

# Predicting and Dissecting the Seats-Votes Curve in the 2006 U.S. House Election\*

Jonathan P. Kastellec<sup>†</sup>      Andrew Gelman<sup>‡</sup>      Jamie P. Chandler<sup>§</sup>

January 3, 2007

## Abstract

The Democrats' victory in the 2006 election has been compared to the Republicans' in 2004. But the Democrats actually did a lot better in terms of the vote. The Democrats received 54.8% of the average district vote for the two parties in 2006, whereas the Republicans only averaged 51.6% in 1994. The 2006 outcome for the Democrats is comparable to their typical vote shares as the majority party in the decades preceding the 1994 realignment. Nevertheless, the size of the Democrats' victory in the 2006 House elections has obscured the sizable structural disadvantages they faced heading into the elections. In this paper we document the advantages the Republicans had, examine how and to what extent the Democrats overcame it, and offer predictions as to whether the results of the 2006 election leveled the electoral playing field for 2008. Our calculations showed that the Democrats needed at least 52% of the vote to have an even chance of taking control of the House of Representatives. More formally, prior to the election we estimated the seats-votes curve for 2006 by constructing a model to predict the 2006 election from 2004, and then validating the method by applying it to previous elections (predicting 2004 from 2002, and so forth). We found that the Democrats in 2006 were always destined to receive fewer seats than their corresponding average vote share. They were able to gain control of the House by winning the largest average district vote by either party since 1990. Has the 2006 election removed the Republicans' structural advantages? While Republicans continue to win more close races, a preliminary analysis of the 2008 election suggests that the switch in incumbency advantage from the Republicans to the Democrats may nevertheless level the electoral playing field.

---

\*We thank Gary Jacobson and Walt Borges for generously sharing their data, Robert Erikson, Eduardo Leoni, Robert Shapiro, David Epstein, and two anonymous reviewers for helpful comments, and the National Science Foundation for financial support.

<sup>†</sup>Department of Political Science, Columbia University, New York, [jpk2004@columbia.edu](mailto:jpk2004@columbia.edu)

<sup>‡</sup>Department of Statistics and Department of Political Science, Columbia University, New York, [gelman@stat.columbia.edu](mailto:gelman@stat.columbia.edu), [www.stat.columbia.edu/~gelman/](http://www.stat.columbia.edu/~gelman/)

<sup>§</sup>Department of Political Science, City University of New York, Graduate School and University Center, [jchandler@gc.cuny.edu](mailto:jchandler@gc.cuny.edu)

# 1 Introduction

As the 2006 midterm elections approached, pollsters, scholars, and journalists attempted to predict whether the Democrats would take back the House or Senate, or both. Much media attention was paid to President Bush's declining popularity and the public's dissatisfaction with the Republican-controlled Congress (for example, see Cook 2006). With most attention paid to the immediate political dynamics of the campaign, less noticed (though not entirely neglected) was the fact that the Democrats faced a significant structural disadvantage in their effort to retake both houses of Congress. The Democrats faced an uphill battle to control the Senate simply due to the small number of seats that were seriously contested. Their hurdle to taking over the House was subtler but perhaps just as high: as we show here, they needed to win well over half the vote share in order to have an even chance of winning 50% of the seats, thereby overcoming a structural advantage enjoyed by Republicans leading up to Election Day.

The Democrats were seeking to end 12 years of Republican control of the House. As Erikson (2006) notes, control of Congress is not simply a function of the preferences of the electorate; factors such as geography, incumbency advantage, and partisan bias in the seats-votes curve influence which party controls Congress, and for how long. The Democrats in 2006 had to overcome both a built-in incumbency advantage for the Republicans and a significant partisan bias that had emerged in favor of the Republicans following their takeover of the House in 1994. In fact, in four of the five elections between 1996 and 2004, Democratic candidates won more votes on average than their Republican opponents, yet were unable to retake the House.<sup>1</sup> This bias is consistent with the well-known tendency that the party in power tends to win a greater share of seats than votes (Tuftes 1973). Figure 1 illustrates this phenomenon, depicting the average district vote for Democrats as well as their share of House seats from 1946-2006. The Democrats enjoyed a large partisan bias during their 40-year reign over the House. The bias switched to the Republicans following their victory in 1994, and they enjoyed that advantage leading up to the 2006 elections.

In this paper we document this partisan bias, examine how and to what extent the Democrats overcame it, and offer predictions as to whether the results of the 2006 election leveled the elec-

---

<sup>1</sup>A similar dynamic has existed in Senate elections: in the 2000, 2002 and 2004 elections (during which 100 senators were elected), Democratic candidates earned a combined three million more votes than Republican candidates, yet the Republicans retained control of the chamber.

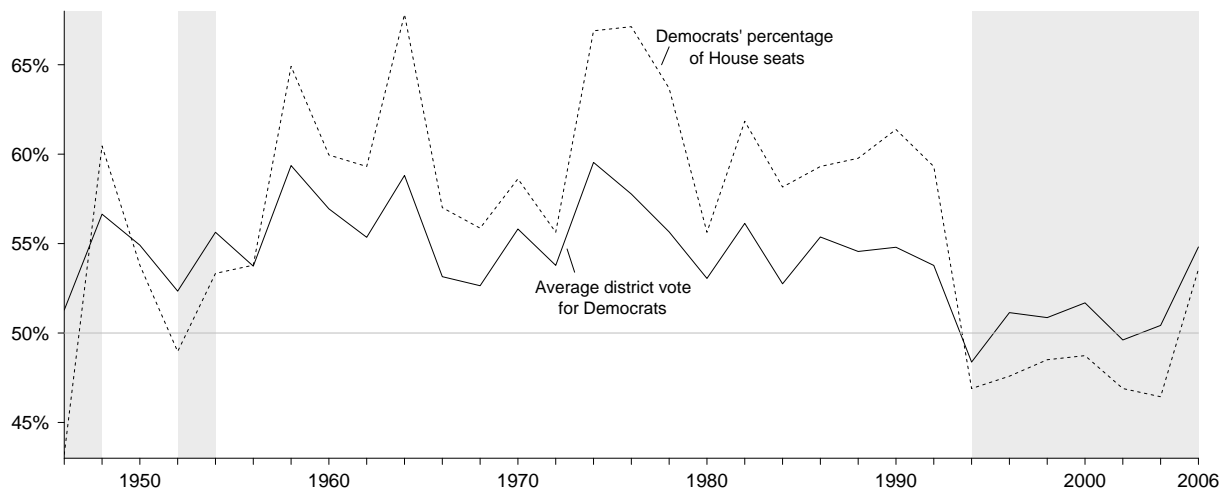


Figure 1: Average district vote versus percentage of House seats for Democrats, 1946–2006, with shaded areas representing periods of Republican control. The solid black line depicts the average district vote for Democrats in each House election from 1946 to 2006. The dotted black line depicts the Democrats' share of House seats in the same period. The gray line is a reference for 50% of the vote. The graph illustrates how the party in control of Congress tends to win a greater share of seats than votes, as had the Republicans in every election following their victory in 1994.

toral playing field for 2008—or perhaps even tilted it in the Democrats' favor. Before the election, we predicted the seats-votes curve using a district-level model of elections to translate the average district vote for Democrats into the number of seats they would gain (or lose) in the House. This strategy allowed us to predict systematically the outcome in each district across a range of hypothetical swings in the average district vote from 2004 to 2006. Our prediction of the translation between votes and seats was largely on the mark: The Democrats earned 54.8% of the average district vote, which our model predicted would lead to them winning 53.1% of the seats, or 231—just two short of their actual total of 233.

More generally, we found that in the range of plausible national election swings, the Democrats were always destined to receive fewer seats than their corresponding average vote share. They were able to gain control of the House by winning a commanding 55% of the average district vote, the largest margin for either party since 1990. Had their vote share been just two percent less, on average, a significant probability existed that the Republicans would have retained control of the House.

Did the Democrats gain in 2006 eliminate the partisan bias in favor of the Republicans? Analysis of each party's vote share among their winning candidates suggests that the Republicans still hold a natural advantage in that their winning margins are smaller than the Democrats, meaning they

“waste” fewer votes. We use our model to make predictions about the 2008 elections, and find that the switch in incumbency advantage from the Republicans to the Democrats may nevertheless level the electoral playing field.

## 2 Predicting House Elections Nationally or by Districts

Forecasts of congressional elections typically focus on the production of a point prediction: the party out of power will pick up  $x$  seats, meaning they either will or will not control the House or Senate or both in the next Congress. The 2006 campaign witnessed no shortage of such predictions, with scholars basing their forecasts on such variables as presidential approval, economic conditions, and the generic congressional ballot (Abramowitz 2006, Bafumi, Erikson and Wlezien 2006, Brandt and Brunell 2006, Campbell 2006, Cuzán and Bundrick 2006, Klarner and Buchanan 2006).<sup>2</sup> For the most part, these predictions were based on inputting national-level predictors into regression-based models in order to produce a predicted seats outcome.<sup>3</sup>

The use of macro-level variables to predict either the national vote or the seat division has a long lineage in the forecasting literature (see, e.g., Tufte 1975, Lewis-Beck and Rice 1984, Oppenheimer, Stimson and Waterman 1986, Marra and Ostrom 1989, Abramowitz, Cover and Norporth 1986, Erikson and Sigelman 1995, *inter alia*). And to be sure, macro-level predictions have their advantages. The point predictions that result from them are intuitive and lead to a dichotomous answer—which party will control the House—that is of interest to scholars, politicians, the media, and the public alike. It is also much easier to collect data at a single level of analysis per election, allowing one to ignore idiosyncrasies at the district level.

The district level, however, incorporates a wealth of information that macro-level predictions throw away by definition. Variables such as the predisposition of voters for one party or the other, incumbency, and whether a race is uncontested play crucial elements in determining the vote division in each race. This, in turn, determines who wins each race, which in turn affects the ultimate seat division. Thus, predicting the vote at the district level is true to the actual process by which

---

<sup>2</sup>The point predictions among these forecasts ranged from a low of Democrats winning 216 seats (Campbell 2006) to a high of them winning 235 seats (Bafumi, Erikson and Wlezien 2006). The latter, which is similar to our approach in that the authors estimate a seats-vote curve based on the generic congressional ballot, came closest to the actual mark of 233.

<sup>3</sup>Two exceptions are Bafumi, Erikson and Wlezien (2006) and Klarner and Buchanan (2006), who incorporate both district-level and national-level information.

seats are aggregated. It is also the only way to estimate a seats-votes curve for a single election. We employ this estimate to study the level of partisan bias that favored the Republicans in 2006 and how the Democrats overcame that bias.

### **3 Predicting the Seats-Votes Curve for a Single Election**

We draw our prediction of the seats-votes curve in 2006 from a statistical model based on the framework for evaluating legislative elections advanced in Gelman and King (1994). The strength of this approach lies in its ability to forecast elections under a range of possible conditions through the use of simulation analysis that incorporates both a range of possible national swings and the predictive uncertainty inherent in any forecast. This incorporation of uncertainty allows us to make probabilistic predictions about the Democrats' chances of retaking the House. Finally, the approach is also quite parsimonious, employing as predictors only the vote in the prior election in each district, whether the incumbent is running for reelection and whether the race is contested or not.

#### **3.1 Simulating Election Outcomes Under a Range of Hypothesized National Swings**

Our method can be summarized informally as follows. (A formal description, along with details about variable coding, appears below.) We begin with the consideration that there is a range of possible national swings in the average district vote from one election year to the next. A swing in the direction of the Democratic party will help the Democrats pick up more seats; more specifically, every possible swing will have different implications in more closely contested elections and will lead to a different translation of seats for the Democrats. Our general strategy is to allow the swing to vary across a range of values, and then predict the outcome of every district at each interval in the range. Aggregating these results produces an estimated seats-votes curve.

The partisan swing, however, is unlikely to affect each district in the same manner. We account for district-level heterogeneity by incorporating incumbency and the district-level vote from the previous election, which serves as a proxy for the district's overall predisposition to support the Democratic candidate.<sup>4</sup> Within each interval of the possible national swing, we first use the results

---

<sup>4</sup>For all analyses that appear in this paper, we impute the district vote for uncontested races to 0.75 for Democratic winners and 0.25 for Republican winners. These values are derived in King and Gelman (1991) and Gelman and King (1994), based on the vote shares received in the last election before a district became uncontested and the first election

of the previous House election to estimate each district’s baseline support for the Democratic candidate. We do so by reducing the observed vote total in each district in which an incumbent ran in the previous election by the estimated advantage of incumbency; this measure can be thought of as the “normal vote” for the previous election. We then shrink the normal vote toward 50% to represent that some of this vote arises from election-specific events, and we are interested in the part that continues on to the next election. The next step is to add back in the estimated incumbency advantage to the normal vote for incumbent candidates in the current election. This procedure results in a hypothetical proportion of the Democratic share of the two-party vote in each district.

To account for uncertainty in the prediction, we simulate the outcome in each district 1,000 times, adding independent normally-distributed errors whose standard deviation is based on the residual error from historical House election forecasts. Finally, for each simulation, we shift the entire election result (that is, add or subtract a constant from all the districts) so that the average district vote matches up to that of the previous election. The result is 1,000 hypothetical election outcomes, each corresponding to a zero national swing (but various district-by-district differences). The next task is to add the national swing to each set of district-level results. We do so in the form of adding a constant to the predicted value within each simulation; we set the constant to a value that ensures that the range of national swings is centered around the average district vote from the previous election.

For every interval in the national swing (in the entire range of average district vote), we thus have 1000 predicted outcomes in each congressional district. We summarize each simulation in two ways. First, we compute the average district vote; second, we compute the proportion of House seats won by the Democrats. Taking the mean seat proportion over all the simulations yields an expected seat share for the Democrats.<sup>5</sup> We then simply “connect the dots” of Democratic share in each interval, producing a predicted seats-votes curve.

Formally, we model the vote in each district as follows:

$$v_i^t = \alpha^t + \rho^t(v_i^{t-1} - \psi^t I_i^{t-1}) + \psi I_i^t + \delta^t + \epsilon_i, \quad \epsilon_i \sim N(0, \sigma^2), \quad (1)$$

---

after a district became uncontested. The average of these values was about 0.75 for the incumbent party and represents the average “effective support” for the party in uncontested races. For a measure based on the presidential vote in congressional districts, see Erikson (2006).

<sup>5</sup>Similar results are obtained by using the median instead of the mean to summarize predictive uncertainty.

where  $v$  = the Democratic share of the two-party vote (as a number between 0 and 1),  $i$  indexes districts,  $t$  represents the current election (that in which the seats-votes curve is being predicted) and  $t - 1$  the previous election. The incumbency variable (defined in each district in each election year) is  $I = 1$  if the incumbent in the district is a Democrat and seeking reelection,  $-1$  if the incumbent is a Republican and seeking reelection, and  $0$  if the seat is open. For values of  $\psi$  (the estimated incumbency advantage), we use the trend of the values estimated in Gelman and Huang (in press), who find a steadily increasing incumbency advantage from the 1940s to 1980, and a constant value of around 8% ever since.<sup>6</sup> The coefficient  $\rho$  captures the autocorrelation of votes between elections.<sup>7</sup> The national swing  $\delta$  is set to values ranging from  $-0.10$  to  $0.10$  (that is, considering swings of up to 10% of the vote), with the constant term  $\alpha$  set so that the national vote swing equals 0 when  $\delta = 0$ .<sup>8</sup>

For every value of  $\delta^t$ , we simulate our model 1000 times. For each simulation, we record the share of seats predicted to be won by Democratic candidates (that is,  $\frac{1}{n} \sum_{i=1}^n 1_{v_i^t > .5}$ ), where  $n$  equals the number of districts in a given election year.<sup>9</sup> We then take the average seat share measure to produce an expected seat proportion for the Democrats at the given value of  $\delta^t$ . We repeat this procedure over a range of values of  $\delta^t$ , allowing for a predicted translation between votes and seats at a plausible range of national swings.

