

Forecasting Congressional Elections

Carl E. Klarner

Department of Political Science

Indiana State University

cklarner@isugw.indstate.edu

This article will briefly examine the history of election forecasting models, focusing on the mere two forecasts made for the 2008 congressional elections with the aid of statistical models, as well as six forecasts for the 2006 midterm elections. Particular attention is given to how the varying competitiveness of House elections should, and should not be, dealt with. How the district level forecasts of Klarner (2008) stand up against on-the-ground assessments by Pollster.com, the Cook Political Report, and the Rothenberg Report made at the same time (circa July 28th, 2008) is then assessed. The Klarner district level forecasts perform well when pitted against polls and expert judgments for the House. For the Senate, Pollster.com wins out. Last, a brief argument for how election forecasting might advance our knowledge of congressional elections is presented.

Forecasting Models

Gary Jacobson likes to tell a story in which he would go to election night parties wearing a coat with many pockets. In each pocket would be the results of a different forecasting model, with a different prediction about the number of seats that would change hands. Once the total number of seats was announced, he would pull out the requisite predictions and say “I’ve got the answer right here!” The message of the story highlights our continuing inability to accurately forecast election results. Jacobsen’s story also points to the fact that if there are numerous models, some are bound to be correct just by chance. A shortcoming of the congressional forecasting subfield is that there is no easily documented history of congressional forecasts because they are often not published. This makes assessing which models lead the pack difficult.

For the sake of this article, “election forecasting” will be defined as a model that has a point estimate for the election before the fact. Published forecasting articles before the mid-1990s almost always had “contingent” forecasts (Campbell and Garand 2000). These would often be presented as a table with a variety of scenarios where interested readers could plug in, say, the current value of presidential approval or some measure of the economy as they become available before the election. The pioneering work of Tufte (1975) is a close example of this. The problem with these contingent forecasts is that twenty years later, it is unclear what exact data went into the later forecast, or when the data was available. This makes establishing a track record for forecasts difficult.

The first time a collection of scholarly forecasts were published in academic journals *before* the election was held was in the October 1996 issue of *American Politics Quarterly* (Campbell and Garand 2000), although other academics had published their point-predictions before elections (Fair 1978). The quicker publication of government statistics as well as a faster publication process for academic journals has made this possible.

The vast majority of attempts at forecasting U.S. elections have been done for the presidency. For example, the most prominent collection of election forecasting articles is referred to as the Campbell collection, after James Campbell, first appearing in *American Political Quarterly* for the 1996 election, then appearing in the March 2001, October 2004 and October 2008 issues of *PS: Political Science and Politics*. These collections contained six, six, eight and nine articles, respectively, regarding the presidential election. In contrast, only one, zero, zero and two articles pertaining to congressional elections were included in these issues for those years. There are other venues that report election forecasts, of course, but in general these also emphasize the presidency.

Many scholars do not focus on forecasting congressional elections as the spotlight is put on the presidency. As one example, Abramowitz (2006) published his congressional forecasts for the 2006 election, and his presidential forecasts for the 2008 election (2008), but not his congressional forecasts for 2008. Also, there are fewer congressional forecasting models in midterm years than there are presidential ones in on-years. In 2006, *PS: Political Science and Politics* only had three articles forecasting congressional elections. Often notable scholars such as Gary Jacobson, Alan Abramowitz, or Stephen Ansolabehere will make forecasts for congressional elections—which they all did in 2008—but not publish these in peer-reviewed publications. The lack of congressional forecasting models has been unfortunate because it prevents scholars of forecasting to build a track record for a diversity of models over time. Still, a brief discussion of the most recent congressional election forecasting models follows.

The 2008 congressional elections saw very few published forecasting attempts from political scientists. Web sites devoted to polls or expert judgments were created and are discussed in the next section. But only two published statistical models forecasting the 2008 congressional elections were publicized. One of these was published by Lockerbie (2008) and the other by Klarner (2008). The more accurate of these two was Lockerbie's. An extended search to find descriptions of other congressional forecasting models for 2008 was unsuccessful.

Lockerbie's (2008) forecast was made 127 days before the election. Lockerbie predicted that the Democrats would pick up 25 seats in the House, while they actually picked up 20 or 21 (one race is as yet unresolved). The dependent variable in his model was the number of seats picked up by the Democrats. His predictor variables were survey respondents' prospective evaluations of the economy as well as a variable measuring how long the party of the President has been in the White House. A final predictor variable measured the number of open seats,

adjusted for which party is at advantage during the election year. This adjustment is done by multiplying the figure by “-1” if survey respondents judge it to be a good year for the Republicans, and “1” for a Democratic year. The variable takes a value of “0” if the public forecasts no clear favorite. Lockerbie theorized that open seats present the biggest opportunities for an advantaged party to pick up seats since incumbents so often win. Lockerbie did not make forecasts for the Senate in 2008.

A new type of forecasting model for congressional elections is the panel model. Panel models pool all 435 House elections across years into one large dataset. This was the model used by Klarner for the 2008 House and Senate elections. Pooling means that predictions are made not only for the overall number of seats that a party will pick up, but also what percent of the vote each party can expect in each district. Bafumi, Erikson, and Wlezien (2007) used district-level information to forecast the outcomes of House elections, although their historical analysis is not of a panel dataset. Stephen Ansolabehere also used district level information for his House model, but has not reported the details. Klarner and Buchanan also reported two such models for both the House and Senate for the 2006 elections (2006a, 2006b).

