A Long-Range, State-Level Presidential Election Forecast Model

Jay A. DeSart
Utah Valley University

Abstract
This paper presents an adaptation of the DeSart and Holbrook presidential election forecast model for the purpose of making longer-range forecasts of presidential elections up to a year in advance of the election. Relying upon state electoral histories, home state advantage, and regime age variables, the model produces in-sample forecasts similar to that of the DeSart and Holbrook September forecast. On the basis of this model, a series of forecasts are generated for the 2016 election from October 2015 through August 2016.

A Long-Range, State-Level
Presidential Election Forecast Model

Election forecasting has grown into a significant subfield of political science over the past decade or so. When Tom Holbrook and I first published our own election forecast model (1999), only a handful of models existed and ours was one of the first that was specifically created to generate state-level forecasts for the purpose of generating a forecast for the Electoral College outcome.

Now, with the much greater availability of polling data at both the national and state levels, and increasingly sophisticated methodologies, election forecasting has grown into a cottage industry that has attracted a great deal of attention, and even a non-academic statistician like Nate Silver has practically become a household name during election seasons. In the past couple of elections we've seen a significant expansion in the development of election forecast models to predict the outcome of American presidential elections. Not only has the number of models grown, but they've also become increasingly complex and sophisticated in their nature.

In 2004, the first Symposium in on presidential election forecasts in *PS: Political Science and Politics* reported predictions from 7 different academic models (Campbell, 2004). In 2008, it reported on 9 models (Campbell, 2008). In 2012, that number had grown to 13 (Campbell, 2012), and the picture presented from this last set of forecasts suggested that the race between Barack Obama and Mitt Romney would likely be a nail-biter. A number of models predicted an Obama victory by margins ranging from about a half of a percentage point to a rather comfortable margin of over 7½ percentage points. But making
the picture a bit fuzzier was the fact that a number of models also generated predictions that Romney would win, some by a margin as big as 6 percentage points.

In the midst of all of the mixed signals being sent by this explosion in the number of models, there is one significant shortcoming that we must face: The lead time for such models has appeared to have an upper limit. Most models rely upon polling data and/or economic indicators at both the national and state levels to generate those forecasts and, unfortunately, there has been a paucity of polling data very far in advance of the elections, and the important economic variables featured in many models do not get measured and made available until a few months ahead of the elections.

The model developed by Thomas Holbrook and myself is no exception. The critical variable in our model is state-level trial heat polls. One significant limitation of our model is that such polling data is typically not widely collected and available until September of the election year, which effectively narrows the lead time for our prediction to roughly one month before the election, at least a full month behind most other prominent academic forecasts models.

Currently, the academic forecast model of note that has one of the longest lead times is that of Helmut Norpoth. His Primary Model generates a prediction of the outcome based on the candidates' performance in the New Hampshire Primary (Norpoth, 2004). The logic of his model is quite simple: The more divisive a party's nomination contests are, the more difficulty its eventual nominee will have in leading a united party and securing enough votes to win the general election. With this model Norpoth has been able to generate reasonably accurate estimates of the national popular vote 10 months in advance of the
election. Furthermore, Norpoth has been able to extend the lead time of his forecast by examining early polls of the New Hampshire primary electorates to attempt to project a winner even further in advance of the election, depending when such polls come out.

Recently, Norpoth outdid himself by developing an even longer range forecast model that would purportedly generate a prediction as long as 4 years in advance of the election (Norpoth, 2013). In other words, relying upon a “Time For Change” sentiment advanced by Alan Abramowitz (1988) in his own forecast model, Norpoth argues that Presidential Elections go through cycles, wherein the longer a party occupies the White House, the more difficult it is for them to hold on to it as regime fatigue sets in with the voters. Based on this model, which simply looks at the national-level outcome of the past two Presidential Elections, he predicts that the Republican candidate should be expected to receive 51.4% of the national popular two-party vote in 2016.

This paper represents my own attempt at pushing the lead-time envelope by adapting the state-level forecast model developed by Tom Holbrook and myself to generate a prediction up to a year or more in advance of the election. Before I can present that model, however, it would be most helpful to examine the short-range model upon which it is based.

**The DeSart and Holbrook Model**

The forecast model that we initially developed after the 1996 election (Holbrook and DeSart, 1999) was a simple and quite parsimonious model comprised of just two variables: the average Democratic share of support among the major party candidates in all trial-heat polls taken in each state during the month of September (POLLS), and the
average Democratic share of the two party popular vote across the two previous elections (PRIOR VOTE). These two variables are meant to capture both the short-term and long-term influences that determine the election outcomes in each of the states. The model, therefore, is represented by the following equation:

\[ VOTE_i = \alpha + \beta_1(POLL)_i + \beta_2(PRIOR VOTE)_i \]

This model generates predictions of the Democratic share of the two-party popular vote (VOTE) in each of the states, i.

