

Abramowitz House and Senate Forecast Model

Harry Bendekgey

July 13, 2018

Alan Abramowitz proposes a simple structuralist model for house and senate prediction. To predict the net change in Republican-held seats in a midterm election year, we need three pieces of information: which party currently holds the Presidency, the generic ballot margin in early September, and the number of seats Republicans currently hold.

Each one of these predictors is heavily rooted in theory. That incumbents lose seats during the midterms is a well-established phenomenon; in fact, 1998 and 2002 were the only midterm election years since World War 2 where the incumbent party did not lose seats in the house. Using the currently held number of seats is a kind of exposure model, which observes that parties with more to lose are likely to lose. Finally, the generic ballot margin in August and September is a reflection of voter sentiment, which directly impacts seat swing.

An older version of Abramowitz's model used the Republican margin in the previous presidential election instead of a dummy variable for who won. This is less rooted in theory, but better fits the data, partly because one of the only two years the incumbent party picked up seats was 2002, and Bush lost the popular vote in 2000 which improves the consistency of the predictor. But it's important to not overfit the data; despite Trump losing the popular vote in 2016, we expect Republicans to likely lose seats this election, not gain them because they "lost" the previous Presidential vote.

Another problem is with consistency of generic ballot data. One of Abramowitz's predictors is the generic ballot margin in early September. I interpreted this to mean the average margin in polls taken 60 to 90 days before the election. In many of the earlier elections being regressed on, this takes into account only a single poll, usually by Gallup. In recent years polling have proliferated massively. Reuters/Ipsos releases a new poll every day. In order to maintain consistency, I will only take into account live polls, including those taken over the telephone by real people. I exclude polls taken online, robotically, or by IVR (Interactive Voice Response).

I start by running regression on these predictors, using data kindly shared with me by Professors Bafumi, Erikson, and Wlezien:

```
fit <- lm(chrseats ~ prevrseats + gendiff + midterm, data=model)
summary(fit)$coef
```

##		Estimate	Std. Error	t value	Pr(> t)
##	(Intercept)	105.6463	30.8847	3.421	0.0041392
##	prevrseats	-0.4468	0.1527	-2.926	0.0110518
##	gendiff	1.6780	0.3901	4.301	0.0007317
##	midterm	-17.7492	4.0558	-4.376	0.0006331

These results are very similar to those of Abramowitz, who proposes:

$$\Delta_{\text{R seats}} = 116.4 - 0.49\text{PrevRSeats} - 13.7\text{Midterm} + 1.795\text{Gendiff}$$

Given that Republicans won 241 seats in 2016, the combination of the first two terms of our regressions give the same result. The only substantive difference is that I estimate incumbency to cost a party 17.7 seats instead of 13.7, which Abramowitz finds.

I do not yet have access to data 60-90 days before the election (considering it hasn't happened yet) so I will use regression data from May-June and produce a forecast. I am using Fivethirtyeight's data, and they include a number of methods for aggregating polls, either by weighting them or adjusting them:

```
c(genpollsimple, genpolladj, genpollweight, genpolladjweight)

## [1] -8.000 -7.855 -8.005 -7.857
```