The theoretical motivation of the Klarner (2008) House and Senate models is based on variables that have been identified by scholars in the field, such as incumbency, the previous office holding experience of non-incumbents (Jacobson 2004), the partisan disposition of the district (Highton 2000), and incumbent ideological moderation (Erikson and Wright 2005). The national level factors used (midterm penalty, presidential approval, change in real per capita income, vote intention) are close to those used by Abramowitz (2006), although he uses no measure of the health of the economy.

One of the strengths of district-level models is that they take the degree of inter-party competition in each district into consideration. Jacobson (2006) argues that the decline of competitive House seats has caused the number of seats that change hands in the aggregate to decrease. As a result, Jacobson says, forecasting models that do not take this into account are lumping elections with much different contexts together. How a wave breaks across districts is dependent on how many competitive districts there are. A three percent rise in the vote for the Republicans will not help them much if there are few competitive districts. If there are many competitive districts, they can pick up a lot. A district-level model measuring the partisan disposition of districts does just this.

Overall, Klarner's 2008 House and Senate models underestimated the percent of seats that the Democrats actually picked up, as did Klarner and Buchanan's (2006a; 2006b) models of the 2006 elections. The House model forecast that the Democrats would have a total of 247 seats in the House of Representatives after the 2008 election, indicating a gain of 11 seats. Although as of December 12, 2008, one seat had not been called, the Democrats had at least 256 seats in the House after the election, and may have as many as 257, meaning the model underestimated Democratic gains by nine or ten seats. At the district level, the model called 409 out of 434 (94.2%) races correctly.¹ In 2006, Klarner's House model also underestimated Democratic gains by nine seats.

The Klarner 2008 Senate model forecast that the Democrats would win fifteen of this year's 35 Senate races, a gain of three seats for the Democrats to hold a total of 54 seats after the election (including Democratic-caucusing independents Bernie Sanders and Joseph Lieberman). The Democrats have picked up at least seven seats in the Senate and may pick up as many as eight, depending on how the Minnesota election is resolved. State-level forecasts called 29 out of

34 elections correctly (85.3%), missing the Senate races in Alaska, New Hampshire, North Carolina, Oregon, and Virginia.² Overall, the Senate model was not effective at forecasting the number of seats Democrats picked up or how the vote would go in individual states. In 2006, the Senate model also predicted a three seat Democratic gain in the Senate, and the Democrats similarly picked up seven seats in actuality. As in 2006, the Senate model performed worse than the House model.

A number of notable political scientists did make predictions for the House and Senate at a roundtable at the 2008 American Political Science Association annual meeting over Labor Day weekend. Abramowitz utilized his 2006 model (2006) to forecast a Democratic gain of thirteen seats in the House, and a gain of seven seats in the Senate for the same party. Gary Jacobson forecast that the Democrats would pick up between ten and twenty House seats, and four to seven Senate seats. Robert Erikson forecast that the Democrats would pick up fifteen seats in the House, and six seats in the Senate, although his predictions were not made with a quantitative model. Again, in 2008, the Democrats picked up 20 or 21 seats in the House and seven or eight seats in the Senate.

Pollyvote.com publicized fifteen forecasts for the 2006 House elections, seven of them purely quantitative in nature. Their success at forecasting is hard to compare because they differ radically in lead time, with the earliest made in late April (Klarner and Buchanan) and the latest made in late October (Bafumi, Erikson and Wlezien 2007). When Klarner's and Buchanan's model is excluded, 88 percent of the variation in seat prediction error is explained by a variable measuring the number of days before the election the forecasts were made. Authors of six of these models fully explained the methods by which they generated their forecasts and their work will be discussed here.

The most accurate of the 2006 models was authored by Bafumi, Erikson and Wlezien (2007) which forecast a Democratic gain of 32 seats, while actually the Democrats took 31. Their model relied on surveys asking respondents their vote intention in the upcoming election to measure the strength of the national wave. They altered their predictions at the district level on the basis of the partisan disposition of the district (measured by past voting history) and whether an incumbent was running or not.

The other five models were not driven by late poll results, and had several commonalities. From the most accurate of these models to the least, these are the models of Abramowitz (2006), (a 29 seat pickup for the Democrats), Klarner and Buchanan (2006a) (22 seats), Cuzan and Bundrick (2006) (twelve seats), Campbell (2006) (thirteen seats), and Brandt and Brunell (2006) (twelve seats). Excepting Klarner and Buchanan, all use the national election as the unit of analysis, using the aggregate number of seats that change hands between the parties, or some variant, as their dependent variables. All five models employed a variable taking the midterm penalty into account. Political scientists have still not been able to replace this dummy variable with substantive variables that explain the strong empirical regularity that the President's party has a tendency to lose seats in a midterm year. Tufte's early work (1975), cited above, used measures of both presidential approval and economic health in his model, as many forecasting models still do. Accordingly, all models save Cuzan and Bundrick used some measure of presidential approval as an independent variable. Surprisingly, both Campbell and Abramowitz did not have independent variables measuring the health of the economy, although the other three models do.

Of the five aggregate models for the House discussed above (for both 2006 and 2008) three of them took the competitiveness of districts into account in some way. Lockerbie (2008)

and Abramowitz (2006) measured the number of open seats as one way of taking into account how much “low hanging fruit” there is for the other party to pluck. The problem with this is that some open seats may be safe seats and some incumbents’ seats may be vulnerable. In fact, a very low proportion of open-seats are won by five percent or less and the proportion of open-seats that are competitive varies considerably from election to election (Jacobson 2006, 31). Most tellingly, when the proportion of seats that are competitive in an election (i.e., won by five percent or less) is regressed on the proportion of seats that are open, only nine percent of variance is explained (analysis not shown).³

Campbell (2006) utilized three different methods of dealing with the competitive district problem. Although he acknowledged that all three methods rest on problematic assumptions, he nevertheless averaged the results of the three. The assumptions underlying these methods are 1) that the number of competitive districts in the upcoming election will be the same as the average of the last two elections, 2) it will be the same as the last election, or 3) “that seat change will be estimated as a percentage of the median absolute seat change in the previous eight elections” (Campbell 2006, 6). The first and second assumptions are problematic because, as Campbell reports, only 44 and 38 percent of the variance, respectively, in the number of competitive districts in a year were explained by the last election’s number or the level in the election before that.

Forecasters who do not control for the changing level of competitiveness in congressional elections are failing to take an important aspect of congressional elections into account. But the aggregate method of forecasting has yet to offer a valid way to do so. Utilizing district level models is a promising method for doing this. Still, in both 2006 and 2008, and for both the House and Senate, the pooled models of Klarner and Buchanan (2006a, 2006b) and Klarner

(2008) understated the extent of the Democratic wave. It will be argued here that this is a problem with pooled models themselves, but a problem that can be overcome with good modeling. The solution to this problem may also improve our understanding of congressional elections.

Figure 1 shows a scatter-plot of the 2008 Democratic percent of the two-party vote forecast by Klarner (2008) for the House on the horizontal axis, and the actual vote on the vertical axis. It is evident that the model did a good job overall except in those districts in which it forecast a narrow Republican win (aside from some elections in the southeast quadrant), which were often won by Democrats. A similar scatter-plot from Klarner and Buchanan's 2006 House forecasts reveals the same pattern (not shown). Figure 2 shows the same scatter-plot as Figure 1, but for the Senate. A similar pattern holds there. There is a fairly strong relationship between forecasts and actual votes, except in districts forecast to be narrowly won by Democrats. A slight deviation from the House figure is that a number of states forecast to be safely Republican were won by narrower margins than forecast (looking at about "eight o'clock" from the origin), not to mention Alaska (looking at about "nine-o-five o'clock" from the origin).

It is reasonable to think about the forecasts as representations of how elections would go based on the normal causes of election outcomes unrelated to campaigns—incumbency, previous-office holding, and so forth. In an election year that favors one party, that party campaigns heavily in opponents' districts that are closest to swinging toward the favored party. These areas see more losses for the disfavored party—in this case going toward the Democrats. Or perhaps it is because the voters there are more moderate and thus especially influenced by the national wave. Although suggestive—see the dangers of analyzing residuals in King (1986)—the scatter-plots draw attention to a major shortcoming of district-level models. Aggregate

models only forecast the net change in seats. If massive and consequential campaign activity occurs in these battleground districts, that activity is implicitly modeled to some degree. District-level forecasting models will have to take this into account in the future to fully live up to their potential.

Last, it has been notoriously difficult to forecast Senate elections accurately (Abramowitz and Segal 1986). Many fewer attempts have been made at forecasting Senate elections than at forecasting House elections. Evidence of this was provided by Abramowitz (2006).

Abramowitz presented both House and Senate forecasting models with similar explanatory variables in each model. The dependent variable in each model measured seat change. The House and Senate models explained 87 and 65 percent of the variation in their dependent variables, respectively. Similarly, a comparison of House and Senate models in Klarner (2008) found larger uncertainty bounds for the Senate, and many more independent variables failing to achieve statistical significance. The difficulty of forecasting Senate elections is perhaps why there are so few models for the Senate. Hopefully scholars will rise to this challenge and increase their efforts to devise accurate forecasting models for Senate elections.

Comparison with Pollster.com, the Cook Political Report, and the Rothenberg Report

How congressional elections are developing is widely monitored by political junkies and political elites alike, and a number of organizations run by political notables call specific races. They do this using expert judgments and/or district or state level polls, which have become more common over time. In this section, the accuracy of the Klarner (2008) district and state level forecasts from July 28th are compared with those of Pollster.com (July 28th), the Cook Political

Report (made July 31st), and the Rothenberg Report (made July 29th). Pollster.com is a well-known site created and run by polling experts such as Mark Blumenthal and Charles Franklin.⁴

The comparisons with Klarner's forecasts has wider relevance than just assessing one forecasting model, because the factors included in the model are motivated by many political scientists' work.

In short, the Klarner forecasts of which party will win districts hold their own very well against either the polls or expert judgments made at the same time for the U.S. House.

