Election Fundamentals and Polls Favor the Republicans

Benjamin Highton, *University of California, Davis*
Eric McGhee, *Public Policy Institute of California*
John Sides, *George Washington University*

Our congressional forecasting model provides predictions of individual House and Senate races as well as aggregate party seat shares in each chamber. It does so by marrying an underlying structural or "fundamentals"-based model with available polling data—an approach similar to Linzer (2013).

The structural portion of the model is based on contested House and Senate elections from 1980 to 2012, excluding those when an independent or third-party candidate won a significant share of the vote.[1] The dependent variable is the Democratic candidate's share of the major-party vote. The independent variables are drawn from the extensive literature that has identified significant national and state or district correlates of congressional election outcomes (e.g., Jacobson 2012). These include:

- The nonannualized change in real gross domestic product (GDP) in the first two quarters of the election year.
- The president's average approval rating in June of the election year.
- Whether it is a midterm or presidential election year.
- Whether the seat is being contested by a Republican incumbent, a Democratic incumbent, or no incumbent (i.e., it is an open seat).
- The Democratic Party's share of the two-party presidential vote in each political unit (states or districts) in the concurrent or most recent election (mean-deviated by election year).
- Relative candidate quality, which is the difference between the Democrat's level of previous elective office experience and the Republican's. For House races, experience is coded as a dichotomy, distinguishing candidates who have held at least some elective office from candidates who have held no such office. For Senate races, each candidate is coded on a six-point scale: no elective office; a local office or state legislator; a statewide office other than governor (including former US senators); US House or large city mayor; governor; or incumbent US senator.
- The Democratic candidate's share of spending by the two major-party candidates.

For Senate elections, the model also includes:

- The previous Democratic vote share for the seat. The effect of this variable is allowed to vary by whether the incumbent senator is running for reelection. (Including lagged vote share in the model of House elections requires excluding election years that follow redistricting.)
- Whether an appointed senator is running for his or her first election, and that senator's party.

All variables are coded so that higher values capture better circumstances for Democratic candidates. For the three national-level variables, we multiply each by a variable for the president's party (–1=Republican; +1=Democrat) and also include that indicator in the model.

We account for uncertainty in our forecasts by assigning probabilities to all our predictions based on the error parameters in our models (Lauderdale and Linzer 2014). Likewise, because the structural model cannot account for all the factors that make each state and election year unique, we estimate a multilevel model with random intercepts for states and years so we can propagate that uncertainty more precisely into our predictions.[2] The results of the structural model for both House and Senate elections are shown in table 1.

Across the two types of elections notable similarity exists in the apparent effects of presidential approval and the midterm penalty. The estimated effects of campaign spending are also almost identical across the models. The estimate for presidential vote is larger for House elections, but that may be due to the inclusion of previous vote in the Senate models because both variables tap the local partisan balance. Also note that several candidate-related variables—incumbency, appointed, and quality differential—are not comparable across the election types because of the differences in model specification and measurement noted previously.

These parameter estimates may be used to make predictions for the 435 House and 36 Senate elections this year. We do this by substituting the 2014 values for each of the elections and computing the predicted vote shares. (Because final spending amounts will not be available until well after November, we use the most recent figures for fundraising during the 2013–2014 cycle.) Our estimates take into account uncertainty about the model parameters (the standard errors of the coefficients) along with the overall uncertainty in the model. In this process, the random error terms for years and states can be thought of as "offsets" that identify how each year or state differed from all others for unmeasured reasons. The multilevel model constrains these offsets to collectively form a normal distribution with

  doi:10.1017/S1049096514001279

mean zero and a variance estimated from the data. In addition to capturing the unmeasured variance, the offsets also contain estimation uncertainty—that is, uncertainty as to the precise value of each offset—which we also include in the predictions.