### 3.2 Validating by Fitting to Previous Elections

To demonstrate the validity of our method, we applied it to the 19 non-redistricting election years (i.e. those not ending in “2”) from 1958 to 2004. Figure 2 plots the predicted seats-votes curve for each year, along with the actual average district vote and seats share for the Democrats. The model performs fairly well, with a root mean-squared error of about 0.02, or about 9 seats out of 435. Two elections (1960 and 1980) contribute disproportionately to the predictive error, as the model was

<sup>6</sup>Specifically, we set  $\psi = 0.08((\text{year} - 1946)/(1980 - 1946))$  if  $\text{year} < 1980$ , or  $\psi = 0.08$  if  $\text{year} \geq 1980$ .

<sup>7</sup>We obtain values of  $\rho$  by regressing  $v_i^t$  on  $v_i^{t-1}$ , along with incumbency and party indicators, for each non-redistricting election year (i.e., those not ending in “2”) from 1946–2004. For each election year analyzed, we take the average of the coefficient on the lagged vote from the previous five years. The estimates of  $\rho$  have a mean of 0.68 with a standard deviation of 0.11.

<sup>8</sup>We set  $\sigma$  to 0.066, which is the mean of the residual standard deviations from the yearly regressions discussed in the previous footnote.

<sup>9</sup>In 1958 and 1960, respectively, 436 and 437 seats were contested due to the addition of Alaska and Hawaii as states. We also dropped the handful of races featuring third-party candidates (with the exception of Bernard Sanders—an independent who held Vermont’s congressional seat from 1998–2004 and who organized with the Democratic party—whom we classified as a Democrat).