Additionally, the Klarner model proves just as accurate—it is really a tie—as Pollster.com and the Rothenberg Report when forecasting which districts will be competitive. But all three are slightly less accurate than the Cook Political Report. For Senate forecasts, the Klarner model does moderately worse than Pollster, Cook or Rothenberg in correctly calling races. It is much worse at calling competitive states than these three organizations, especially Pollster.com.

Table 1 compares the number of districts in the November election that Pollster.com predicted correctly on July 28th with correct predictions for House seats made by the Klarner model. Because of the expense of conducting district level polls, Pollster.com had run polls in only 57 districts on that date. Here, the Klarner model actually outperforms Pollster.com, calling 77 percent (44 of 57) of the districts polled by Pollster.com correctly. Pollster.com called just 72 percent (41 of 57) correctly.

Comparison with the Rothenberg Report and the Cook report are tricky, because they do not make forecasts for “toss-up” districts. Rothenberg is agnostic on ten districts, while Cook does not call 27. Another surprising shortcoming in their scores is that they do not specifically say for which party safe districts are safe—i.e., projected to be Democratic or Republican. Of

the safe districts that Rothenberg failed to call for the Democrats or Republicans, seven turned out to be won by five percent of the vote or less (including Virginia's fifth district) while Cook does this with three. But this short-coming is ignored for the rest of the comparison.

For the 71 districts Cook called, Klarner called 85 percent correctly, while Cook called 86 percent correctly (Table 1). For the 54 districts Rothenberg does call, the Klarner forecast got 67 percent correct, while Rothenberg called 70 percent correctly. It is also interesting to note that of the 71 predictions made by Cook, the Klarner and Cook models have the same predictions for 70 of the races.⁵ However, if Cook made such forecasts for toss-up districts, the agreement would probably be lower. Similarly, Rothenberg and Klarner forecast 50 of 54 races identically. In sum, in the districts they all forecast, the political science forecasting model is evenly matched with those of Rothenberg and Cook three months before the election. Klarner's track record among districts that Cook and Rothenberg defined as "toss-ups" is also reported in Table 1.

A multiple regression was run with the actual Democratic percent of the two-party vote as the dependent variable (subtracted by 50). The following three independent variables were used: the percent of the two-party vote predicted by Klarner (subtracted by 50), the seven-point ranking system used by Cook (subtracted by 4), and the nine-point ranking system used by Rothenberg (subtracted by 5). Centering these variables around zero means the regression equation's constant can be examined for signs of biased forecasts. Every district Cook and Rothenberg define as safe is assumed to be won by the correct party (which is a generous assumption to them, as indicated above). The results of these regressions for the House are displayed in model one of Table 2. Pollster.com is not used in model one because of the small number of districts with scores, but the percent of the two-party vote predicted for the Democrats by Pollster.com is added for model two (again, centered around zero). Both the Klarner and

Cook forecasts are statistically significant ($p < .001$), while the Rothenberg forecasts are not, and actually have the wrong sign. The y-intercept in model one is statistically significant, indicating that the model has a pro-Republican bias, as is also evidenced from three bivariate regressions using each independent variable alone (analyses not shown). “Bias” of course can also include a shift in vote intention between a forecast and the election. Overall, these results present evidence that political science forecasting models have something to contribute to Washington insiders when it comes to forecasting elections.

In model two of Table 2, which examines just 58 House districts, only the Cook forecasts retain statistical significance when Pollster’s forecasts are added to the model.⁶ Model two is also statistically significant at the .05 level (one-tailed). For the few districts where polls were conducted, the Klarner forecasts did not appear to be helpful, but on the other hand, Pollster is of no help in districts without polls, either. When Pollster’s forecasts are used as the only independent variable, no bias is evident when looking at the y-intercept.

Another test of the models is to see how well these methods call competitive elections. Interested observers focus their attention on the basis of the prognostications of pundits as well as on surveys. Again, it is hard to assess what percentage range in an actual election the Cook and Rothenberg ratings mean by “toss-up.” But Cook defines 27 districts as “toss-ups,” while Rothenberg’s three middle scores (called either “toss-up” or “toss-up/leaning” districts) were applied to 28 districts. (This will give Rothenberg a slight advantage in the following analysis). The 27 races forecast by Klarner as being won by the narrowest amounts were also defined as “competitive” (which are races forecast to be won by 9.3 percent or less). The same was done for Pollster.com’s most closely forecast races (forecast to be won by 13.0 percent or less). Twenty-seven House races in 2008 were won by five percent or less and are also considered

“competitive.” The 27 most competitive districts from each forecasting method (but 28 for Rothenberg) are now each compared with the 27 districts that were actually the most competitive.

Again, the Klarner forecasts stack up favorably against the polls or expert judgments, as seen in the bottom two rows of Table 1. Klarner correctly calls eight of 27 competitive, while Pollster only gets seven. Rothenberg ties Klarner by also calling eight. Cook does slightly better than everyone else, correctly calling ten competitive races “toss-ups,” and so calling seven percent more of the actually competitive races than the runners-up.

It is probable that polls and expert judgments would outperform a statistical model as election day approaches. Idiosyncratic factors introduced by the campaign are sure to have an impact the statistical model does not measure and the polls might. However, the fact that statistical forecasting models for the House hold their own against these on-the-ground assessments is especially noteworthy because they are vastly cheaper in time and other resources.