To produce our predictions, we simulate 1,000 outcomes for each election by drawing from the distributions estimated from the structural model discussed previously.[3] When doing so, we treat the year and state error terms somewhat differently. Because we do not know what kind of election year we will have, we draw from a normal distribution with the same mean (i.e., zero) and variance as the year error term and add these values to our predictions. By contrast, we know that the elections will be conducted in the same set of states as before, so we can simply add those state offsets to our predictions in something more closely approximating a "fixed effects" approach, while also accounting for the estimation uncertainty in the offsets. With these simulated elections, we make predictions about individual races and overall partisan seat distributions by aggregating our individual predictions and the uncertainty that comes with them.

Following the suggestion of Sides (2014) that forecasters estimate "heuristic" models that focus on different sets of key factors, we present the forecasts of different versions of the structural model in table 2. For each model, we present the predicted number of Democratic seats, the associated 95% prediction interval, and the estimated probability that the chamber will have a Democratic majority.[4] Models 1 and 2 in table 2 present two versions of the model. The first takes into account background characteristics of the election that are outside of candidates' control. The second builds in characteristics of the candidates.

The most valuable additional information that can be systematically incorporated into our predictions is polling data from individual elections, which are only available for the Senate contests.

At some points in the past, these two models produced different results, at least for the Senate. But as of now their predictions are very similar. The point prediction for Democratic seats hardly changes; the main difference is in the size of the confidence interval. Adding information about the candidates decreases the confidence interval, but the absence of movement in the point prediction also suggests that candidate quality and spending are roughly balanced between the two parties in the aggregate.[5] On the House side, both models give the Republicans overwhelming odds of retaining control. For the Senate,

Table 1

Parameter Estimates of Contested House and Senate Election Outcomes, 1980–2012

	HOUSE		SENATE	
	β	σ	β	σ
Intercept	40.16	0.51	39.77	0.79
Presidential vote	0.46	0.01	0.27	0.04
Incumbency	3.61	0.18	−2.36	0.75
Open	–	–	−0.85	0.56
Lagged vote	–	–	0.20	0.04
Open X Lagged vote	–	–	−0.17	0.05
Appointed	–	–	2.07	1.46
Candidate quality differential	1.99	0.18	1.34	0.18
Campaign spending	0.22	0.00	0.23	0.01
Party of president	0.71	0.60	−0.25	0.58
Real GDP growth	0.05	0.26	0.42	0.25
Presidential approval	0.13	0.03	0.10	0.03
Midterm	−3.41	0.66	−2.99	0.63
Error terms:				
State	2.34		0.96	
Year	1.16		0.61	
Residual	5.74		5.53	
N	6351		558	
Deviance	40393.4		3499.8	

Table 2

Predicted Party Division of Seats for the House and Senate

	HOUSE	SENATE
Model 1 (Background only)		
Median predicted Democratic seats	191	47
95% prediction interval	[180, 202]	[43, 51]
Estimated probability of Democratic control	<1%	14%
Model 2 (model 1 + candidate characteristics)		
Median predicted Democratic seats	190	48
95% prediction interval	[181, 199]	[45, 51]
Estimated probability of Democratic control	<1%	12%
Model 3 (model 2 + polling update)		
Median predicted Democratic seats	–	50
95% prediction interval	–	[48, 52]
Estimated probability of Democratic control	–	51%

Note: Model 1 includes all the structural variables except the candidate quality differential and campaign spending. Model 2 includes those variables as well, and model 3 updates the model 2 forecast in a Bayesian fashion using the available polling data. See text for details. Note that, because Democratic vice president Joe Biden can serve as a tie-breaking vote in the Senate, the Democrats would retain control of that chamber with 50 seats.

both models are also favorable to the Republicans, giving the Democrats only about a one-in-five chance of retaining control.

The most valuable additional information that can be systematically incorporated into our predictions is polling data from individual elections, which are only available for the Senate contests. To do this we first compute the Democratic share of major-party vote intentions in each poll. Then, race by race, we estimate the current Democratic poll share and its standard error based on a local regression smoother (loess) with a bandwidth of 1.5. Standard errors for the predictions become progressively larger as a function of the variance in the polls themselves and the number of days since the last poll. In addition, to ensure that we properly account for uncertainty in these polls, we also bootstrapped 1,000 loess models based on the means and sample sizes of each poll. This sampling variance was then added to the loess model variance to obtain a total variance for the polling estimate.