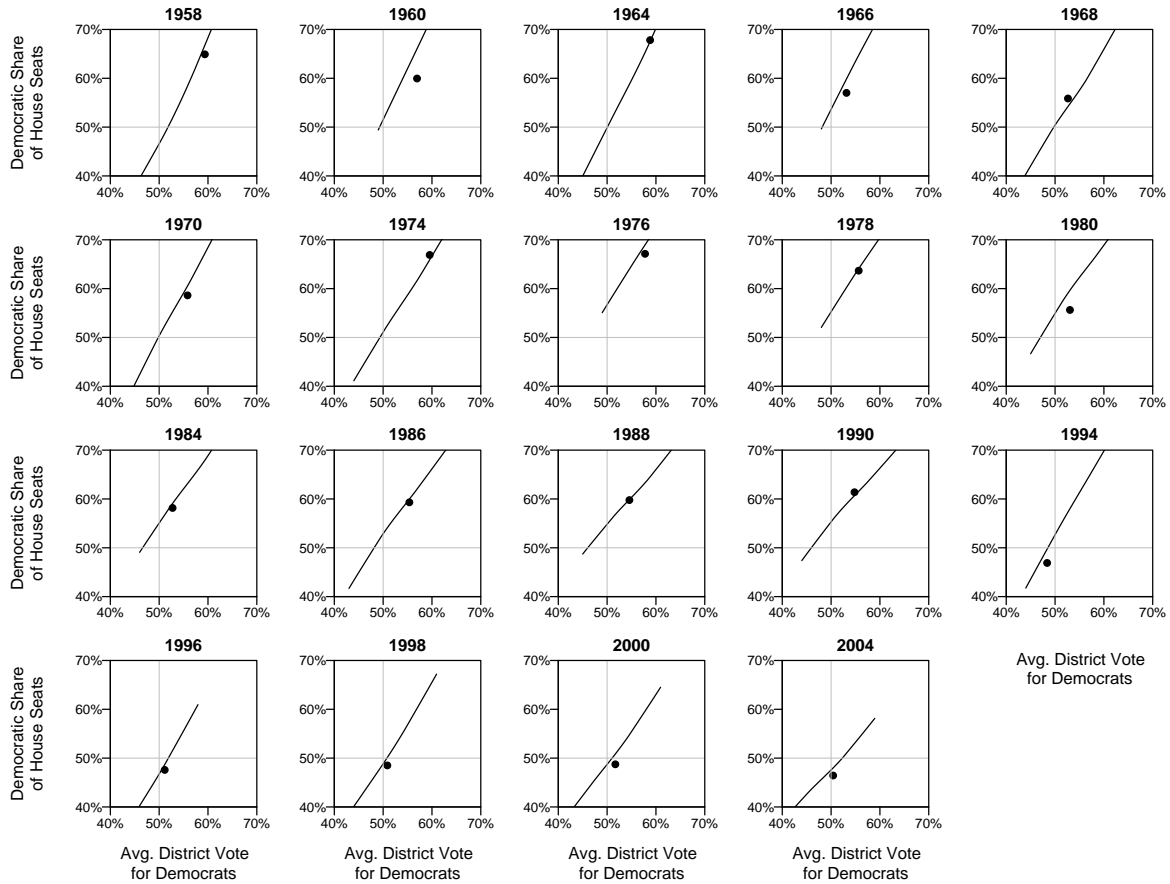


Figure 2: *Seats-votes curve prediction and actual election results, 1958–2004 (excluding redistricting years). For each year, the line depicts the estimated seats-votes curve, and the dot shows the actual election result. The light gray lines are references for 50% of the vote and 50% of the seats. As is well known (King and Gelman 1991), seats-votes curves have become less steep in recent decades—a given swing in votes will typically result in a lower swing in seats than before. As the plots show, the actual election result typically conforms closely to the predicted seats-votes curve.*

off by 26 and 19 seats, respectively, in those years. For 10 of the elections, the prediction error was less than 4 seats. The root mean squared error of the predictive standard deviation of seats given votes is 0.01, or about 4 seats, implying that the rest of the predictive error is due to year-to-year variation unexplained by the model. Thus, the model is not perfect, but we believe it demonstrates ample validity to serve as a useful tool for studying the relationship between seats and votes in the 2006 elections.



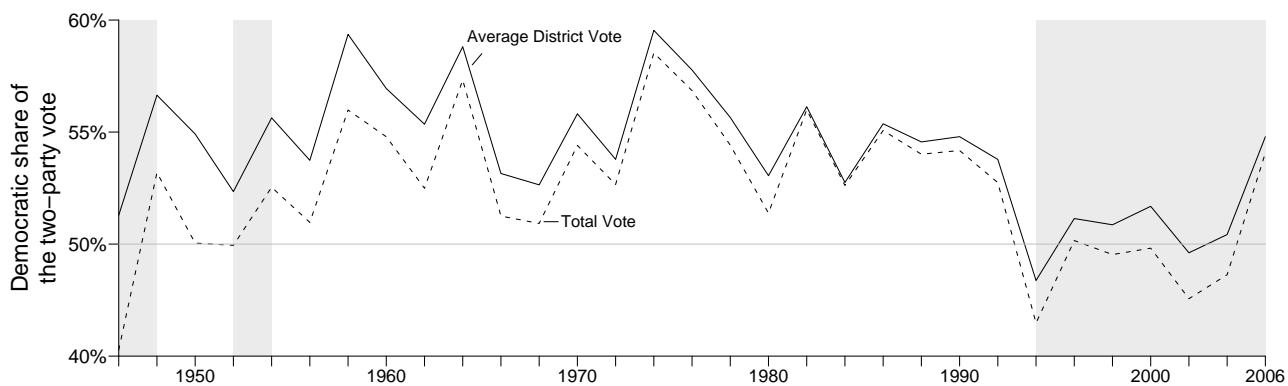


Figure 3: *Democrats’ share of the two-party total vote and average district vote in House elections, 1946–2006. Shaded areas represent periods of Republican control of the House.*

### 3.3 Average District Vote and Total Vote

In this paper we have summarized election results using the average district vote—that is, the parties’ share of the two-party vote, averaged over the 435 congressional districts: for the Democrats, this is  $\frac{1}{n} \sum_{i=1}^n v_i$ . This is the usual way that political scientists aggregate Congressional vote (King and Gelman 1991, Gelman and King 1994) and is a reasonable summary in that it is relatively immune to turnout variation and represents the national level of support for the party.

An alternative measure is the share of total vote that goes to the Democrats, which can be written as  $\sum_{i=1}^n T_i v_i / \sum_{i=1}^n T_i$ , where  $T_i$  is the turnout (the number of votes for the two major parties) in district  $i$ . Total vote differs from average district vote because of district-by-district variation in turnout, which is correlated with the parties’ vote shares.

Figure 3 shows the time series of average district vote and total vote for the Democrats over the past sixty years. In the early period, the Democrats’ average district vote was much higher than their total vote share, which can be attributed to their large number of uncontested, low turnout districts in the South. In the past fifteen years, a gap has again opened up, this time of about three percentage points, because the Democrats do better in low-turnout districts (for example, in inner cities). This trend continued in 2006: the Democrats won 54.8% of the average district vote, but only 53.3% of the total vote.

As noted above, we prefer average district vote as a summary measure of national support for the parties, and so that is what we have used in this paper.

## 4 Predicting and Dissecting The 2006 Seats-Vote Curve

Using this method, we predicted prior to the election the seats-votes curve for 2006, along with estimated probabilities of the Democrats taking control of the House given their average district vote.<sup>10</sup> Figure 4a presents the estimated seats-votes curve, along with the actual average district vote and the percentage of seats the Democrats won.<sup>11</sup> Figure 4b depicts the predicted probability of the Democrats winning the House given their average district vote.<sup>12</sup> The points respectively show the actual average district vote and seat share for the Democrats, and the predicted probability that the Democrats would win the House, based on the actual average district vote. As the seats-votes curve demonstrates, our prediction was very close to the mark. The Democrats earned 54.8% of the average district vote, an amount that we predicted would lead to them gaining 53.1% of the seats, or 231 seats. In actuality, they won 233 seats. We estimated that the Democrats had a 92% probability of winning the House based on their average district vote.

