 2012 and the vote in the bi-election is used to adjust the bi-election outcome.
8. The RMSEs from equations 2 and 3 provide the standard deviations for u_{k_open} and $u_{k_incumbent}$. For open seats, this value is 5.47; for incumbent races, it is a predictably smaller 3.54.
9. Most of the variation in the seat outcome 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.
10. The adjusted R^2 for the equation is .82 and the RMSE is 1.75.

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. New York: Wiley and Sons, 118–23.
- . 2010a. “Balancing, Generic Polls, and Midterm Congressional Elections.” *Journal of Politics* 72: 705–19.
- . 2010b. “Forecasting House Seats from Generic Congressional Polls: The 2010 Midterm Elections.” *PS: Political Science and Politics* 43: 633–36.
- Chen, Jowei, and Jonathan Rodden. 2013. “Unintentional Gerrymandering: Political Geography and Electoral Bias in Legislatures.” *Quarterly Journal of Political Science* 8: 239–69.

APPENDIX

Because Democrats turn out less than Republicans, the mean district Democratic vote is always greater than the national Democratic vote. And the difference is greatest in midterm years when turnout is lowest. This makes a small difference in the expected mean vote shift conditional on the national vote shift. In 2012 contested seats, the mean Democratic vote was 0.49 greater than the summed votes in the contested seats. In 2010, the comparable differential was 1.54. Thus, we might think that we should subtract 0.49 from 1.54 to claim 1.05 percentage points to be deducted from the mean Republican seat swing.

However, the 2010 differential of 1.54 is distorted because the 2010 districts represent 2000 populations. We estimate the distortion to be 0.64 points, the difference between the mean 2008 Obama vote in 2010-apportioned districts and the new 2012-apportioned districts (the lagged vote in the new districts). In 2014, four years after the census, the amount of distortion would be 40% of the 2010 distortion, so we subtract only 60%, or $.6 \times 0.64 = 0.38$. Thus, we end up with $1.54 - 0.49 - 0.38 = 0.67$. We add 0.67 to the projected vote swing of -3.21 to get our estimate of -2.54 as the projected mean district vote swing, 2012–2014.