Table 3 examines the Senate and compares the Klarner forecasts with those of Pollster.com, Cook, and Rothenberg identically (although there were no “toss-up/tilting” races in Rothenberg’s Senate typology). One caveat is that there were far more polls available by July 28th that assessed Senate races in comparison to House races, which makes Pollster.com more accurate. A less obvious implication of this is that there were a higher proportion of safe states polled than safe districts. (On average, House districts on Pollster.com were actually won by 12.6 percent, while that figure for the Senate is 15.0). Accordingly, of the 28 states for which Pollster.com had data, they called 93 percent correctly, in comparison to 82 percent correctly by Klarner. The Cook Report and Rothenberg also beat Klarner’s state forecasts by about seven

percent. For the Senate, these alternate methods perform better than Klarner, although the differences are ten percent or less.

The multiple regression for the House is duplicated for the Senate and is reported in model three of Table 2. In this model, when all three variables are put head-to-head, none achieves statistical significance at the .05 level. Forecasts by Klarner and by Cook are both statistically significant at the .10 level (one-tailed), however, while the Rothenberg forecast is far from statistical significance. When the Rothenberg variable is dropped, the Cook and Klarner variables achieve statistical significance at the .05 level (one-tailed), although further analyses (not depicted) indicate that Cook performs better. Most of the relationship between Rothenberg and the actual results are shared by Cook, but Klarner's forecasts still have something to add to Cook's.

Because most states had been polled by July 28th for the Senate, Pollster.com's forecasts exist for almost the entire sample (Table 2, model four). When it goes head-to-head against the other three types of forecasts, it beats all of them: none attain statistical significance, while it does. This is also true when any two or any one of the other three independent variables is in a regression with it (analyses not shown). The Senate polls on Pollster.com clearly outclass the other three types of forecasts. However, when all four independent variables are utilized in four bivariate regressions, the y-intercept in each model is statistically significant, indicating that all four have a pro-Republican bias (analyses not shown).

Last, it is important to examine how well forecasts focus our attention on competitive elections. One problem is that Cook calls five Senate seats "toss-ups" while Rothenberg does this for four, giving Cook a comparative advantage. As a result, the top four as well as the top

five most competitive Senate seats from Klarner's and Pollster's forecasts are compared to Rothenberg and Cook, respectively. The last four rows of Table 3 display how well the four types of forecasts do at calling competitive races. For the top four most competitive seats (those won by six percent or less), Pollster.com's model outperforms all the others. Klarner's model does the worst, with Rothenberg's in the middle. The pattern is the same for the top five most competitive races, with the Cook Report in the middle this time. Pollster.com again does better in the Senate races.

Forecasts based on variables long acknowledged to be important in House elections can still add to the ability of polls and expert judgments to forecast who will win elections or identify competitive races. In contrast, the evidence indicates that polls rule the day when it comes to forecasting Senate elections. Again, it cannot be emphasized enough how much cheaper running a forecasting model is.

What Is the Use?

Another question that presents itself with forecasting models is "what is the use?" Should we not just focus on understanding elections after they happen, especially because some of the key variables that we know have an impact on elections are not measured until after the election (Bartels and Zaller 2001), or at least not measured very well? Why should we bother at all with forecasting models?

First and foremost, forecasting highlights the importance of theory. A bad theory will lead to bad forecasts. (See Armstrong [2001, 20] for a summary of studies supporting this assertion.) Work by Cuzon and Bundrick (2008) further emphasized the importance of theory.

But Cuzon and Bundrick also pointed out the problematic effects of “letting the data speak for themselves” when assessing the impact of independent variables on the quantity to be forecast. Instead of estimating the impact of an independent variable by standard regression analysis, Cuzon and Bundrick (2008) standardized all the independent variables and then constrained all their impacts to have the *same weight* in their regression equation. The forecasts based on this method were more accurate than those made by letting ordinary least squares assign each explanatory variable’s weight. (See Cuzon and Bundrick [2008] for citations to studies outside of Political Science.) Samples can tell us a lot about a population, but we as social scientists often push inferences too far.

The institutional incentives that compel traditional political science researchers to “dredge data” are far too great for most of us to resist. In the quest for statistically significant—and more publishable—findings, specification searches are conducted by dropping and adding variables or using slightly different measures of concepts. An article by Gerber, Green and Nickerson (2001) demonstrates that such searches are biasing the conclusions of research in at least one subfield of our discipline. On the other hand, research on forecasting has found that data mining or the use of statistical significance as a criterion for including variables in a statistical model does absolutely no good in improving forecasts (Armstrong 2001). Therefore, forecasters are presented with an institutional incentive to avoid such behavior, which may advance our knowledge of political phenomena considerably.

A common complaint about forecasting models is that they are not based on extensively developed theories (Eijk 2005). Of course there is room for progress in this area. But it should be noted that many forecasters have given the theoretical rationales for their models in prior work, and do not have the space to fully present them each time they present their forecasts.

Furthermore, as argued above, forecasters have a greater incentive to develop strong theoretical models than other political scientists. That aside, it could be very productive to see the most widely distributed forecasting models published in a special collection that examines at length their theoretical underpinnings.