Here we can see that weighting or adjusting the polls causes a minimal shift in generic ballot average. Looking back at our regression model the differences between these values will lead to a different forecast by one-third of a seat. Thus we'll just use the simple aggregation method. We get:

```
dem_win_pct

## [1] 0.7011

interval$fit

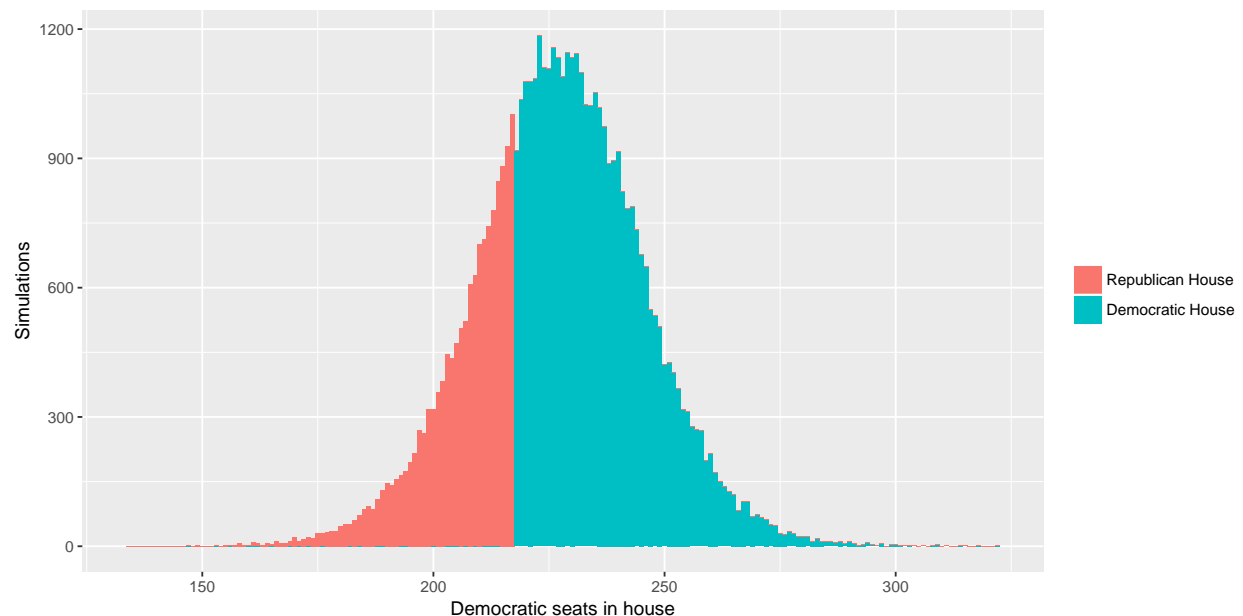
##      fit      lwr    upr
## 1 -33.19 -69.72  3.33

Abramowitz.pred

## [1] -30.26
```

We predict a 70% chance of Democrats taking control of the house. Our point estimate is that Republicans will lose 33 seats (they need to lose 24 to lose control of the chamber) with a 95% confidence interval of losing 69 seats to gaining 3 seats. This is a huge confidence interval. It is worth noting that I am using a prediction interval for the 2018 parameter values instead of just using residual standard error, because these parameters have a higher predictive error than average.

If we want to see it graphically, I take draws from a $t(\mu = -33, \sigma = 17, df = 14)$ distribution and round to the nearest whole number:



Now we move on to the Senate model. The change in Republican senate seats is predicted by the generic ballot margin, (the same one used for the house model) which President controls the white house, and the

difference in Republican seats up minus the number of Democratic seats up. Again, this is an application of an exposure model.

```
sen_fit <- lm(chrseats_sen ~ rsen_exposure + midterm + gendiff, data=model)
summary(sen_fit)$coef

##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.3726    1.19613   1.148 0.270380
## rsen_exposure  -0.4493    0.13562  -3.313 0.005126
## midterm        -2.5410    0.89104  -2.852 0.012807
## gendiff         0.2111    0.08527   2.475 0.026704
```

We note that these coefficients have large standard errors, and somewhat high p-values. If we use this model to predict what will happen in 2018:

```
dem_sen_win_pct

## [1] 0.06961

sen.interval$fit

##      fit      lwr      upr
## 1 4.782 -4.495 14.06

Abramowitz.sen.pred

## [1] 4.165
```

Again, this is a huge confidence interval, but it still only gives Democrats a 7% chance of taking the Senate. In fact, of the 26 Democratic seats up, Cook Political rates 14 of them "Solid D". Of the 9 Republican seats up, 3 are considered "Solid R." Assuming these ratings are trustworthy, that means the range of possible Republican seat changes is $[-6, 12]$. The predictive confidence interval spans the full range of those probabilities.

There are two possible takeaways here. One is that the coefficients all have large standard errors, and so with more elections (aka more data) we can become more confident about their values and the confidence interval will similarly close. Under this interpretation, the most important thing here is the point estimate, of Republicans picking up 5 seats.

The other interpretation is that this model is somewhat lacking. We see it has a relatively small R^2 , when compared to the house model:

```
summary(fit)$r.square

## [1] 0.8361

summary(sen_fit)$r.square

## [1] 0.7222
```

Perhaps this is because Senate races tend to be higher profile, and thus the candidates matter more than in house races. Perhaps this is because not all seats are up every senate election cycle, and this model fails to take into account the partisan tilt of each of the three classes of senate seats. Important pieces of information, like how many of the incumbents are in seats with a lean in the other direction, are notably missing from this model. And that makes it difficult to put too much weight into this prediction.

Graphically, we see the Senate distribution (using the same method previously but truncated by all possible results given how many seats are up for election) as:

