Optimizing State Legislative Election Forecasts

Carl Klarner
Carl.Klarner@gmail.com
Visiting Scholar, Harvard University

November 14, 2014 Draft
Version updated somewhat from version distributed to NPSA Panelists.

Abstract
State legislative elections offer a unique opportunity to test a variety of methodological issues that give insight into forecasting and elections generally. This is because of the large number of such elections. How to best predict state legislative elections is also important in itself. This manuscript has two overall purposes. The first is to present the prediction model utilized for predicting the results of the 2014 state legislative elections. The second is to assess the efficacy of competing modeling decisions with the use of “drop one analyses” in which one biennium of elections were dropped in turn from a sample of elections spanning 1968 to 2012.

Presented at the annual meeting of the Northeast Political Science Association, Boston, MA

The research presented in this manuscript builds off the work of Klarner's (2010a, 2010b) pre-election forecasts of the 2010 state legislative elections. The first section of the paper presents the model that was utilized to make forecasts for the 2014 state legislative elections, publicized on October 31, 2014 and reports on those forecasts. The second section examines numerous modeling decisions and their efficacy at forecasting. The overall approach is to run pairs of models that only differ in one way from each other. The relative forecasting performance of the two models within a pair is then assessed to provide insight into the consequences of a particular modeling approach.

SECTION ONE: 2014 STATE LEGISLATIVE FORECASTS

Forecasts for the 2014 state legislative elections made by Political Scientist Carl Klarner on October 30 indicate that the Democrats will lose majorities in five state senates and nine state houses. Those state senates are Colorado, Iowa, Maine, New York and Washington, while the state houses are Colorado, Kentucky, Maine, Minnesota, Nevada, New Hampshire, New Mexico, Washington and West Virginia. Forecasts were generated from analyses of the 77,844 contested state legislative elections that occurred between 1968 and 2012. National conditions (state of the economy, presidential approval, midterm penalty, Congressional vote intention) were found to have a large effect on state legislative elections and worked against the Democrats in the forecast. Numerous race specific variables were also examined (incumbency, prior vote in the district, etc.). Table 1 reports forecasts at the state legislative chamber level. Specific predictions for all 5,577 state legislative elections up this year, 3,157 of which were contested by at least one Democrat and one Republican, are reported in the Excel sheet "Klarner_State_Leg_Forecast_Districts2014_10_30.xlsx," publicized before the election and sent out to NPSA panelists.

Forty-five states will have elections to their state legislatures this coming Tuesday (see Table 1), although four of those states do not have elections to their state senates.[1] Table 1 presents the percentage of major party legislators who are Democrats going into the 2014 elections in each chamber (column four).[2] The percentage of seats in each chamber the Democrats are forecast to have is reported in column seven. Forecasts are arrived at by taking into account model uncertainty, and then conducting 2,000 simulations of the 2014 elections. The median percentage of seats the Democrats hold across these 2,000 simulations is then taken as the forecast for the election outcome. These simulations indicate that the Democrats will be losing seats in most states. The only gains are forecast to come in the Oregon, North Dakota (strangely enough) and Rhode Island State Senates, as well as the North Dakota House (again, a puzzle). West Virginia is forecast to see especially large losses for the Democrats in both of its chambers. Other states will see losses for the Democrats of over ten percent, including the California Senate, the Nevada House and both chambers of the Maine legislature.

These simulations also indicate the percent probability that the Democrats will have control (i.e., a majority) of seats in a chamber after the election (column six of Table 1). Of the five state senates that Democrats are forecast to lose, the New York State Senate is the most competitive overall, with a 46 percent chance of Democratic control. The Washington State Senate sees a 25 percent chance, with the other three having even lower percentages. Among state houses that the Democrats are forecast to lose, Nevada is the most competitive (46 percent chance of retaining control), New Mexico is the second most (31 percent) with the others at 25 percent or less. The Republican wave could, of course, extend more greatly, but taking many more chambers than those already listed is not likely, according to the model. The Oregon state house has a 54 percent chance of going to the Democrats, so it is a very plausible target for the Republicans, but Table 1 indicates that Democratic majorities in state houses are virtually assured after that. Republicans may also pick off the Connecticut State Senate (67 percent chance of Democratic control) or the Nevada State Senate (79 percent chance of Democratic control) but these are long shots and no other possible targets among state senates show themselves after those.

The upcoming election sees 26 partisan state senates with half (or in one case, one-third) of their seats up. That means that holdover state senators—those elected to four year terms two years ago—will play a major role in which party controls those state senates. The percentages of all state senate seats that are represented by holdover Democrats and holdover Republicans is presented in the accompanying Excel sheet.[3] The surprising seat gains predicted for Democrats in North Dakota is explained by holdover legislators. First, it should be noted that even if the forecast is correct, the Democrats will still be left with a mere 34 percent in the Senate, and 26 percent in the House. In the state senate, 10 out of 23 of the holdover state senators are Democrats. Furthermore, the North

---

[1] Excepting special elections to fill the remaining two years of four year terms.

[2] http://www.ncsl.org/documents/statevote/legiscontrol_2014.pdf accessed October 29, 2014.

[3] Note: these figures do not take into account state senators who left office in the middle of their terms and were replaced by someone of a different party.

Dakota House of Representatives is the only "lower chamber" in the county that has holdover legislators because of its staggered four year terms. Again, the Democrats have a non-trivial minority of holdover state house members (14 out of 46). This means that the Democrats merely need to win more than 4 out of 24 of the state senate seats that are up to make progress, or more than 9 out of 47 of the state house seats that are up. Give the number of Democratic incumbents and legislators who have served in the past (but not the immediate past) who are running, the Democrats are expected to do this by the model.

NATIONAL CONDITIONS IN 2014

National conditions do not look good for the Democrats generally. Congressional vote intention and presidential approval imply a large wave in favor of the Republicans, although the state of the national economy is not as bad as one might expect.

The Congressional vote intention question was not asked by Gallup in 2014, so the most recent two polls from pollingreport.com as of October 22, 2014, conducted between Oct 9 and Oct 14, were averaged and utilized.[4] They indicated that 50.6 percent of registered voters said they would vote for Democrats in the upcoming U.S. House election (of those saying they would either vote for a Democrat or a Republican). Historic statistics for Congressional vote intention are largely in terms of a national sample of voters (not registrants or likely voters), and the 50.6 figure above is estimated here to be equivalent to 51.5 percent of a national sample of voters saying they will vote for a Democrat. Comparing to surveys asked at around the same time, this is very close to other historic lows for the Democrats in the 1968 to 2014 period: in 2010, 49.3 percent said they would vote for a Democrat for Congress, while in 1994 49.5 percent did.[5] (Generally, a good rule of thumb to use with Congressional vote intention is to subtract five percent for the percent of voters who will vote for Democrats on election day.)

The presidential approval rate of 40.6 percent that Obama received on rolling Gallup polls taken between October 13 and October 19 also imply trouble for the Democrats. Between 1968 and 2014, this approval rate is ninth from worst, while it is thirty-ninth from the best. For midterms in the 1968 to 2014 period, only 1986 and 2002 see presidential approval ratings boding this much ill for the Democrats (and in both cases the Republican president's popularity was offset by the midterm election penalty). Obama's current approval rating is almost exactly what Clinton's was immediately before the 1994 elections (41.0 percent).

The national economy—represented here by a weighted average of real quarterly income growth (Hibbs 2012), is actually not as bad as one might expect. Growth was 2.38 percent, which for the twelve midterm election years examined here (i.e., even non-presidential years), is only beat by four other elections, and beats seven. Comparing to all years from 1968 to 2014, 21 years were better, 25 were worse.

Taking the above factors into account, voters also appear to punish the party of the president in state legislative elections in midterm years. For this reason, Democratic state legislative candidates can expect to lose three quarters of a percent compared to two years ago, if that is the last election they ran in.

Taken together, these factors explain why the Democrats are predicted to lose state legislative seats all across the country.

METHODOLOGY

The forecasts presented here build off the work of Klarner's (2010a, 2010b) pre-election forecasts of the 2010 state legislative elections. To arrive at these forecasts, analyses of most state legislative elections in 48 states in the 1968 to 2012 period were conducted. Nebraska was not used because of its non-partisan state legislature, Minnesota was not used when its legislature's elections were non-partisan, and Louisiana was not included because of its unique election system (and it does not have elections this year anyway). Only contested elections were

---

[4] The first poll was a Fox News Poll conducted between October 12 to October 14 (N=1,012), while the second was a ABC News / Washington Post poll conducted between October 9 to October 12 (N=1,006). The former had 43 percent of registered voters saying they would vote for the Democrat and the same percentage saying they would vote for a Republican. The two corresponding percentages were 46 and 44 in the latter poll. Historically, the generic ballot question is reported mostly as a sample of all national adults (445 out of 523 surveys collected, versus 333 surveys reporting for registrants and a mere 88 reporting for likely voters). As a result, surveys that did not report on national samples of adults had figures for a national sample of adults estimated from a regression model. The 2014 figures are also estimated, as only registrants and likely voters were reported from those survey. Two polls were used because of the small samples of each, and the small number of days each poll was conducted.

[5] 1986 also sees a lower amount than the 2014 amount, but in the next survey in 1986 just a few days later, the Democratic proportion went up considerably.

examined, as fully uncontested elections can be called in advance.[6] Past elections with more than 20 percent of the vote going to a non-major party candidate were not considered.

The variables used in Klarner's (2010a) forecasting model were all included as predictor variables (with one exception—whether a seat is open because of a term limit), with the addition of many others. Factors that could be measured before an election were compared to election outcomes for 77,844 state legislative elections. The following is a brief list of the factors used to forecast the 2014 state legislative elections. A more precise and extensive explanation can be found in the appendix. The results of the hierarchical linear models used to make the prediction are reported in Table 2. Seventy-six predictor variables were utilized, many of which were the lagged components associated with a non-lagged independent variable.

Incumbency

How many terms an incumbent has held office

How much the boost from incumbency goes up when funding for the state legislature is higher.

The fact that the boost from incumbency appears to have increased then decreased in the 1968 to 2012 period.

Whether a candidate who held office in the other legislative chamber is running

Whether a candidate who served in the legislature in the past—but not immediately in the past—was running.

Whether the current election is fully contested or partially contested.

The vote in the district in the last election (but not measured when redistricting has occurred).

Whether the immediately prior election was fully contested, partially contested or fully uncontested.

How well the state economy was doing (which was not found to be associated with state legislative election outcomes).

A penalty variable that captures the propensity for the governor's party to lose seats in the middle of their terms.

A penalty variable that captures the propensity for the president's party to lose seats in the middle of their terms.

Presidential approval

Congressional vote intention: this is a question that asks people how they will vote for Congress, but it is associated with how people vote for the state legislature.

How well the national economy is doing

Repeat contest vote from the past: the same two candidates face off against each other in a race. Such elections are more predictable in their outcomes in contrast to other elections.

Repeat loser's vote from the past: the candidate who lost last time is running again (but the candidate who won last time is not running again). Such elections are more predictable in their outcomes in contrast to other elections.

Winner from other chamber's vote: how well did a candidate who won an election last time in the other chamber of the legislature do? This is the only factor that was not related to how well they'll do in the current election.

Loser from other chamber's vote: how well did a candidate who lost an election last time in the other chamber of the legislature do? This is (weakly) related to how well they'll do in the current election.

---

[6] Free-for-all multimembers districts that are "partially contested" (for example, the election will have two winners, but only one Democratic candidate runs, while two Republicans run) were also analyzed.

Vote in the other chamber of the legislature from last time when state house and state senate districts are identical. This especially helps predictions when the last election in a district was uncontested.

For state senate districts only: vote in the state house from last time when state house districts are nested within state senate districts. This especially helps predictions when the last election in a district was uncontested.

Trend variables for the nation as a whole, the south and border states are also taken into account.

SECTION TWO: OPTIMIZING FORECASTS

What is often referred to as “drop-one” analysis was conducted to examine the efficacy of various modeling decisions. Typically, such analyses drop one observation from a sample, make an estimate for that observation based on an analysis of the other cases, and error is computed as the difference between the actual and predicted values (Abramowitz 2012). In the analysis here, election outcomes for a biennium worth of elections are deleted from the dataset (i.e., an even numbered year as well as the odd numbered year that precedes it), forecasts are made for the elections in that year, and forecast error is computed. This is done for each biennium in the dataset. An obvious drawback of such an analysis is that what is being forecast occurred between years the phenomena being forecast has known values for. If there is a “structural break” with the past, whether minor or major, it will plausibly not have the same detrimental effect on forecasts as when a forecast is made at the end of a time series. The strength of these analyses is that they produce a large number of predictions.

These methods of assessing model accuracy diverge from traditional Political Science methods, which typically rely on “in sample” measures of forecasting error such as the standard error of the estimate or, worse, assessing whether variables attain statistical significance. Such methods suffer the drawback of reading too much into a sample. In particular, the standard error of the estimate does not apply to out of sample data (Beck 2000, 162). Of more concern is that if a sample has been analyzed many times, such as for the U.S. House, with literally thousands of articles published explaining these elections, predictor variables will be biased to look more efficacious than they are.

Issues and challenges that present themselves to forecasting state legislative elections, as well as modeling decisions that might help deal with these challenges, are discussed below.

MISSING DATA

The best predictor of vote share in a legislative election is what vote share in that district was last time. This makes missing data a major problem for predicting state legislative elections because of redistricting and the large proportion of races that are uncontested.

When redistricting occurs precise information on prior vote share is unknown. Vote share in prior U.S. House elections mapped onto current election districts in a redistricting year is widely available. But these data are not available for state legislative districts. As a result, all lagged variables measured at the district level must be treated as “system missing.” In the 1968 to 2014 period, 26.8 percent of elections were redistricted since the last time a seat was up. This is an especially serious problem in state senates with four year staggered terms, as 40 percent of the seats up over the course of a decade have been redrawn since the last election. The average state senate sees 33.2 percent of its elections redistricted since the election, while state houses average 24.6 percent. For fully contested races, these figures are 30.9 for state senates, and 21.7 percent for state houses.

If a district was not contested in the last election, one still knows which party won the district, but this is much less information than knowing the percentage of the vote a party obtained. In total, 35.8 percent of state legislative elections are not contested. However, if one is only concerned with forecasting contested elections, the situation is helped considerably, since there is a correlation between whether an election was uncontested last time and whether it is uncontested this time.

The breakdown of the availability of past vote share for the 83,412 fully contested elections from 1968 to 2014 is displayed in the following table.

Table: Percent of Elections by Reason Lagged Values are Observed or Not

	Fully Contested (N=83,412)	Partially Contested (N=2,585)	Fully Contested, Restricted Sample (N=78,323)	Partially Contested, Restricted Sample (N=2,383)
Missing Values at t-1	.09	.04	.04	.00

No Lagged Value, Beginning of Dataset	7.55	10.72	7.25	10.53
No Lagged Value, Redistricting	23.78	17.91	23.83	18.38
Fully Uncontested Prior Election	13.74	13.81	13.09	13.18
Partially Uncontested Prior Election	1.68	27.39	1.57	26.86
Fully Contested Prior Election	53.16	30.14	54.22	31.05
Column Total	100.00	100.00	100.00	100.00

The situation is further helped a little by sensible restrictions on the sample, which entail excluding years in which a legislative chamber had lopsided majorities of one party (more than 84 percent) or a third party candidate attained more than 20 percent in a race. Column four of the above table reports this situation. However, only 54.22 percent of cases have a clear lagged vote share from the last election.

To gauge the problems that not having lagged vote share causes for prediction error, a regression analysis that only utilized the variables reported in Table 2 as being part of the "basic model" was conducted on 23 different datasets. In each dataset, a biennium of data was deleted for the dependent variable. For example, 1983 and 1984 were changed to system missing for the dependent variable for one dataset. The coefficients from other years were then utilized to generate predictions for each biennium, and the absolute value of the amount each prediction was off by was computed. For the following discussion, this was done in two separate analyses. In one, missing values were simply coded "0" and a battery of dummy variables coded "1" where such missing cases occurred for a corresponding variable were included (in practice, only six such variables had to be included because of correspondence in missingness between variables). In a second, Gary King's Amelia program (Honaker et al 2001, King et al 2001) was used to impute missing data. The average absolute value of prediction error for the above types of cases is reported in the Table immediately below. Table 3 reports the average absolute value of prediction error in numerous models that will be reported on shortly. Use of Amelia brings the mean absolute value of prediction error down from 6.934 percent (model 2 of Table 3) to 6.291 percent (model 4 of Table 3), a substantial improvement (.641 percent) for the basic model. The full model also improves with the use of Amelia. Error goes down from 6.915 percent (model 1 of Table 3) to 6.263 percent (model 3 of Table 3, a .622 percent improvement.

Table: Mean Absolute Value of Out of Sample Forecast Error by Status of Lagged Value

	No Amelia Imputation Fully Contested, Restricted Sample (N=75,416)	Amelia Imputation Fully Contested, Restricted Sample (N=75,416)
Missing Values at t-1 (N=31)	15.76	14.63
No Lagged Value, Beginning of Dataset (N=5,648)	7.93	6.87
No Lagged Value, Redistricting (N=17,869)	8.70	7.17
Fully Uncontested Prior Election (N=9,809)	8.16	8.14
Partially Uncontested Prior Election (N=1,189)	5.72	5.71
Fully Contested	5.84	5.47

Prior Election (N=40,870)		

Predictions from drop-one analysis.
Note: the column N does not agree with the column N in the last table because elections for 2014 are not included in these figures.

In column two of the above table, it is evident that prediction error goes up markedly when a prior election is uncontested, or when lagged values are not available because of redistricting. Prediction error in elections that were fully uncontested last time is 2.3 percent higher than in those that were preceded by fully contested elections. Similarly, prediction error goes up by almost three percent when you go from an election that was fully contested last time to one that has been redistricted.

Use of Amelia reduces prediction error in the instance of redistricting and no prior lagged value because a case is at the beginning of the dataset by over one percent. Prediction does not go down for fully uncontested prior elections. In fact, elections with such prior situations see a lot more prediction error than those that have been redistricted or that are at the beginning of the dataset. This is because it is assumed that vote share in the past is equivalent to "100" if no Republican candidates ran, and "0" if no Democratic candidates ran, although this assumption is relaxed somewhat with the inclusion of a variable in the model indicating how contested an election was. The substantial improvement Amelia brings to the redistricted cases and that are missing lagged values implies that the method of recoding prior uncontested elections as "100" and "0" is flawed and will be revisited shortly.

Other methods of dealing with missing lagged variables will also be examined here. One is to bring in information about the other chamber's vote share when state senate and state house districts are identical or when state house districts are nested inside of state senate districts. Table 3 indicates that doing so brings down forecasting error by .035 percent overall (6.256 versus 6.291), a miniscule amount, but every bit helps. The following table displays the improvement in forecasting error among cases that have a nested or identical counter-part.

Table: Mean Absolute Value of Out of Sample Forecast Error

	Basic Model	Basic Model Plus "Other Chamber" Vote Share	Difference	Full Model	Full Model Minus "Other Chamber Vote Share	Difference
Uncontested Last Time, No Other Chamber Equivalent (N=9,186)	8.15	8.14	.01	8.13	8.13	.00
Uncontested Last Time, Other Chamber Also Uncontested (N=197)	7.17	7.16	.01	7.34	7.36	-.02
Uncontested Last Time, Other Chamber Undercontested Last Time (N=232)	7.56	6.99	.57	6.98	7.59	.61
Uncontested Last Time, Other Chamber Fully Contested (N=466)	7.11	6.62	.51	6.73	7.25	.52

The table above indicates that whether you add vote share from the other chamber to the basic model, or subtract it out of the full model, the inclusion of vote share from the other chamber brings down prediction error by about half a percent, but only if the election in the other chamber was also either fully contested, or at least partially contested. Knowing which party took the seat unopposed last time does not help prediction, presumably because it is always the same as the party that took the (unopposed) seat in question.

The small numbers of cases that display nested / identical components (around 900 elections) provides insight into why the inclusion of these variables only brings down prediction error overall by .035 percent. However, the information gleaned above will provide insight into how to use state legislative returns at the county level to reduce forecasting error.

DISTRIBUTIVE LAG MODELS

Distributive lag models include a lagged version of the dependent variable, as well as lagged components for independent variables. Debate exists about whether such models are good to use. Whether they are good to use generally or not, their efficacy in state legislative models seems to have theoretical merit.

The inclusion of a lagged dependent variable creates a model of change in vote share as the coefficient associated with lagged vote share approaches “1” as occurred in the models examined here. Accordingly, change in independent variables must also be assessed, meaning that lagged independent variables should also be included. How the national midterm penalty works is a good illustration of this. Alabama had state legislative elections in 1986 and 1990, but not 1988. In both years, it was a midterm penalty for the Republicans, so we would not expect much of a change in Democratic vote share between the two years because of the midterm penalty (represented by a variable introduced below called “Midterm penalty” that takes on a value of “1” in a midterm with a Democratic president, “-1” with a Republican, and “0” when not a midterm). The sum of the coefficients associated with the lagged and non-lagged midterm penalty variable should therefore cancel out to an extent, as the lagged coefficient would be positive in sign while the non-lagged coattail would be negative. But from 1990 to 1994, we would expect to see a swing in a Republican direction, and the sum of the coattails would be a negative quantity larger in magnitude than the non-lagged coefficient alone. State legislative chambers with two year cycles would have much different scenarios that would be captured by this modeling strategy. Generally, the lagged components will take on signs that are the opposite of their non-lagged components, unless an effect continues for more than one period.

At the district level, consider the following example. Say you know that the Democratic vote in the last election in a district was 65 percent. Would your perception of how Democratic that district is be altered by whether the district was won by a Democratic incumbent, or whether it was an open seat? Including a lagged incumbency variable clearly matches our intuition and similar arguments can be made for all other candidate level variables.

Table 3 displays analyses that exclude lagged independent variables. The first set of these leave out the lagged component of all independent variables if the independent variable in question has both a lagged and non-lagged component (models 5 and 15) (the one exception is lagged contestation advantage: if lagged vote share is left in, it cannot be excluded). The second set of these leave out only the lagged components of the national level independent variables (models 6 and 16). Table 3 gives clear evidence that excluding lagged independent variables is not a good idea. For the full model (model 3), error goes from 6.263 to either 6.448 when all lagged components are dropped (model 15) or 6.410 when national level lagged variables are dropped (model 16). The former difference is .185 percent. The increase in error for the basic model is smaller when dropping lagged components (.094 for dropping all, .081 for just dropping national) but still substantial. Because the full model performs better, what happens when they are dropped from the full model is more important.

COMPLEX VERSUS SIMPLE

With around 100,000 elections available for analysis, does it pay to have numerous variables tracking as many aspects of state legislative elections as possible? Or is it better to have a more parsimonious model? A look at Table 2 indicates that many of the variables that are not in the “basic model” attain statistical significance and have sizeable coefficients. But whether variables attain statistical significance in a model or the size of their t-statistics does not necessarily indicate their ability to aid accurate forecasts (Beck 2000, 165). How effective are the variables reported in Table 2 that attain statistical significance at reducing forecasting error?

“Second Term Incumbent” and “Third or More Term Incumbent” both attained statistical significance, and both have non-trivial coefficients. A second term incumbent can expect to receive .58 percent more than an incumbent running for reelection the first time. Furthermore, these are variables with lots of variation, and therefore a lot of potential predictive power. But their inclusion does not do anything to increase the predictive power of the model. Comparing the basic models (4 versus 13) or the full models (3 versus 23) sees no change in mean absolute error.

The interactions between “Decade” and “Incumbency Advantage” show large coefficients, attain statistical significance, and have sizeable t-values. But their inclusion appears to actually hurt forecasting ability slightly in the case of the basic model (model 4 versus 14) and not improve things for the full model (model 3 versus 24).

Vote share from prior contests that have characteristics that are thought to make such contests better predictors of current conditions often attain statistical significance. "Repeat Contest*Vote Share$_{t-1}$" is statistically significant, with a large t-value, as is "Repeat Loser's Vote Share$_{t-1}$" as is "Loser From Different Chamber's Vote Share$_{t-1}$", but when these variables are added to the basic model (model 4 versus model 11) or taken out of the full model (model 3 versus model 21) mean absolute error barely changes at all.

Both the "South*Trend" and "Border*Trend" variables attained statistical significance. Care was taken that the trajectories of these trends be allowed to slow over time with the inclusion of squared and cubed terms (which attained statistical significance in the case of the Border-interactions with those variables). Yet adding these variables to the basic model (model 4 versus 9) slightly worsened forecast error (.02 percent). Taking these trends out of the full model (model 3 versus 19) gave some indication that they helped the model (by .013 percent) but these effects are extremely small.

By far the most widely cited finding from the study of state legislative elections is the finding that the boost from incumbency goes up when state legislatures have more institutional support (Berry et al 2000), so the performance of the legislative spending / incumbency interaction is worthy of note. The modeling strategy adopted in Table 2 was different than Berry et al, which might account for the fact that these variables (except for "Incumbency*Legislative Expenditures, Cat 3") failed to attain statistical significance with the correct signs (Incumbency*Legislative Expenditures, Cat 5 attained statistical significance and had a strongly negative sign, contrary to expectations). But whatever the case, adding these interactions to the basic model (model 4 versus 7) or taking them out of the full model (model 3 versus 17) barely changes things at all (.004 percent change in mean absolute error).

There is more that can be gleaned from Table 3, but the overriding point is that the addition of numerous variables to the models reported in Klarner (2010a) have not substantially reduced forecasting error. Overall, the mean absolute value of forecasting error went down by only .028 between the basic and full models (model 4 versus model 3).

CONCLUSION

The addition of numerous variables to the "basic model" presented in Klarner (2010a) did not reduce forecasting error by very much, which is mildly disheartening. However, insights from this process will be implemented in the immediate future. A large number of modeling ideas still have to be tested and hopefully those will result in substantial reductions in forecasting error. It is also heartening that two modeling decisions that have been utilized in my past models that are somewhat controversial—use of Amelia and distributive lag models—were shown to be effective at reducing forecasting error.

BIBLIOGRAPHY

Abramowitz, Alan. 2012. "Forecasting in a Polarized Era: The Time For Change Model and the 2012 Presidential Election." *PS: Political Science and Politics* 45(October): 618-619.