Another related complaint is that some forecasting models do not contribute to our theoretical understanding of congressional elections if the variables they use are too close causally to the dependent variable. The use of congressional vote intention is an example of such a variable. This is a poll question asking what party a respondent will vote for in the upcoming congressional election. When vote intention polls taken close to the election are used, they can yield very accurate predictions (Bafumi, Erikson, and Wlezien 2007). Such exercises help us understand polling better, and may even help us create more representative polls. But they do not help us understand the more fundamental questions about what drives congressional elections. However, predicting actual results well can be useful in itself. Models designed to predict an election with the utmost accuracy and models designed to test factors that are earlier in the “funnel of causality” are both interesting and informative. But the field of forecasting would benefit if there was an explicit distinction between the two types of models. Of the six quantitative models from 2006 that were reviewed above, the three most accurate contained a vote intention variable (Abramowitz 2006, Bafumi, Erikson, and Wlezien 2007, Klarner and Buchanan 2006a). As these models may have an unfair advantage over others, the community of forecasters may want to think seriously about putting such models into an entirely separate league.

Another potential advantage of forecasting models is that they focus attention on the more banal aspects of the determinants of election outcomes, instead of concentrating solely on

aspects of elections that may appear at first blush to be more theoretically interesting. Of course, all determinants of election outcomes are important. But we cannot understand broader theoretical issues until we iron out some of the more mundane aspects of how to model congressional elections. Although this is only suggestive at this stage, one example from Klarner (2008) is trying to clean prior district election results of the impact of incumbency. The idea is that if a Democratic incumbent ran in a House district last time, the percent of the vote won by the Democrats in the House district will overstate how Democratic a district is. If a Democratic incumbent had not run in such a district, prior vote would be a sign of a more Democratic district. Mundane? Perhaps. But to understand other variables' impacts on congressional elections, or the impact of incumbency for that matter, these mundane types of relationships need to be understood. Accordingly, when the House model presented in Klarner (2008) does not include the lagged variable measuring incumbency, the estimated impact of many of the other explanatory variables changes greatly. An all-encompassing focus on predicting Y ensures attention to these "mundane" modeling decisions. Another example of delving into what some thought were mundane empirical regularities was a prolonged study of the many aspects of the incumbency advantage, which Fiorina argued was fruitful (2005, 162-3).

Conclusion

The subfield of election forecasting has much to offer political science as a discipline. It encourages progress in understanding how elections work, as well as providing entertainment to people throughout as well as outside political science. Overall, the House forecasts of 2006 were fairly accurate. All of them predicted the Democrats would pick up a sizeable number of seats. The two published House forecasts of 2008 were also at least fairly accurate. However, election

forecasters need to put as much emphasis on forecasting House—and especially Senate—elections as they do presidential elections.

The development of district or state level forecasting models can also potentially improve our discipline's ability to forecast elections. As it is, House district level forecasts can already hold their own against the forecasts of Washington insiders when the election is about three months off. Pooled models allow an elegant solution to taking the competitiveness of districts into account, and also point the way to understanding the role of campaigns in elections better.

Bibliography

- Abramowitz, Alan. 2002. "Using the Generic Vote Question to Forecast the Outcome of the 2002 Midterm Election." Elections, Public Opinion and Voting Behavior section of the American Political Science Association Symposium on Mid-Term Elections.
- 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 and Politics* 39(October):863-66.
- Abramowitz, Alan I. 2008. "Forecasting the 2008 Presidential Election with the Time-for-Change Model." *PS: Political Science and Politics* 41(October):691-95.
- Abramowitz, Alan I., and Jeffrey Segal. 1986. "Determinants of the Outcomes of U.S. Senate Elections." *Journal of Politics* 48(2):433-39.
- Armstrong, J. Scott. 2001. "Standards and Practices for Forecasting." In *Principles of Forecasting: A Handbook for Researchers and Practitioners*, ed. J. Scott Armstrong. Norwell, MA: Kluwer Academic Publishers.
- Bafumi, Joseph, Robert S. Erikson, and Christopher Wlezien. 2007. "Forecasting House Seats from Generic Congressional Polls." *Extension of Remarks, Legislative Studies Section Newsletter* 30(1). www.apsanet.org/~lss/Newsletter/jan07/EOR.pdf.
- Bartels, Larry, and John Zaller. 2001. "Presidential Vote Models: A Recount." *PS: Political Science and Politics* 34(March):9-20.

