

Methodology

Our Senate forecasts proceed in seven distinct stages, each of which is described in detail below. For more detail on some of the terms below please see our FiveThirtyEight glossary.

Stage 1. Weighted Polling Average

Polls released into the public domain are collected together and averaged, with the components weighted on three factors:

* *Recency*. More recent polls receive a higher weight. The formula for discounting older polling is based on an exponential decay formula, with the premium on newness increasing the closer the forecast is made to the election. In addition, when the same polling firm has released multiple polls of a particular race, polls other than its most recent one receive an additional discount. (We do not, however, simply discard an older poll simply because a firm has come out with a newer one in the same race.)

* *Sample size*. Polls with larger sample sizes receive higher weights. (Note: no sample size can make up for poor methodology. Our model accounts for diminishing returns as sample size increases, especially for less reliable pollsters.)

* *Pollster rating*. Lastly, each survey is rated based on the past accuracy of "horse race" polls commissioned by the polling firm in elections from 1998 to the present. The procedure for calculating the pollster ratings is described at length here, and the most recent set of pollster ratings can be found here. All else being equal, polling organizations that, like The New York Times, have staff that belong to The American Association for Public Opinion Research (A.A.P.O.R.), or that have committed to the disclosure and transparency standards advanced by the National Council on Public Polls, receive higher ratings, as we have found that membership in one of these organizations is a positive predictor of the accuracy of a firm's polling on a going-forward basis

The procedure for combining these three factors is modestly complex, and is described in more detail here. But, in general, the weight assigned to a poll is designed to be proportional to the predictive power that it should have in anticipating the results of upcoming elections. Note that it is quite common for a particular survey from a mediocre pollster to receive a higher weight than one from a strong pollster, if its poll happens to be significantly more recent or if it uses a significantly larger sample size.

Certain types of polls are not assigned a weight at all, but are instead dropped from consideration entirely, and not used in FiveThirtyEight's forecasts nor listed in its polling database. from the firms Strategic Vision and Research 2000, which have been accused – with compelling statistical evidence in each case – of having fabricated some or all of their polling, are excluded. So are interactive (Internet) polls conducted by the firm Zogby, which are associated with by far the worst pollster rating, and which probably should not be considered scientific polls, as their sample consists of volunteers who sign up to take their polls, rather than a randomly-derived sample. (Traditional telephone polls conducted by Zogby *are* included in the averages, as are Internet polls from firms other than Zogby.)

Polls are also excluded from the Senate model if they are deemed to meet FiveThirtyEight's definition of being "partisan." FiveThirtyEight's definition of a partisan poll is quite narrow, and is limited to polls conducted on behalf of political candidates, campaign committees, political parties, registered PACs, or registered 527 groups. We do not exclude polls simply because the pollster happens to be a Democrat or a Republican, because the pollster has conducted polling for Democratic or Republican candidate in the past, or because the media organization it is polling for is deemed to be liberal or conservative. The designation is based on who the poll was conducted *for*, and not who conducted it. Note, however, that there are other protections in place (see Stage 2) if a polling firm produces consistently biased results.

Stage 2. Adjusted Polling Average

After the weighted polling average is calculated, it is subject to three additional types of adjustments.

* *The trendline adjustment*. An estimate of the overall momentum in the national political environment is determined based on a detailed evaluation of trends within generic congressional ballot polling. (The procedure, which was adopted from our Presidential forecasting model, is described at more length here.) The idea behind the adjustment is that, to the extent that out-of-date polls are used at all in the model (because of a lack of more recent polling, for example), we do not simply assume that they reflect the present state of the race. For example, if the Democrats have lost 5 points on the generic ballot since the last time a state was polled, the model assumes, in the absence of other evidence, that they have lost 5 points in that state as well. In practice, the trendline adjustment is designed to be fairly gentle, and so it has relatively little effect unless there has been especially sharp change in the national environment or if the polling in a particular state is especially out-of-date.