The last step is to combine the estimated poll standing and predicted vote shares. Following Jackman (2004), we treat the structural prediction as a Bayesian prior, which is then updated with information from the polls. The weight attached to each component—the structural forecast and the polling estimate—is a function of their relative variances as well as a scaling parameter that inflates or deflates the variance of the structural forecast based on our confidence in this prior.[6] Simply put, estimates with less variance receive greater weight.

The prediction that includes the estimated poll standing (model 3) diverges from the prediction from the structural model alone. This has not always been the case throughout this election cycle. On the Senate side in particular, there have been close contests in an unusually large number of states (Arkansas, Colorado, Georgia, Kentucky, Louisiana, Michigan, and North Carolina), combined with a relatively narrow Democratic margin of control in the Senate itself. That has made our aggregate forecast particularly responsive to shifts in polling, as even a small change in a few of these races can produce large shifts in the predicted probability of a Democratic majority. At the moment, the polls suggest that Democratic candidates in some key races are outperforming the structural forecast. Incorporating the polls improves the chances of a Democratic Senate majority, at least as of this writing.

At this point, our forecasts give the Democrats virtually no chance (<1%) of taking control of the House but an even chance (51%) of holding the Senate. The median of our 1,000 House simulations gives the Republicans 245 seats, for a gain of 11 seats. The 95% prediction interval is 236 to 254 Republican seats. For the Senate our current estimate is for the Republicans and Democrats to each control 50 seats with a 95% prediction interval of 48 to 52 seats. Of course, movements in the polls that occur after the publication of this article could shift our forecast, which we will routinely update and display on the Washington Post's ElectionLab website (http://www.washingtonpost.com/wp-tran/politics/election-lab-2014). ■

NOTES

1. For the Senate, we exclude four races that featured prominent independent or third-party candidates: Connecticut (2006), Alaska and Florida (2010), and Maine (2012).
2. We also estimated a model for House elections that used a random intercept for districts instead of states. The predictions and corresponding uncertainties were virtually identical, but the model with random district intercepts was more computationally intensive and created complications for redistricting years.
3. Uncontested races are assigned a Democratic vote share of 0% or 100% for all simulations.
4. For the Senate 64 of the 100 seats are not up for election this year. The partisan distribution of those seats is 30 Republican and 34 Democratic (counting the independents Bernie Sanders and Angus King as Democrats because that is the party with which they caucus). Because the vice president is a Democrat, the Republicans need to win at least 21 of the 36 elections to reach 51 seats and gain control of the Senate.
5. It is interesting to note that the probability of Democratic Party control in the Senate actually increases slightly in model 2 compared to model 1, despite the fact that the seat prediction does not change and the prediction interval shrinks. This is because, while the median prediction does not change in model 2, the *mean* prediction is higher, with 49 and 50 Democratic seats more likely than in model 1.
6. Based on auxiliary analysis of the 2008–2012 Senate elections, we use a scaling parameter that grows gradually over the course of the summer and reaches a maximum value of 5 by this point in the election cycle. This places most of the weight on the polls. In races with scant polling and thus a larger variance associated with the polling average, the final forecast reflects more weight on the forecast. In races without any polling, the structural forecast represents our prediction.

REFERENCES

Jacobson, Gary. 2012. *The Politics of Congressional Elections.* New York: Pearson.

Jackman, Simon. 2004. "Bayesian Analysis for Political Research." *Annual Review of Political Science* 7: 483–505.

Linzer, Drew A. 2013. "Dynamic Bayesian Forecasting of Presidential Elections in the States." *Journal of the American Statistical Association* 108 (501): 124–34.

Lauderdale, Benjamin E., and Drew A. Linzer. 2014. "Under-performing, Over-performing, or Just Performing? The Limitations of Fundamentals-based Presidential Election Forecasting." Unpublished manuscript.

Sides, John. 2014. "Four Suggestions for Making Election Forecasts Better, and Better Known." *PS: Political Science and Politics* 47 (2): 339–41.