- Brandt, Patrick T., and Thomas L. Brunell. 2006. "Forecasting the 2006 House Election." http://www.utdallas.edu/~pbrandt/pbrandt/2006%20House%20Forecasts_files/Forecast2006.pdf, accessed December 2, 2008.
- Campbell, James E. 2006. "Forecasting the 2006 National Elections to the U.S. House of Representatives." *The Forum* 4(2): article number 2.
- Campbell, James E., and James C. Garand. 2000. "Forecasting U.S. National Elections." In *Before the Vote: Forecasting American National Elections*, eds. James E. Campbell and James C. Garand. Thousand Oaks, CA: SAGE Publications, Inc.
- Cuzan, Alfred, and Charles M. Bundrick. 2006. "Will the Republicans Retain Control of the House of Representatives in 2006?" Paper presented at the Northeastern Political Science Association annual meeting, Boston.
- Cuzan, Alfred, and Charles M. Bundrick. 2008. "Predicting Presidential Elections with Equally Weighted Regressors in Fair's Equations and the Fiscal Model." Manuscript under review, but also available at <http://www.uwf.edu/govt/documents/CuzanandBundrick-2008-PredictingPresidentialElections6-23-08.pdf>.
- Eijk, Cees van der. 2005. "Election Forecasting: A Sceptical View." *British Journal of Politics and International Relations* 7:210-14.
- Erikson, Robert S., and Gerald C. Wright. 2005. "Voters, Candidates, and Issues in Congressional Elections." In *Congress Reconsidered*, 8th Edition, ed. Lawrence C. Dodd and Bruce I. Oppenheimer. Washington, DC: CQ Press.
- Fair, Ray C. 1978. "The Effect of Economic Events on Votes for President." *Review of Economics and Statistics* 60(2):159-73.

- Fiorina, Morris. 2005. "Keystone Reconsidered." In *Congress Reconsidered*, 8th Edition, ed. Lawrence C. Dodd and Bruce I. Oppenheimer. Washington, DC: CQ Press.
- Gerber, Alan S., Donald P. Green, and David Nickerson. 2001. "Testing for Publication Bias in Political Science." *Political Analysis* 9(4):385-92.
- Highton, Benjamin. 2000. "Senate Elections in the United States: 1920-94." *British Journal of Political Science* 30:483-506.
- Jacobson, Gary C. 2004. *The Politics of Congressional Elections*. 6th ed. New York: Longman.
- Jacobson, Gary C. 2006. "Competitive in U.S. Congressional Elections." In *The Marketplace of Democracy: Electoral Competition and American Politics*, eds, Michael P. McDonald and John Samples. Washington, DC: CATO Institute and Brookings Institutional Press.
- King, Gary. 1986. "How Not to Lie with Statistics: Avoiding Common Mistakes in Quantitative Political Science." *American Journal of Political Science* 30:666-87.
- Klarner, Carl. 2008. "Forecasting the 2008 U.S. House, Senate, and Presidential Elections at the District and State Level." *PS: Political Science and Politics* 41(October): 723–8.
- Klarner, Carl E., and Stan Buchanan. 2006a. "Forecasting the 2006 Elections for the U.S. House of Representatives." *PS: Political Science and Politics* 39 (October): 857–61.
- . 2006b. "Forecasting the 2006 Elections for the United States Senate." *PS: Political Science and Politics* 39 (October): 849–55.
- Lockerbie, Brad. 2008. "Election Forecasting: The Future of the Presidency and the House." *PS: Political Science and Politics* 41 (October): 713-16.
- Tufte, Edward R. 1975. "Determinants of the Outcomes of Midterm Elections." *American Political Science Review* 69 (September): 812–26.

¹ These forecasts were posted shortly after August 23rd, 2008 on

www.forecastingprinciples.com/PollyVote/index.php/pollyblog/3/85-new-model-forecast.html.

² The fact that the model under-estimated the aggregate gain by four, and called five specific races incorrectly is because probabilities are assigned to each seat's outcome and these are then aggregated in a simulation.

³ This analysis was done for all House elections between 1946 and 2006, utilizing data shared by Gary Jacobson.

⁴ Although Pollster's forecasts were not available per se, the polls they were based upon are, and the methodology used to aggregate poll results described on the Web site were used to recreate Pollster's predictions from July 28th, 2008. For three or fewer polls, the average is taken. For between four and seven polls, a regression line is fit, and the estimated value for the last poll is taken. For eight polls or more, a Loess line is fit, and the last estimated value is saved. Although Pollster.com is vague about this, the independent variable in the last two instances is probably a count representing days. The poll results from www.fivethirtyeight.com were not available from that time, nor were the ratings from Congressional Quarterly.

⁵ I know Cook did not have access to my district level predictions at that time, and I was not monitoring his.

⁶ Before, 57 districts were compared, because the outcome of one district has still not been resolved. A close approximation to the vote percentages is known, however. It should also be noted that the final tallies for many races have not been established, but evidence noted above indicates this is not a problem.

Table 1: Accuracy of Different House Forecasting Methods About 100 Days before the 2008 Election

	Klarner Forecasts July 28	Pollster.com	Cook Political Report	Rothenberg Report
Districts Called Correctly	409 of 434 (94.2%)			
Districts Called Correctly of Those Called by Pollster.com	44 of 57 (77.2%)	41 of 57 (71.9%)		
Districts Called Correctly of Those Called by Cook	60 of 71 (84.5%)		61 of 71 (85.9%)	
Districts Called Correctly of Those Called Toss-Ups by Cook	14 of 27 (51.9%)		0 of 27 (0.0%)	
Districts Called Correctly of Those Called by Rothenberg	36 of 54 (66.7%)			38 of 54 (70.4%)
Districts Called Correctly of Those Called Toss-Ups by Rothenberg	6 of 10 (60.0%)			0 of 10 (0.0%)
Number of Top 27 Competitive Districts Called Correctly	8 of 27 (29.6%)	7 of 27 (25.9%)	10 of 27 (37.0%)	8 of 27 (29.6%)
False Positives	19 of 27	20 of 27	17 of 27	20 of 28