* *The house effects adjustment*. Sometimes, polls from a particular polling firm tend consistently to be more favorable toward one or the other political party. Polls from the firm Rasmussen Reports, for example, have shown results that are about 2 points more favorable to the Republican candidate than average during this election cycle. It is not necessarily correct to equate a house effect with "bias" – there have been certain past elections in which pollsters with large house effects proved to be more accurate than pollsters without them – and systematic differences in polling may result from a whole host of methodological factors unrelated to political bias. This nevertheless may be quite useful to account for: Rasmussen showing a Republican with a 1-point lead in a particular state might be equivalent to a Democratic-leaning pollster showing a 4-point lead for the Democrat in the same state. The procedure for calculating the house effects adjustment is described in more detail here. A key aspect of the house effects adjustment is that a firm is not rewarded by the model simply because it happens to produce more polling than others; the adjustment is calibrated based on what the highest-quality polling firms are saying about the race.

* *The likely voter adjustment*. Throughout the course of an election year, polls may be conducted among a variety of population samples. Some survey all American adults, some survey only registered voters, and others are based on responses from respondents deemed to be "likely voters," as determined based on past voting behavior or present voting intentions. Sometimes, there are predictable differences between likely voter and registered voter polls. In 2010, for instance, polls of likely voters are about 4 points more favorable to the Republican candidate, on average, than those of registered voters, perhaps reflecting enthusiasm among Republican voters. And surveys conducted among likely voters are about 7 points more favorable to the Republican than those conducted among all adults, whether registered to vote or not.

By the end of the election cycle, the majority of pollsters employ a likely voter model of some kind. Additionally, there is evidence that likely voter polls are more accurate, especially in Congressional elections. Therefore, polls of registered voters (or adults) are adjusted to be equivalent to likely voter polls; the magnitude of the adjustment is based on a regression analysis of the differences between registered voter polls and likely

voter polls throughout the polling database, holding other factors like the identity of the pollster constant.

Step 3: FiveThirtyEight Regression

In spite of the several steps that we undertake to improve the reliability of the polling data, sometimes there just isn't very much good polling in a race, or all of the polling may tend to be biased in one direction or another. (As often as not, when one poll winds up on the wrong side of a race, so do most of the others). In addition, we have found that electoral forecasts can be improved when polling is supplemented by other types of information about the candidates and the contest. Therefore, we augment the polling average by using a linear regression analysis that attempts to predict the candidates' standing according to several non-poll factors:

- A state's Partisan Voting Index
- The composition of party identification in the state's electorate (as determined through Gallup polling)
- The sum of individual contributions received by each candidate as of the last F.E.C. reporting period (this variable is omitted if one or both candidates are new to the race and have yet to complete an FEC filing period)
- Incumbency status
- For incumbent Senators, an average of recent approval and favorability ratings
- A variable representing stature, based on the highest elected office that the candidate has held. It takes on the value of 3 for candidates who have been Senators or Governors in the past; 2 for U.S. Representatives, statewide officeholders like Attorneys General, and mayors of cities of at least 300,000 persons; 1 for state senators, state representatives, and other material elected officeholders (like county commissioners or mayors of small cities), and 0 for candidates who have not held a material elected office before.

Variables are dropped from the analysis if they are not statistically significant at the 90 percent confidence threshold.

Step 4: FiveThirtyEight Snapshot

This is the most straightforward step: the adjusted polling average and the regression are combined into a 'snapshot' that provides the most comprehensive evaluation of the candidates' electoral standing at the present time. This is accomplished by treating the regression result as though it were a poll: in fact, it is assigned a poll weight equal to a poll of average quality (typically around 0.60) and re-combined with the other polls of the state.

If there are several good polls in race, the regression result will be just one of many such "polls", and will have relatively little impact on the forecast. But in cases where there are just one or two polls, it can be more influential. The regression analysis can also be used to provide a crude forecast of races in which there is no polling at all, although with a high margin of error.

Step 5. Election Day projection

It is not necessarily the case, however, that the current standing of the candidates – as captured by the snapshot — represents the most accurate forecast of where they will finish on Election Day. (This is one of the areas in which we've done a significant amount of work in transitioning FiveThirtyEight's forecast model to The Times.) For instance, large polling leads have a systematic tendency to diminish in races with a large number of undecided voters, especially early in an election cycle. A lead of 48 percent to 25 percent with a high number of undecided voters, for example, will more often than not decrease as Election Day approaches. Under other circumstances (such an incumbent who is leading a race in which there are few undecided voters), a candidate's lead might actually be expected to expand slightly.