Beck, Nathaniel. 2000. "Evaluating Forecasts and Forecasting Models of the 1996 Presidential Election." In *Before the Voting: Forecasting National Elections*, ed. James E. Campbell and James C. Garand. SAGE Publications, Inc.: Thousand Oaks, CA. 161-167.

Berry, William D., Michael B. Berkman, and Stuart Schneiderman. 2000. "Legislative Professionalism and Incumbent Reelection: The Development of Institutional Boundaries." American Political Science Review 94(December): 859-74.

Erikson, Robert S., and Christopher Wlezien. 2012. *The Timeline of Presidential Elections: How Campaigns Do (and Do Not) Matter.* Chicago: University of Chicago Press.

Folke, Olle, and James M. Snyder. 2012. "Gubernatorial Midterm Slumps." *American Journal of Political Science* 56(October):931-48.

Hibbs, Douglas E. 2012. "Obama's Reelection Prospects under 'Bread and Peace' Voting in the 2012 US Presidential Election." *PS: Political Science and Politics* 45(October):635-39.

Holbrook, Thomas M. 2012. "Incumbency, National Conditions, and the 2012 Presidential Election." *PS: Political Science and Politics* 45(October): 640-643.

Honaker, James, Anne Joseph, Gary King, Kenneth Scheve, and Naunihal Singh. 2001. Amelia: A Program for Missing Data. (Gauss Version). Cambridge, MA: Harvard University.

King, Gary, James Honaker, Anne Joseph, and Kenneth Scheve. 2001. "Analyzing Incomplete Political Science Data: An Alternative Algorithm for Multiple Imputation." *American Political Science Review* 95 (March): 49–69.

Klarner, Carl E. 2010. "Forecasting Control of State Governments and Redistricting Authority After the 2010 Elections." *The Forum* 8(3), Article 14.

Klarner, Carl E. 2010. "Forecasting the 2010 State Legislative Elections." *PS: Political Science and Politics* 43(October):643-48.

Table 1: Forecasts for the 2014 State Legislative Elections by Chamber

state	Cham ber	Seats Up in 2014	Pre-Election % of Seats Held by Democrats	Pre-Election Party Control	Forecast % Chance of Democratic Majority	Forecast Democratic % of Seats	Party Forecast to Have a Majority	Democrats Forecast to Lose Control of Chamber	% Democra Seat Lo Forecas
Alabama	SEN	all	32	R	0	28.6	R	No Change	
Alabama	HS	all	36	R	0	28.6	R	No Change	
Alaska	SEN	half	35	R	0	30.0	R	No Change	
Alaska	HS	all	35	R	0	32.5	R	No Change	
Arizona	SEN	all	41	R	0	40.0	R	No Change	
Arizona	HS	all	40	R	0	36.7	R	No Change	
Arkansas	SEN	half	37	R	0	37.1	R	No Change	
Arkansas	HS	all	48	R	0	41.0	R	No Change	
California	SEN	half	72	D	98	57.5	D		1
California	HS	all	69	D	94	57.5	D	No Change	1
Colorado	SEN	half	51	D	19	45.7	R	Dem Loss	
Colorado	HS	all	57	D	10	44.6	R	Dem Loss	1
Connecticut	SEN	all	60	D	67	52.8	D		
Connecticut	HS	all	65	D	99	57.6	D	No Change	
Delaware	SEN	half	62	D	99	61.9	D		
Delaware	HS	all	66	D	99	58.5	D	No Change	
Florida	SEN	half	35	R	0	32.5	R	No Change	
Florida	HS	all	38	R	0	31.7	R	No Change	
Georgia	SEN	all	32	R	0	32.1	R	No Change	
Georgia	HS	all	34	R	0	32.8	R	No Change	
Hawaii	SEN	half	96	D	100	92.0	D		
Hawaii	HS	all	86	D	100	82.4	D	No Change	
Idaho	SEN	all	20	R	0	14.3	R	No Change	
Idaho	HS	all	19	R	0	12.9	R	No Change	
Illinois	SEN	third	68	D	100	66.1	D		
Illinois	HS	all	60	D	100	55.9	D	No Change	
Indiana	SEN	half	26	R	0	24.0	R	No Change	
Indiana	HS	all	31	R	0	29.0	R	No Change	
Iowa	SEN	half	52	D	5	48.0	R	Dem Loss	
Iowa	HS	all	47	R	0	43.0	R	No Change	
Kansas	SEN	none	20	R	0	20.0	R	No Change	
Kansas	HS	all	26	R	0	22.4	R	No Change	
Kentucky	SEN	half	38	R	0	34.2	R	No Change	
Kentucky	HS	all	54	D	2	45.0	R	Dem Loss	
Louisiana	SEN	none	33	R					
Louisiana	HS	none	43	R					
Maine	SEN	all	56	D	12	42.9	R	Dem Loss	1
Maine	HS	all	61	D	25	47.0	R	Dem Loss	1
Maryland	SEN	all	77	D	100	66.0	D		1
Maryland	HS	all	70	D	100	58.2	D	No Change	1
Massachuset ts	SEN	all	90	D	100	82.5	D		
Massachuset ts	HS	all	82	D	100	76.9	D	No Change	
Michigan	SEN	all	32	R	0	31.6	R	No Change	
Michigan	HS	all	46	R	1	42.7	R	No Change	
Minnesota	SEN	none	58	D					
Minnesota	HS	all	54	D	3	43.3	R	Dem Loss	1