Our model confirms that the Democrats faced a stacked electoral deck heading into the 2006 elections. They needed 52% of the average district vote just to have a 50% probability of winning the House, and needed nearly 55% (which they obtained) to ensure control. To borrow Krugman's (2006) metaphor, the Democrats needed a "storm surge" to overcome the "high Republican levee" erected by the Republicans' structural advantages.

The source of the bias could be found largely in the difference between districts controlled by the two parties: some of this difference was due to geographic factors and favorable redistricting, and some to the Republicans' incumbency advantage. Figure 5 presents separate histograms of the vote shares for Democratic and Republican winners (excluding uncontested races) from each election from 1996 to 2006, along with the mean vote percentage for each. Consider first the five

---

<sup>10</sup>Information on incumbency status and retiring members was obtained from [www.cq.com](http://www.cq.com). Information on uncontested candidates was obtained from Giroux (2006)—uncontested candidates were Democrats facing no Republican challenger and vice versa.

<sup>11</sup>Vote totals for the Democratic and Republican candidates are based on official certified results from each state. As noted earlier, for each uncontested race, we imputed the vote to 0.25 for districts where no Democratic candidate ran and 0.75 for districts where no Republican candidate ran. Three districts warrant explanation. In Louisiana's 2nd district, where incumbent William Jefferson (D) won in a runoff election held in December, we used the results of the original election to measure the Democrats' vote in the district because the runoff only featured Democratic candidates. In Texas' 23rd district, incumbent Henry Bonilla (R) was defeated in a December runoff. Here we used the results from the runoff election since it was contested by both parties. Finally, we used the official certified results from the election in Florida's 13th district, which was marred by apparent voter machine malfunctions.

<sup>12</sup>We generated predicted probabilities by taking the deterministic seats prediction (the expected value from a probabilistic district-by-district prediction) at each interval of the national swing and then adding predictive uncertainty based on the root mean squared error of the historical seats votes curves presented in Figure 2.

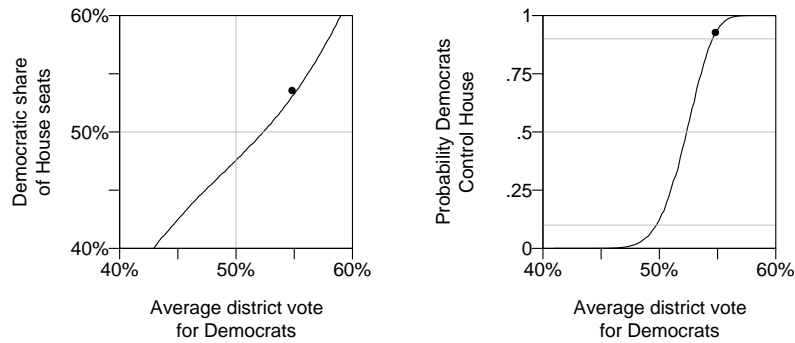


Figure 4: (a) *Predicted seats-votes curve for 2006; the point depicts the actual average district vote for Democrats (54.8%) and their share of House seats. Our point prediction was that they would win 231 seats (or 53.1% of seats) based on their average district vote of 54.8%, which turned out to be just two less than their actual total of 233 (53.6%) of seats.* (b) *The predicted probability the Democrats would control the House given their average district vote. Light lines are references for 10%, 50% and 90% probabilities. We estimate that the Democrats had a 90% probability of winning the House given the vote share they received.*

elections preceding 2006. In each year, the distribution of Democratic winner vote shares is more symmetric than that of Republicans, with a large percentage of Democratic winners earning more than 75% of the vote each year. In contrast, the distribution of Republican winner vote shares is much more skewed toward 50%, meaning they won their races by much smaller margins. In each year the mean Democratic winning vote share is also larger than the mean Republican winning vote share. Consider 2004: the Democrats won their victories with an average of 69% of the vote, while the Republicans averaged 65% in their contests, thus “wasting” fewer votes. The Republicans won 47 races with less than 60% of the vote; the Democrats only 28. Many Democrats have been in districts where they win overwhelmingly, while many Republicans have been winning the close races.

So how did the Democrats’ overcome the Republicans’ advantages in 2006? In one sense, the answer is easy: they simply obtained a large enough share of the votes to guarantee control of the House. Their average district vote of 54.8% was the largest since 1990, and three points higher than the 51.6% Republicans received in their landslide year of 1994. If the Democrats had received that same average district vote in 2006, according to our seats-votes curve the Democrats would have had only a 44% probability of winning the House. Thus, it took a large national swing for the Democrats to ensure their victory.