The poll variable is derived from data obtained from various internet sources like NationalJournal.com’s PollTrack, PollingReport.com, and Pollster.com. In our previous work, we have shown that these September polls generally are a fairly good indicator of how the election will turn out in each of the states (Holbrook and DeSart, 1999; DeSart and Holbrook, 2003). Indeed, it is the case that candidates win the states in which they hold a significant lead in the polls in the month of September. The PRIOR VOTE variable is intended simply as a measure indicating the overall partisan tendency of a state. It serves as a stabilizing influence in the model. While the polls might respond to the short term stimuli of the specific campaign, the prior vote variable reflects the tendency for states to regress back to their typical behavior on Election Day.

Given a less than stellar performance of this model in its first attempt at generating an \textit{a priori} forecast of the 2000 election, we went back to the drawing board and made some adjustments to it (DeSart and Holbrook, 2003). In spite of the fact that it did quite well in 2004, we did modify the model in advance of the 2008 election by adding a third
term, a national poll variable and extending the time frame of the PRIOR VOTE variable from the average of the two previous elections to four (DeSart and Holbrook, 2010).

The reasoning behind making these changes was fairly straightforward. First, the change to the PRIOR VOTE variable would help us mitigate the effects of any home-state advantage that might have been enjoyed by a candidate in a previous election and “normalizing” the average somewhat. This problem was evidenced by the fact that the model over-estimated the Democratic share of the vote in Arkansas in both 2000 and 2004, no doubt due to Clinton’s strong “favorite son” showing in that state in 1992 and 1996.

The purpose of the national poll variable was to address an issue that we first identified in our original presentation of the model (Holbrook and DeSart, 1999). In that analysis we determined that there were significant year-specific effects in the forecasts that weren’t being picked up by the state-level polls. We added the national poll variable in an attempt to capture those year-specific contextual effects.

Thus, the model we used to generate state-level forecasts for both the 2008 and 2012 elections was:

\[ \text{VOTE}_i = \alpha + \beta_1(POLL)_i + \beta_2(\text{PRIOR VOTE})_i + \beta_3(\text{NATIONAL POLLS})_i \]

From those state-level forecasts, we can then generate predictions of both the Electoral College vote and national popular vote. For the Electoral College forecast it is a simple matter of awarding a state’s Electoral College votes to the candidate that the model’s point estimate suggests will win in that state.

The national popular vote extrapolation is a little more complex. Since each state’s contribution to the total number of votes cast in the election varies considerably according
to its population and level of turnout, simply averaging the state-level predictions across the 50 states would produce a biased estimate favoring the less populated/lower turnout states. To account for this, each state needs to be weighted according to its overall contribution to the total number of votes cast nationwide. To calculate this weight for the purposes of a prediction, a state’s total number of votes in the previous election is taken as a proportion of the total number of votes cast in those elections.\(^1\) The national popular vote forecast is thus a simple matter of taking the weighted total of the state-level forecasts. With this model, then, we can generate a fairly accurate estimate of the outcome a month in advance of the election, once all of the polls in September have been administered and released.

**Data Availability and Longer Range Forecasts**

As well as this model has performed, its biggest drawback has been that it is limited by the fact that its most important variable, the state-level polls, are simply not available for all 50 states prior to September. As such, it comes out too late to be included in the APSA’s quadrennial forecast symposium. Finding a way to break past the time constraint imposed by the unavailability of data, would require us to find alternative measures for our variables; measures that are available much longer in advance of the election.

As it turns out, the solution to this problem presented itself to us in the midst of the 2012 campaign. For the first time since the beginning of our efforts, September came and

---

\(^1\) Previously, we have used a state’s average vote total across the two previous elections in an effort to mitigate any potential issues where a particularly contentious down-ticket race in the previous election would spike turnout, however, further analysis showed that any such effect was negligible. Indeed, averaging turnout across two previous elections actually tended to produce less accurate estimates of the national popular vote.
went and there were a number of states in which no polls had been conducted. Left without a complete set of data with which to generate predictions, we imputed the POLL variable by averaging poll figures from the 3 or 4 states that were “most similar” to them based on their outcomes in 2008. This proved to be a fairly effective strategy because it allowed us to generate a forecast that missed the national popular vote result by only 1.1% and scored a perfect 51 out of 51 in predicting the state-level winners (and the District of Columbia).