Mississippi	SEN	none	38	R					
Mississippi	HS	none	47	R					
Missouri	SEN	half	27	R	0	26.5	R	No Change	
Missouri	HS	all	33	R	0	27.6	R	No Change	
Montana	SEN	half	42	R	0	40.0	R	No Change	
Montana	HS	all	38	R	0	36.0	R	No Change	
Nebraska	SEN	half	100	D					
Nebraska	HS	unicameral							
Nevada	SEN	half	52	D	79	52.4	D		
Nevada	HS	all	64	D	46	50.0	R	Dem Loss	1
New Hampshire	SEN	all	46	R	0	37.5	R	No Change	
New Hampshire	HS	all	55	D	0	42.3	R	Dem Loss	1
New Jersey	SEN	none	60	D					
New Jersey	HS	none	60	D					
New Mexico	SEN	none	60	D					
New Mexico	HS	all	53	D	31	48.6	R	Dem Loss	
New York	SEN	all	52	D	46	49.2	R	Dem Loss	
New York	HS	all	71	D	100	65.3	D	No Change	
North Carolina	SEN	all	34	R	0	32.0	R	No Change	
North Carolina	HS	all	36	R	0	32.5	R	No Change	
North Dakota	SEN	half	30	R	0	34.0	R	No Change	-
North Dakota	HS	half	24	R	0	25.5	R	No Change	-
Ohio	SEN	half	30	R	0	24.2	R	No Change	
Ohio	HS	all	39	R	0	34.3	R	No Change	
Oklahoma	SEN	half	25	R	0	20.8	R	No Change	
Oklahoma	HS	all	29	R	0	26.7	R	No Change	
Oregon	SEN	half	53	D	96	56.7	D		-
Oregon	HS	all	57	D	54	51.7	D	No Change	
Pennsylvania	SEN	half	46	R	0	42.0	R	No Change	
Pennsylvania	HS	all	45	R	0	43.3	R	No Change	
Rhode Island	SEN	all	86	D	100	86.8	D		-
Rhode Island	HS	all	92	D	100	86.7	D	No Change	
South Carolina	SEN	none	39	R	0	39.1	R	No Change	
South Carolina	HS	all	37	R	0	33.9	R	No Change	
South Dakota	SEN	all	20	R	0	20.0	R	No Change	
South Dakota	HS	all	24	R	0	20.0	R	No Change	
Tennessee	SEN	half	21	R	0	12.1	R	No Change	
Tennessee	HS	all	28	R	0	23.2	R	No Change	
Texas	SEN	half	39	R	0	35.5	R	No Change	

Texas	HS	all	37	R	0	35.3	R	No Change	
Utah	SEN	half	17	R	0	13.8	R	No Change	
Utah	HS	all	19	R	0	13.3	R	No Change	
Vermont	SEN	all	75	D	100	73.3	D		
Vermont	HS	all	66	D	100	62.7	D	No Change	
Virginia	SEN	none	49	R					
Virginia	HS	none	33	R					
Washington	SEN	half	51	D	25	46.9	R	Dem Loss	
Washington	HS	all	56	D	8	45.9	R	Dem Loss	1
West Virginia	SEN	half	71	D	88	55.9	D		1
West Virginia	HS	all	53	D	0	37.0	R	Dem Loss	1
Wisconsin	SEN	half	45	R	5	45.5	R	No Change	
Wisconsin	HS	all	39	R	0	38.4	R	No Change	
Wyoming	SEN	half	13	R	0	13.3	R	No Change	
Wyoming	HS	all	13	R	0	11.7	R	No Change	

Table 2: Determinants of State Legislative Election Outcomes, 1968-2012: Full Model

Independent Variable	Coefficient	Standard Error	Variable Included in "Basic Model"
South	2.066*	0.921	
Border	0.143	0.847	
Trend	0.289	0.176	
Trend Squared	-0.008	0.008	
Trend Cubed	0.000	0.000	
South*Trend	-0.296*	0.169	
South*Trend Squared	0.008	0.009	
South*Trend Cubed	0.000	0.000	
Border*Trend	0.285*	0.171	
Border*Trend Squared	-0.019*	0.009	
Border*Trend Cubed	0.000*	0.000	
Vote Share$_{t-1}$	0.727*	0.005	X
Candidate Advantage	29.717*	0.345	X
Candidate Advantage$_{t-1}$	-23.805*	0.378	X
Incumbency Advantage	3.856*	0.143	X
Incumbency Advantage$_{t-1}$	-1.292*	0.140	X
Incumbency Advantage*Uncontested$_{t-1}$	-0.185	0.311	
Legislative Expenditures, Cat 2	0.056	0.256	X
Legislative Expenditures, Cat 3	0.462*	0.279	X
Legislative Expenditures, Cat 4	0.759*	0.315	X
Legislative Expenditures, Cat 5	-0.199	0.524	X
Incumbency*Legislative Expenditures, Cat 2	-0.182	0.135	X
Incumbency*Legislative Expenditures, Cat 3	0.422*	0.134	X
Incumbency*Legislative Expenditures, Cat 4	0.035	0.153	X
Incumbency*Legislative Expenditures, Cat 5	-0.717*	0.368	X
Second Term Incumbent	0.583*	0.135	
Third or More Term Incumbent	0.277*	0.132	
Dummy: Years 1979 to 1990	-1.362	1.076	
Dummy: Years 1991 to 2002	-2.491*	1.404	
Dummy: Years 2003 to 2014	-1.851	1.638	
Incumbency*Dummy: Years 1979 to 1990	0.885*	0.117	
Incumbency*Dummy: Years 1991 to 2002	1.092*	0.121	
Incumbency*Dummy: Years 2003 to 2014	0.313*	0.136	
Dummy: Years 1979 to 1990$_{t-1}$	-1.219	0.792	
Dummy: Years 1991 to 2002$_{t-1}$	-0.270	0.932	
Dummy: Years 2003 to 2014$_{t-1}$	-0.744	1.012	
Incumbency*Dummy: Years 1979 to 1990$_{t-1}$	-0.966*	0.127	
Incumbency*Dummy: Years 1991 to 2002$_{t-1}$	-0.577*	0.115	
Incumbency*Dummy: Years 2003 to 2014$_{t-1}$	-0.385*	0.140	
Other Chamber Advantage	4.987*	0.190	X
Other Chamber Advantage$_{t-1}$	-0.813*	0.221	X
Other Chamber Advantage*Uncontested$_{t-1}$	-0.401	0.779	
Prior Legislator Advantage	3.073*	0.137	X
Prior Legislator Advantage$_{t-1}$	-1.145*	0.175	X
Prior Legislator Advantage*Uncontested$_{t-1}$	-2.396*	0.919	
State Senate Election	0.006	0.090	X
Governor's Party	-0.080	0.194	X
Governor's Party$_{t-1}$	-0.058	0.137	X
State Midterm Penalty	-0.227	0.159	X