At the same time, the size of the Democrats’ vote gain and seat gain obscures the structural

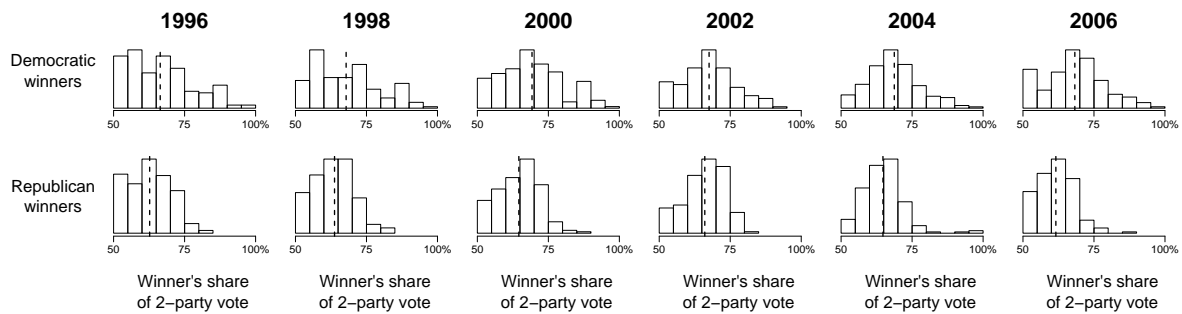


Figure 5: *Distribution of vote margins for Democrats and Republicans in winning races, 1996-2006. For each year, the top histogram depicts the distribution of the Democratic share of the two-party vote in districts where Democrats won, while the bottom histogram gives the Republican share of the two-party vote in districts won by Republicans. The dotted lines represent mean vote shares. We exclude uncontested races to avoid clustering around 75%. For all years, the distribution of vote shares for Democratic winners is much more symmetric than for Republican winners, who are more skewed toward 50%. Likewise, the mean vote share for Democratic winners is systematically lower than that of Republican winners. In these years, Republicans “wasted” fewer votes, which contributed to the partisan bias against Democrats. This trend became even more pronounced in 2006.*

disadvantages that prevented them from making even greater gains in 2006. For example, had the Republicans received 54.8% of the vote in 2006, we estimate they would have won about 249 seats, or 16 seats more than the Democrats obtained. Less obvious is that the incumbency advantage likely kept many Republicans in office. Of the 211 Republicans who ran for reelection, 189 won (or 90%); of those 211, 47 (or 25%) earned less than 58% of the vote, meaning their victory can be attributed to the incumbency advantage (recall that the estimated average incumbency advantage is eight percentage points (Gelman and Huang in press)). By contrast, all 191 Democrats incumbents who ran won; and only seven won by less than 58% of the vote. The last column in Figure 5 paints a more general picture, showing that in 2006 the Democratic victors continued their pattern of winning larger vote margins than their Republican winners. In fact, the difference between the mean vote shares of the two parties’ winning candidates exceeded those of the previous five elections. And of the 118 races where the victor won by less than 60% of the vote, Republican candidates won 78 of them (or 66%).

## 5 Looking Ahead to 2008

It thus seems possible that the size of the Democratic victory hides a continuing structural disadvantage in House elections.<sup>13</sup> Looking ahead to 2008, one advantage the Democrats will have is that with their takeover of the House, the incumbency advantage will shift to them. Will this level the electoral playing field? While it is too early to make firm predictions about the 2008 elections, we can use our model to construct a preliminary estimation of the 2008 seats-votes curve to predict whether the Democrats might still face a partisan bias. Recall that our model depends on knowing which incumbents are running for re-election; because that is unknowable at this point, for each simulation we simply assume every incumbent had a 90% probability of running.<sup>14</sup> We also assume there will be no uncontested races. For consistency, we examine the probability of the Republicans retaking the House in 2008, given their average district vote.

The results are presented in Figure 6. The results suggest that the Democrats' victory in 2006 did indeed remove the bias that favored the Republicans. Our model estimates that the Democrats will need just 49% of the vote to have a 50% probability of retaining control of the House, meaning the Republicans will need 51% to have an even chance of retaking it. Thus, while the Democrats may enjoy a partisan bias in 2008, it will likely be much less than the Republicans enjoyed in 2006. In addition, two other countervailing forces will affect the Republicans' chances in 2008. On the one hand, if our model is correct, the 51% of the average district vote they will need to have a 50% probability of retaking the House is larger than any percentage they have received since 1994. On the other, the Democrats will need to defend the seats of their first-term members who won in Republican-leaning districts in 2006, and are unlikely to receive as large an average vote share as they did in 2006. For this reason, the Democratic leadership has given these freshman favorable committee assignments in an effort to boost their incumbency advantage (Nagourney 2006).

---

<sup>13</sup>The bias the Democrats face in the Senate is even more severe due to the Republicans' geographical advantage. As Hacker and Pierson (2006, 12) point out, "[a]ssuming senators represent half their states' residents, the 49 Democrats in the [110th] Senate represent approximately 40 million more Americans than the 49 Republicans."

<sup>14</sup>In each year since 1996, no fewer than 88% of House incumbents have run for reelection.

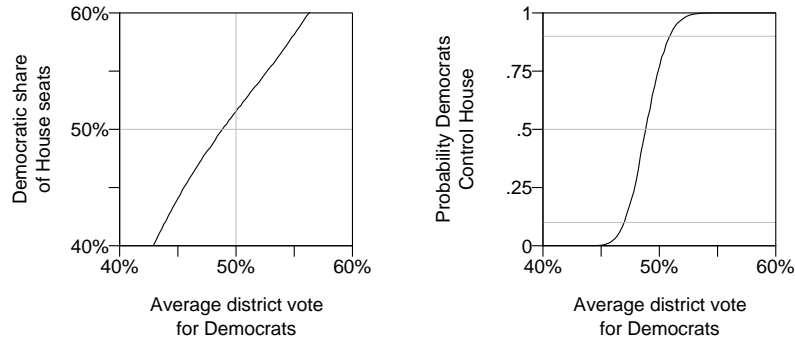


Figure 6: (a) Predicted seats-votes curve for 2008; (b) The predicted probability the Democrats will control the House given their average district vote. Light lines are references for 10%, 50% and 90% probabilities.