Of course, that still doesn’t solve the current problem because it still requires us to have state-level polling data which simply isn’t available. However, it does suggest that another way to do a long-range imputation of a state’s short-term partisan proclivity is to simply rely upon its result from the previous election. This does, of course, change the nature of the variable somewhat. The purpose of the POLL variable is to capture the short-term influences in each of the states. Can that really be picked up by simply looking at how the state voted in the previous election? Probably not, and it is obviously one of the risks of producing a long-range forecast.

The real question, however, is how well does a state’s previous result help us know how it will vote in the next? Simply looking at the bivariate correlation between the states’ results with that of their result in the previous election suggests that it doesn’t work very well at all. For the elections of 1952-2012, the Pearson’s correlation is only .48. However, if we drill down a bit further into the data we can see that not all elections are created equally in terms of their ability to predict the outcome of the next.

Figure 1 presents the simple bivariate scatterplots for each pair of elections from 1952 to 2012. One can see that for a number of elections, the previous result does not do
very well at all in indicating what will happen the next time around. However, it is also very clear to see that there is a definite pattern as to when the previous result works and when it doesn’t.

The plots from 1960 to 1976 show a tremendous amount of flux from one election to the next, but from that point on the plots have settled into fairly tight and predictable patterns. It is probably no coincidence that the 1960s and 70s have been identified as the time during which the US went through a significant partisan realignment (Carmines and Stimson, 1980, Petrocik, 1981). Indeed, this is pretty clear evidence of the partisan upheaval during that time. More to the point, this shows that during periods of partisan realignment, a state’s prior electoral history is understandably not a very good indicator of how hit will go in the next election.

However, we can also see that once the new partisan alignment settled in, state outcomes have become quite predictable simply by looking at the previous election. Since 1980, state outcomes have been correlated with the result from the previous election at an average of .89, and since 1996 the average correlation is .94.

What all of this suggests is that during periods of “normal” partisan alignments state outcomes are fairly consistent from one election to the next, and can be quite useful to project the outcome in the next election. Even so, they are likely not going to be very good to pick up the short-term influences that will be unique from one election to the next. We will still need to find some way of capturing the unique effects of each election context if we are going to be able to generate accurate predictions of outcomes.
Recall that one way in which we capture the unique contextual effects of each election year in our model is a national poll variable. Fortunately, long range national match-up polls are far more common many months in advance of the election than are state-level polls.

It is difficult to find long-range national poll data prior to 2000. The PollingReport.com archives only have national polling data back to 1998. Supplementing this with some relatively limited 1996 data from Gallup. The resulting dataset I was able to compile consists of national polling data for each month from October of the previous year up through September of the election year for each election from 1992 to 2012. Figure 2 presents the scatterplot showing the relationship between the national popular vote result and the average national match-up polls between the two eventual nominees taken in each month. Granted there’s only five data points in each of the plots, but they clearly demonstrate an obvious pattern of a strong positive association. Indeed, Pearson’s $r$ ranges from a low of .75 in February and March to a high of .92 in September. What this suggests is that while certainly not perfect or as strong as those taken much later, national polls taken up to a year ahead of the election appear to be somewhat indicative of what will happen at the polls the following year, and may indeed help generate reasonably accurate estimates.

**Building a model**

In an analysis I’ve presented elsewhere (DeSart, 2015) I ultimately concluded that a simple four variable model can be used to generate reasonably accurate predictions of
state-level presidential election outcomes. To generate a prediction for each state, \( i \), the model is:

\[
VOTE_i = \alpha + \beta_1(PRIOR\ RESULT)_i + \beta_2(NATIONAL\ POLLS)_i + \\
\beta_3(HOME\ STATE)_i + \beta_2(REGIME\ AGE)_i
\]

PRIOR RESULT is simply the Democratic share of the 2-party popular vote (2PPV) in the previous presidential election, and NATIONAL POLLS is the average Democratic 2-party share of polls taken within a particular month. The other two terms in the model are additional terms that are necessary to account for the longer lead time and the absence of state-level polling data that can account for specific state-level effects that can't be captured by the national polling data.

*Home State Advantage*

Perhaps one of the most important, and most easily addressed, factors is the home state advantage. Previous authors have noted a significant home state advantage for presidential candidates in explaining state-level outcomes (Holbrook, 1991; Campbell, 1992). Accounting for this in our September Forecast Model is typically not necessary, because it most likely already comes through in the September polls. When we were developing our model, we consistently found that a Home State dummy variable failed to achieve statistical significance. Simply put, voters in a state most likely already know that a candidate is from their state in September, and any such advantage that brings will already be reflected in how they respond to a trial heat poll.