Separate equations are used for incumbent and open-seat races, the formula for the former being somewhat more aggressive. There are certain circumstances in which an incumbent might actually be a slight underdog to retain a seat despite of having a narrow polling lead — for instance, if there are a large number of undecided voters — although this tendency can sometimes be overstated.

Implicit in this process is distributing the undecided vote; thus, the combined result for the Democratic and the Republican candidate will usually reflect close to 100 percent of the vote, although a small reservoir is reserved for independent candidates in races where they are on the ballot. In races featuring three or more *viable* candidates (that is, three candidates with a tangible chance of winning the lection), however, such as the Florida Senate election in 2010, there is little empirical basis on which to make a "creative" vote allocation, and so the undecided voters are simply divided evenly among the three candidates.

Step 6. Error analysis

Just as important as estimating the most likely finish of the two candidates is determining the degree of uncertainty intrinsic to the forecast.

For a variety of reasons, the magnitude of error associated with elections outcomes is higher than what pollsters usually report. For instance, in polls of Senate elections since 1998 conducted in the final three weeks of the campaign, the average error in predicting the margin between the two candidates has been about 5 points, which would translate into a roughly 6-point margin of error. This may be twice as high as the 3- or 4-percent margins of error that pollsters typically report, which reflects only sample variance, but not other ambiguities inherent to polling. Combining polls together may diminish this margin of error, but their errors are sometimes correlated, and they are nevertheless not as accurate as their margins-of-error would imply.

Instead of relying on any sort of theoretical calculation of the margin of error, therefore, we instead model it directly based on the past performance of our forecasting model in Senatorial elections since 1998. Our analysis has found that certain factors are predictably associated with a greater degree of uncertainty. For instance:

- The error is higher in races with fewer polls
- The error is higher in races where the polls disagree with one another.
- The error is higher when there are a larger number of undecided voters.
- The error is higher when the margin between the two candidates is lopsided.
- The error is higher the further one is from Election Day.

Depending on the mixture of these circumstances, a lead that is quite safe under certain conditions may be quite vulnerable in others. Our goal is simply to model the error explicitly, rather than to take a one-size-fits-all approach.

Step 7. Simulation.

Knowing the mean forecast for the margin between the two candidates, and the standard error associated with it, suffices mathematically to provide a probabilistic assessment of the outcome of *any one given race*. For instance, a candidate with a 7-point lead, in a race where the standard error on the forecast estimate is 5 points, will win her race 92 percent of the time.

However, this is not the only piece of information that we are interested in. Instead, we might want to know how the results of particular Senate contests are related to one another, in order to determine for example the likelihood of a party gaining a majority, or a supermajority.

Therefore, the error associated with a forecast is decomposed into local and national components by means of a

sum-of-squares formula. For Congressional elections, the 'national' component of the error is derived from a historical analysis of generic ballot polls: how accurately the generic ballot forecasts election outcomes, and how much the generic ballot changes between Election Day and the period before Election Day. The local component of the error is then assumed to be the residual of the national error from the sum-of-squares formula, i.e.:

$$Local\ Error = \sqrt{Total\ Error^2 - National\ Error^2}$$

The local and national components of the error calculation are then randomly generated (according to a normal distribution) over the course of 100,000 simulation runs. In each simulation run, the degree of national movement is assumed to be the same for all candidates: for instance, all the Republican candidates might receive a 3-point bonus in one simulation, or all the Democrats a 4-point bonus in another. The local error component, meanwhile, is calculated separately for each individual candidate or state. In this way, we avoid the misleading assumption that the results of each election are uncorrelated with one another.

A final step in calculating the error is in randomly assigning a small percentage of the vote to minor-party candidates, which is assumed to follow a gamma distribution.

A separate process is followed where three or more candidates are deemed by FiveThirtyEight to be viable in a particular race, which simulates exchanges of voting preferences between each pairing of candidates. This process is structured such that the margins of error associated with multi-candidate races are assumed to be quite high, as there is evidence that such races are quite volatile.

-
-
-

Search This Blog

Search

All NYTimes.com Blogs »

- **Follow This Blog**

- Twitter
- RSS

Reads and Reactions

- **PEWRESEARCH.ORG**

  Twitter Reaction to Events Often at Odds with Public Opinion

  Pew finds that Twitter is sometimes to the left and sometimes to the right of public opinion as measured in polling. - Micah Cohen

- **WASHINGTONPOST.COM**