## 6 Conclusion

We used a parsimonious model of elections to predict and analyze the translation between votes and seats in the 2006 House elections, and to make a preliminary assessment of the 2008 election. We validated the model on the previous half-century of House elections, as shown in Figure 2. As a methodological matter, it is crucial that we both took a district-level approach and estimated a separate seats-votes curve for each election, rather than combining several election results to estimate a time-averaged curve, as is often done. Our results show that it took a large wave of Democratic support in 2006 to overcome the Republicans' structural advantages, and that the Democrats' newfound incumbency advantage will help them in their efforts to retain control of the House in 2008, while by no means guaranteeing it.

## References

- Abramowitz, Alan I. 2006. "National Conditions, Strategic Politicians and U.S. Congressional Elections: Using the Generic Vote to Forecast the 2006 House and Senate Elections." *PS: Political Science & Politics* 39(4):863–66.
- Abramowitz, Alan I., Albert D. Cover and Helmut Norpoth. 1986. "The President's Party in Midterm Elections: Going from Bad to Worse." *American Journal of Political Science* 30(3):562–76.
- Bafumi, Joseph, Robert S. Erikson and Christopher Wlezien. 2006. "Forecasting House Seats from Generic Congressional Polls." Dartmouth University technical report.
- Brandt, Patrick T. and Thomas L. Brunell. 2006. "Forecasting the 2006 House Election." University of Texas, Dallas technical report. available at <http://www.utdallas.edu/pxb054000/USHouseElectionForecast2006/HouseForecasts2006.html> [accessed 2 January 2007].
- Campbell, James E. 2006. "Forecasting the 2006 National Elections to the U.S. House of Representatives." *The Forum* 4(2). available at <http://www.bepress.com/forum/vol4/iss2/art2> [accessed 2 January 2007].
- Cook, Charlie. 2006. "The Gathering Storm." *The Cook Political Report*. August 5th. available at <http://www.cookpolitical.com/column/2006/080506.php> [accessed 2 January 2007].
- Cuzán, Alfred G. and Charles M. Bundrick. 2006. "Will the Republicans Retain Control of the House of Representatives in 2006?" University of West Florida technical report. available at <http://uwf.edu/govt/documents/WilltheRepublicansRetainControlin200610-24-06.pdf> [accessed 2 January 2007].
- Erikson, Robert E. 2006. "Party Control of the U.S. House of the Representatives." Columbia University technical report.
- Erikson, Robert S. and Lee Sigelman. 1995. "Poll-Based Forecasts of Midterm Congressional Election Outcomes: Do the Pollsters Get it Right?" *Public Opinion Quarterly* 59(4):589–605.
- Gelman, Andrew and Gary King. 1994. "A Unified Method of Evaluating Electoral Systems and Redistricting Plans." *American Journal of Political Science* 38(2):514–54.
- Gelman, Andrew and Zaiyang Huang. in press. "Estimating Incumbency Advantage and its Variation, as an Example of a Before-After Study." *Journal of the American Statistical Association*
- Giroux, Greg. 2006. "For 55 House Incumbents, 'Safe' is an Understatement." *CQPolitics.com*. October 3.
- Hacker, Jacob S. and Paul Pierson. 2006. "Enter Center." *The New Republic*. December 25. pp. 11-13.
- King, Gary and Andrew Gelman. 1991. "Systemic Consequences of Incumbency Advantage in U.S. House Elections." *American Journal of Political Science* 35(1):110–38.
- Klarner, Carl and Stan Buchanan. 2006. "Forecasting the 2006 Elections for the U.S. House of Representatives." *PS: Political Science & Politics* 39(4):857–61.

- Krugman, Paul. 2006. "Will the Levee Break?" *New York Times*. October 13. p. A27.
- Lewis-Beck, Michael S. and Tow W. Rice. 1984. "Forecasting U.S. House Elections." *Legislative Studies Quarterly* 9(3):475–86.
- Marra, Robert F. and Charles W. Ostrom, Jr. 1989. "Explaining Seat Change in the U.S. House of Representatives." *American Journal of Political Science* 33(3):541–69.
- Nagourney, Adam. 2006. "With Their Eye on 2008, Democrats Nurture Vulnerable House Freshmen." *The New York Times* December 22. p. A30.
- Oppenheimer, Bruce I., James Stimson and Richard W. Waterman. 1986. "Interpreting US Congressional Elections: The Exposure Thesis." *Legislative Studies Quarterly* 11(2):227–47.
- Tufte, Edward R. 1973. "The Relationship Between Seats and Votes in Two-Party Systems." *American Political Science Review* 67(3):540–554.
- Tufte, Edward R. 1975. "Determinants of the Outcomes of Midterm Congressional Elections." *American Political Science Review* 69(3):812–26.