For a long-range forecast, however, the idea of a home state advantage is problematic for two reasons: First, the lack of state-level polling data means that such an
effect is not likely to be picked up in the model unless we specifically specify for it. The second is that a year in advance of the election we often do not know for sure who the eventual nominees will be, let alone the state from which they hail.

These problems become exacerbated by the fact that our long-range surrogate for the state poll variable, PREVIOUS RESULT, is doubly sensitive to the influence of this effect because not only does it not pick up the effect it will have in the upcoming election, but it is retroactively affected by it given who the candidates were in the previous election and where they were from. Therefore, any attempt to specify it in this model must account for both affects.

To do this, I created a simple dummy variable that would attempt to capture these fluctuations in the home state advantage from one election to the next. Clearly, we would expect the support for a candidate to go up in the state wherein they reside, so a simple dummy variable that indicates that the state is the candidate’s home state would work. However, since the dependent variable in the model is the Democratic share of the two-party popular vote (2PPV), it needs to be adjusted to whether or not the state is the home state of the Democratic candidate or the Republican candidate to correctly capture the estimated effect. We would expect the Democratic share of the 2PPV to increase in the home state of the Democratic candidate and decrease in the home state of the Republican candidate. So, to capture this effect the home state of the Democratic candidate in any given election is coded 1, and the home state of the Republican candidate is coded -1, to keep it consistent.
In addition, the retroactive home state effect needs to be accounted for as well given the PREVIOUS RESULT variable. Therefore, the opposite coding scheme is used to account for the expected decline the Democratic share of the 2PPV in the home state of the Democratic candidate in the previous election and the expected increase of the dependent variable in the home state of the Republican candidate in the previous election. If we have a president running for reelection, like we did in 2004 and 2012, those increases and decreases would not be present. Therefore, Texas in 2004 and Illinois in 2012 are coded 0, along with the rest of the non-home state states.

*Regime Age and "Time for Change"

One additional factor that needs to be addressed given the lack of state-level polling data is the impact of a “Time for Change” influence. Both Abramowitz (1988) and Norpoth (2013) have demonstrated that the longer a party holds on to the White House, the more difficult it becomes for them to hold on to it. Again, in our September Model, we assume that we are capturing the effects of this factor through the use of the poll variables, since if there is indeed such an effect, it would most likely exhibit itself in the responses of voters a month or two in advance of the election. It may be picked up somewhat by the NATIONAL POLL variable, but to be certain it would make sense to add a term to the model to account for this.

A simple variable which counts the number of consecutive terms one party has held the White House going into the election can be used to operationalize the age of a party’s occupancy of the White House. However, like the Home State variable, this will need to be
adjusted depending upon whether or not the current occupant of the White House is a Democrat or a Republican. For example, the more consecutive terms the Democrats occupy the White House, the lower we would expect the Democratic candidate to receive in the current election. On the other hand, the more consecutive terms a Republican occupies the White House, we would expect the Democratic candidate to receive a higher percentage of the vote. Therefore, to keep the coding consistent with our expectations the REGIME AGE variable added to the model is the count variable mentioned above, but given a negative sign if we are counting consecutive Republican terms and positive if we are counting consecutive Democratic terms.

With these four variables in the model, I generated coefficients using the average of the national polls for each month from the prior October through August. We would expect the PRIOR RESULT and HOME STATE variables would have fairly consistent effects from one month to the next, but it is reasonable to expect that the NATIONAL POLLS and REGIME AGE effects may show some variability over the course of the campaign. We would expect this given that details about the candidates’ qualities and proposals are coming to light over the course of the campaign, and the inevitable focus on the “failures” of the current administration will be discussed by those in the opposition. Those coefficients, along with the corresponding model performance statistics, are presented in Table 1.

**Model Performance**

Each of the monthly models are fairly consistent with what I found in my initial presentation last Fall. All four variables achieve statistical significance at the .01 level. As
expected, the PRIOR RESULT and HOME STATE variables both remain fairly stable from one month to the next, with HOME STATE showing that a candidate can expect around a three point increase in the vote in his or her home state.

On the other hand, the POLLS and REGIME AGE variables vary a bit from one month to the next. This was to be expected, since both are intended to measure the election context and the polls are likely to be influenced by at least some of the factors that are presumed to be behind the REGIME AGE variable.

Overall, we can see that over each of the 12 months the model produces reasonably accurate in-sample forecasts. Model R²s are consistently around .89 for each monthly model. The standard error of the estimate hovers around 3.20. By way of comparison, this R² is the same for that of the DeSart and Holbrook September Model. Perhaps not surprisingly, the September model has a slightly lower standard error of the estimate, 2.95.