State Midterm Penalty$_{t-1}$	0.228*	0.109	X
State Economy	-0.018	0.013	X
State Economy$_{t-1}$	-0.001	0.010	X
President's Party	-1.342*	0.368	X
President's Party$_{t-1}$	1.045*	0.211	X
National Midterm Penalty	-0.674*	0.365	X
National Midterm Penalty$_{t-1}$	0.648*	0.175	X
Presidential Approval	0.073*	0.018	X
Presidential Approval$_{t-1}$	-0.046*	0.010	X
Congressional Vote Intention	0.081	0.050	X
Congressional Vote Intention$_{t-1}$	-0.076*	0.021	X
National Economy	0.227*	0.088	X
National Economy$_{t-1}$	-0.244*	0.041	X
Repeat Contest	-3.662*	0.414	
Repeat Contest*Vote Share$_{t-1}$	0.071*	0.008	
Repeat Loser	-1.032*	0.367	
Repeat Loser's Vote Share$_{t-1}$	0.022*	0.006	
Winner From Different Chamber	0.005	0.267	
Winner From Different Chamber's Vote Share$_{t-1}$	0.004	0.006	
Loser From Different Chamber	-1.951*	0.611	
Loser From Different Chamber's Vote Share$_{t-1}$	0.046*	0.010	
Nested or Identical Other Chamber: Candidate Advantage$_{t-1}$	-9.07*	0.496	
Nested or Identical Other Chamber: Incumbency$_{t-1}$	-1.783*	0.180	
Nested or Identical Other Chamber: Other Chamber Legislator$_{t-1}$	-2.466*	0.328	
Nested or Identical Other Chamber: Prior Legislator$_{t-1}$	-1.168*	0.449	
Nested or Identical Other Chamber: Vote Share$_{t-1}$	0.241*	0.011	
Dummy: Nested or Identical Other Chamber Exists$_{t-1}$	-12.556*	0.546	
Constant	11.521*	2.755	X
Year Level Error Term	0.515*	0.160	X
State-Year Level Error Term	1.901*	0.056	X
District Level Error Term	7.371*	0.021	X

Note: Cell entries in column two are the unstandardized regression coefficient produced from a hierarchical linear model with standard errors in parentheses in column three. The dependent variable is the percent of the two-party vote for the Democrats. N=77,844, Number of Years=45, Number of State-Years=982. Cases in chambers where the majority party had more than 84 percent of the vote were excluded, as were elections in which third party candidates received more than 20 percent of the vote.
*=p<.05, two-tailed hypothesis tests for all variables.

Table 3: Drop-1 Analyses: Model and Absolute Value of Prediction Error Averaged Across 1968-2012

Model	Brief Description of Model	Mean Absolute Prediction Error
1	Full Model, No Amelia	6.915
2	Basic Model, No Amelia	6.934
3	Full Model	6.263
4	Basic Model	6.291
5	Basic Model, Minus all Lagged Variables	6.384
6	Basic Model, Minus Lagged National Variables	6.372
7	Basic Model, Minus Incumbency/Expenditure Interactions	6.295
8	Basic Model, Minus State Economy/Governor's Party	6.290
9	Basic Model, Plus Regional Trends	6.310
10	Basic Model, Plus Identical/Nested Variables	6.256
11	Basic Model, Plus Candidate Specific Lagged Variables	6.286
12	Basic Model, Plus takes into account whether inc, other and prior were in uncontested races in the past	6.291
13	Basic Model, Plus Incumbent Terms Served	6.289
14	Basic Model, Plus Incumbency/Decade Interactions	6.313
15	Full Model, Minus all Lagged Xs (with a Few Exceptions)	6.448
16	Full Model, Minus Lagged National Variables	6.410
17	Full Model, Minus Incumbency/Expenditure Interactions	6.259
18	Full Model, Minus State Economy/Governor's Party	6.258
19	Full Model, Minus Regional Trends	6.276
20	Full Model, Minus Identical/Nested Variables	6.298
21	Full Model, Minus Candidate Specific Lagged Variables	6.267
22	Full Model, but doesn't take into account whether inc, other and prior were in uncontested races in the past	6.263
23	Full Model, Minus Incumbent Terms Served	6.263
24	Full Model, Minus Incumbency /Decade Interactions	6.267

Note: Each analysis recodes vote share to system missing for a biennium (i.e., a two year period such as elections taking place in 1981 and 1982), runs an analysis of the other years, saves predicted values, and compares those to the actual values. Column two gives a brief description of the model in question. Column three reports the average absolute value of prediction error for each election across all years for a model.

APPENDIX: VARIABLE DEFINITIONS

The variables used in Klarner's (2010a) forecasting model were all included as predictor variables, with the addition of many others.

Variables measured at the level of the election (i.e., one contest in one district in one election year) were as follows. The first was lagged vote share (lagged seat share in model five). A second variable measured "under-contestation" in contests which occurred in Free-For-All Multimember Districts (FFA-MMD). It was computed by first dividing the number of Democratic candidates by the number of seats up for election, and subtracting that by a quantity which was computed by dividing the number of Republican candidates by the number of seats up for election, and is called "Contestation Advantage." The third state legislative variable measured incumbency, and was coded "1" in single-member districts (SMDs) when the Democratic candidate was an incumbent, "-1" for a Republican incumbent, and "0" when either both or neither of the major party candidates were incumbents, and is referred to as "Incumbency Advantage." In FFA-MMDs, this and all other similar variables were measured by subtracting proportions, as with the under-contestation variable.[7] The fourth was coded "1" in SMDs when the Democratic candidate served in the other chamber of the state legislature immediately prior to the election, "-1" for such Republican candidates, and "0" otherwise, and is called "Other Chamber Advantage."[8] A fifth was coded analogously to incumbency, etc., if a candidate had held office in the legislature in the past, but not immediately in the past.

Because of the documented tendency of the incumbency effect to increase with the level of state legislative professionalization (Berry et al 2000, Carey, Niemi and Powell 2000), four dummy variables which measured the financing of the state legislature (adjusted for inflation) per capita were included, and each was also interacted with the incumbency variable. The first of these dummy variables represented legislatures that spend between $5.00 and $10.00 per capita, the second between $10.00 and $15.00, the third between $15.00 and $30.00, and the fourth between $30.00 and $127.00, the maximum observed to be spent on the state legislature per capita. The omitted category was between $1.60 and $5.00 spent on the state legislature. A series of dummy variables was utilized instead of one variable which measured spending on the state legislature to take the non-linear nature of the interaction into account, and the highly skewed nature of the data.

The possibility that incumbents increase their vote share over time was also considered. A variable that added the proportion of Democratic candidates who had served two terms minus the proportion of Republican candidates who had served two terms was included. An analogous variable measuring three terms or more was also included.

The incumbency effect has been documented to have first risen and then declined in the period examined here. Accordingly, a dummy variable coded "1" for the years 1979 to 1990 was created, as was a dummy variable coded "1" for 1991 to 2002, and a third variable coded "1" for 2003 to 2014. These in turn were interacted with the "Incumbency Advantage" variable, with the 1968 to 1978 as the reference period. Similar to other studies, incumbency went up in the 1979 to 1990 period, went up even more in the 1991 to 2002 period, and has been reduced greatly in the 2003 to 2012 period.

A control variable coded "1" when a race was in the state senate was included, coded "0" for state house races.

A dummy variable coded "1" when an election was preceded by the same two candidates in SMDs, and labeled "Repeat Contest." This was interacted with lagged vote share. "Repeat Loser" was a dummy variable that captured whether a loser ran again in the next election, and was also interacted with lagged vote share. Two dummy variables for a winner from the other chamber and a loser from the other chamber were also included and interacted with lagged vote share.