Table 2: Predicting the Democratic Percent of the Two-Party Vote (Centered)

	Model one: House	Model two: House	Model three: Senate	Model four: Senate
Klarner Forecast of Democratic % of Two-Party Vote	.80*** (.04)	.09 (.11)	.19 (.14)	.06 (.09)
Cook seven-point scale	1.66*** (.51)	2.03** (.90)	2.67 (2.04)	.56 (1.23)
Rothenberg nine-point scale	-.37 (.34)	.56 (.47)	1.38 (2.12)	-.69 (1.38)
Pollster.com Forecast of Democratic % of Two-Party Vote		.15* (.09)		.97*** (.19)
Constant	1.75*** (.28)	1.90** (.65)	2.22** (1.00)	1.23 (.73)
Standard Error of the Estimate	5.04	4.23	5.58	3.28
R-Squared	.92	.72	.84	.93
N	377	58	34	28

Note: The dependent variable in both models is the actual percent of the two-party vote going to Democrats, subtracted by fifty. The cell entries are the unstandardized regression coefficient with their standard error in parentheses. *** = $p < .001$, two-tailed, ** = $p < .05$, two-tailed, * = $p < .05$, one-tailed.

Table 3: Accuracy of Different Senate Forecasting Methods About 100 Days before the 2008 Election

	Klarner Forecasts July 28	Pollster.com	Cook Political Report	Rothenberg Report
Seats Called Correctly	29 of 34 (85.3%)			
Seats Called Correctly of Those Called by Pollster.com	23 of 28 (82.1%)	26 of 28 (92.9)		
Seats Called Correctly of Those Called by Cook	27 of 30 (90.0%)		29 of 30 (96.7%)	
Seats Called Correctly of Those Called Toss-Ups by Cook	2 of 4 (50.0%)		0 of 4 (0.0%)	
Seats Called Correctly of Those Called by Rothenberg	27 of 31 (87.1%)			29 of 31 (93.5%)
Seats Called Correctly of Those Called Toss-Ups by Rothenberg	2 of 3 (66.7%)			0 of 3 (0.0%)
Number of Top 4 Competitive Seats Called Correctly	0 of 4 (0.0%)	3 of 4 (75.0%)		2 of 4 (50.0%)
False Positives	4	1		2
Number of Top 5 Competitive Seats Called Correctly	0 of 5 (0.0%)	3 of 5 (60.0%)	2 of 5 (40.0%)	
False Positives	5	2	3	

Figure 1: Forecast versus Actual Democratic Percent of the Vote in House Districts

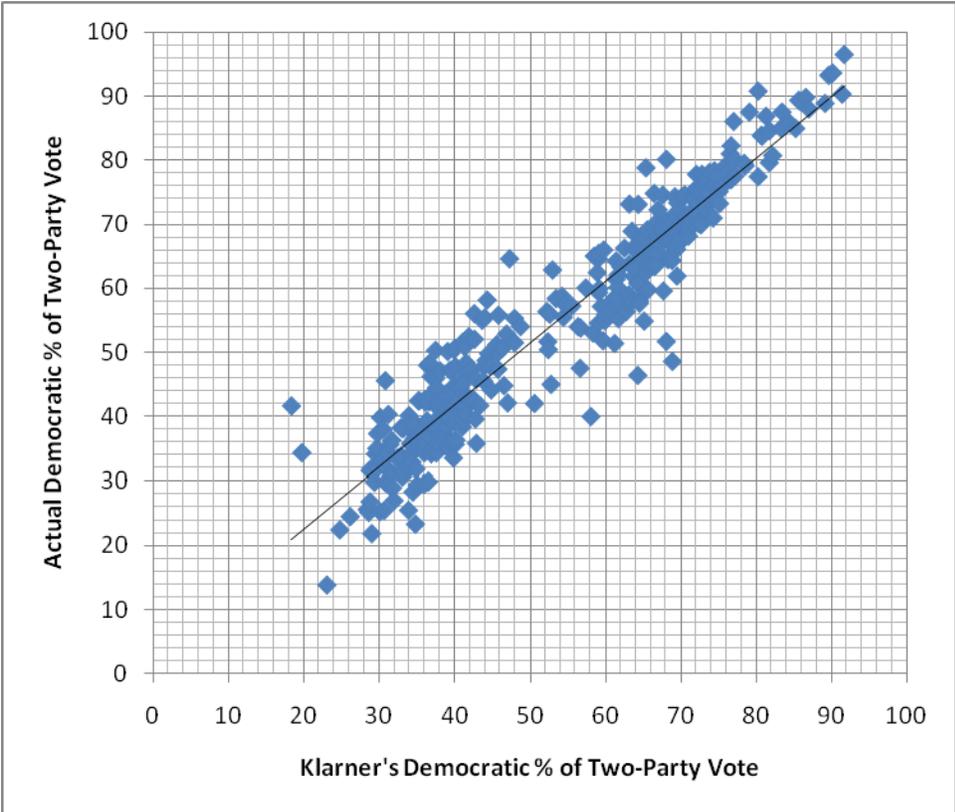


Figure 2: Forecast versus Actual Democratic Percent of the Vote in Senate Races