To help deal with missing vote share when prior elections were uncontested, vote share from the other legislative chamber in the instance of identical state senate and house districts were utilized. This was also done for state senate districts that had nested state house districts in them (but not vice versa). Whether such elections were uncontested or uncontested was controlled for (Nested or Identical Other Chamber: Candidate Advantage$_{t-1}$), as was incumbency, whether a candidate from the other chamber was running, and whether a prior legislator was running. All of these variables were coded the same as above, except applied to the other chamber of the legislature. A dummy variable coded "1" when such data existed was included (coded "0" otherwise).

[7] The more specific definition of this variable that also applies to FFA-MMDs is as follows: Democratic incumbents as a proportion of the number of seats in the district, subtracted by Republican incumbents as a proportion of seats. All variables which measured candidate attributes generalized to FFA-MMDs in this way.

[8] Distinguishing between state senate candidates who served in the state house and state house candidates who served in the state senate revealed no difference in impact.

The impact of four national level factors examined in prior research was examined. The first was a variable capturing the midterm penalty against the party of the president, coded "1" in non-presidential election years when a Democrat was in office, "-1" in such years when a Republican was in office, and "0" in other years. The second was a variable reporting the percentage of people who reported they would be voting for the Democratic candidate in the upcoming U.S. House election. The third was a variable which measured the percentage of people approving of the job of the president when a Democrat was in office, or 100 minus the percentage approving of the job of the president when a Republican president was in office. A fourth measured the health of the economy with percent change in real disposable income per capita. Specifically, the percent change in real income growth for the last four quarters was computed (annualized, to make more intuitive). Because the last month for which national per capita income data was available was August 2014, quarters ran from December to February, March to May, etc., instead of as standard. A weighted average of these was then taken, with less weight given to the earlier quarters to reflect voters giving less weight to quarters further in time from the election (Hibbs 2012; Erikson and Wlezien 2012). The quarter closest to the election was given a weight of "4," the next "3," then "2," and the earliest quarter a weight of "1." This was also multiplied by "-1" for Republican presidents. Last, a variable coded "-1" in the presence of a Republican president and "1" for Democratic presidents was necessary to include as a control because of the last two variables.

State level variables analogous to the national level variables were also examined when available. First was a variable which measured the impact of a state midterm penalty, based on the party and term of the governor (Folke and Snyder 2012). The second was an economic variable measured analogously to the national economic variable for the same time period, and multiplied by "-1" when there was a Republican governor. Last, a control variable coded "-1" for Republican governors and "1" for Democratic governors was also included.

The health of state's economies was measured from quarterly real disposable per capita income data (seasonally adjusted), ending in the second quarter of the election year, the last quarter for which such data are available in 2014. Two measures were examined both looking at percent change in income. The first examined the percent change in per capita income from four quarters ago. The second was a weighted average of per capita income growth (annualized) in one quarter over the last four quarters, with weights of "4" given to the most recent change, "3" to the second most recent, then "2," then "1."

A dummy variable coded "1" for the eleven states of the confederacy was included (called "South"), as was a dummy variable coded "1" for Maryland, West Virginia, Kentucky, Missouri, and Oklahoma, which were defined as "Border" states. A trend variable coded "0" in 1968 and advancing by one every year was included, along with its squared and cubed components. These were all then interacted with both "South" and "Border." Squared and cubed trend terms were included to allow any overall trend during the time period to slow. Not allowing such slowing could be very damaging to predictions at the end of a time series.

Lagged versions of the dependent variable and all independent variables were also included in the models. Excluding lagged independent variables from a time series when they are statistically significant is not appropriate (De Boef and Keele 2008), but additionally, they often have substantive rationales for inclusion. The inclusion of a lagged dependent variable creates a model of change in vote share as the coefficient associated with lagged vote share approaches "1" as occurred in the models examined here. Accordingly, change in independent variables must also be assessed, meaning that lagged independent variables should also be included. How the national midterm penalty works is a good illustration of this. Alabama had state legislative elections in 1986 and 1990, but not 1988. In both years, it was a midterm penalty for the Republicans, so we would not expect much of a change in Democratic vote share between the two years because of the midterm penalty (represented by a variable introduced below called "Midterm penalty" that takes on a value of "1" in a midterm with a Democratic president, "-1" with a Republican, and "0" when not a midterm). The sum of the coefficients associated with the lagged and non-lagged midterm penalty variable should therefore cancel out to an extent, as the lagged coefficient would be positive in sign while the non-lagged coattail would be negative. But from 1990 to 1994, we would expect to see a swing in a Republican direction, and the sum of the coattails would be a negative quantity larger in magnitude than the non-lagged coefficient alone. State legislative chambers with two year cycles would have much different scenarios that would be captured by this modeling strategy. Generally, the lagged components will take on signs that are the opposite of their non-lagged components, unless an effect continues for more than one period.

Lagged state legislative vote share was coded "100" when no Republican candidates ran in the prior election and "0" if no Democratic candidates ran in the prior election. The assumption that prior uncontested elections were equivalent to elections in which one party received all the votes was relaxed with the inclusion of lagged contestation advantage, defined above.

The lagged versions of all three of the above variables (Incumbency Advantage, Other Chamber Advantage and Prior Legislator Advantage) were interacted with a variable that indicated how uncontested an election was last time. For incumbency, this would be called "Incumbency Advantage*Uncontested $_{t\text{-}1.}$" The motivation for this was that lagged incumbency (as an example) would plausibly have a much different impact on vote share if the last election was uncontested versus not. For a SMD, this variable would be equal to lagged "Incumbency Advantage" if only one major party candidate ran. For a two seat FFA-MMD with under-contestation (say one non-incumbent Democrat and two Republicans, both of whom were incumbents) the incumbency score of each party would be computed separately (that is, the number of candidates of a party who were incumbents divided by the number of seats to be won), and then would be subtracted by the following quantity. The quantity would equal one minus the number of candidates of the other party divided by the number of seats to be won. So in the example above, "Incumbency Advantage*Uncontested$_{t\text{-}1}$" would have a score of "-.5" while "Incumbency Advantage$_{t\text{-}1}$" would still have a score of "-1." If a two seat FFA-MMD had two incumbent Republicans running against no Democrats, "Incumbency Advantage*Uncontested$_{t\text{-}1}$" would have a score of "-1."

Because of redistricting, missing data was a problem for the district-level lagged variables. Extensive archival research and consultation with maps identified when districts were redrawn. Because of court ordered inter-censual redistricting on the one hand, and some districts literally not being redrawn ("one person one vote" rules for SLDs are laxer than those for U.S. House districts, even today), one cannot merely assume that all districts in years ending in "1" or "2" are redrawn and districts in other years are not redrawn.