For the specific state-level predictions, we find an average mean absolute error of around 2.5 points, which is slightly higher than that of our September model (2.30). If we simply use the point-estimates of the state-level predictions and determine winners and losers, the long-range models for each month correctly predict the winners of around 92% of the states. This is only slightly lower than that for our September model (93%).

There is a bit more difference in the accuracy of the national-level projections from the various monthly models and our September model. While an average mean absolute error of around 1.15 is reasonably good, it is roughly two times higher than that for our September model. This is, of course, not surprising given the longer lead-time of the long-range models.
What is interesting, however, is that the model seems to perform somewhat better in October, November and December of the year prior to the election than it does from January onward. For those three months, the $R^2$, standard error of the estimate, mean absolute error, and states correctly predicted are, for the most part, modestly better than they are for January onward. While these differences are not great, and one must be cautious to not overstate those differences, but it does suggest that this long-range model may work a little better the longer the range is.

The model is primarily designed to assess the context of the campaign, even before the nominees have been determined. The campaign enters a period of flux once the primaries and caucuses begin in January and February. Candidates rise and fall as the voters begin to see them perform on the electoral stage. The model’s accuracy ticks back up modestly in May and June, presumably when the nominations are locked up and voters can see the November landscape more clearly before the conventions start and temporarily skew the view.

Figure 1 seems to support this notion, with the polls appearing to be less predictive of the outcome from January through April and then again after June until they snap into line in September. Figure 3 also demonstrates how the REGIME AGE variable performs relative to the NATIONAL POLLS variable. REGIME AGE is supposed to tap into the context of the election year and it increases in importance in the model when the polls fall into flux. REGIME AGE, to be sure, is a fairly blunt instrument with which to measure context and certainly by July, when the candidates are known and the race starts to take shape, it declines in its utility. But, then again, the point of this model is to serve as a long-range
forecast. By July and August, other models come out which are much more fine-tuned in measuring the context in the short range.

**The 2016 Prediction**

Generating a prediction of how the election will turn out in November is a simple matter of plugging the values of the variables into the regression equations presented in Table 1. In the initial presentation of this model last Fall and for several months following, this process was complicated by the fact that we didn’t know for certain who the nominees would be. As a result, projections needed to be made in a matrix format showing the projections for each combination of candidate matchups, which I presented on a monthly basis on my research website. (http://research.uvu.edu/DeSart/forecasting).

This process was obviously made much simpler as candidates gradually dropped out of the race. Looking back now, we can examine what has transpired in the matchup between the two nominees. Figure 3 shows how the model has projected the result of a Clinton v. Trump matchup since October.

For the most part, the model has fairly consistently projected a very close election each month with the exception of February, April, and July. In those three months, the model projected a clear victory for Trump in both the popular vote as well as in the Electoral College. The projections for the rest of the months are a little less clear. The only month where the model projected an outright victory for Clinton over Trump in both the popular vote and the Electoral College was the very first one using polling data from
October 2015. Based on that forecast, Clinton would be expected to win 50.3% of the national two-party popular vote and win an Electoral College majority, 285 to 253².

Since that time, and excluding the months already mentioned where Trump victories were projected, the model has rather consistently projected a very close race in both the popular vote and Electoral College. Furthermore, in each of those instances, the model projected Electoral College misfires where Trump won narrow victories in the popular vote, but Clinton won Electoral College majorities. The final forecast, suggests that Clinton will win 49.82 of the national 2PPV, but win a slim Electoral College majority 272 to 266³.

---

² This projection is different from the one I presented on my website last November because after I generated that initial forecast, I located and incorporated 1996 polling data into the analysis. This changed the coefficients enough to yield a different forecast.

³ Given that there is still several days left in August 2016 at the time of this writing, the forecast may be modified somewhat as more August polling data come out.
Discussion

The goal of this work was to try extend the traditional lead-time of most election forecast models from a couple months to at least a year. While it succeeds in developing a model that generates such forecasts and does reasonably well in post-casting past elections, it remains to be seen whether it generates accurate forecasts. There are a few caveats to consider as we move forward to November 2016.

The first is that this model is only based on five elections worth of data. While that results in 250 data points, it is still uncertain that those five election contexts constitute a significant enough representation of the variation that is likely to occur across time. Furthermore, the analysis of Figure 1 suggests that this model would only work reasonably well during periods of relatively stable partisan alignments. During periods of partisan upheaval, it is quite likely that this model would generate wildly incorrect predictions.

In addition, the model does not have the ability to take into account any idiosyncratic factors that may make any given election unique. For example, it seems fairly clear at this point that one of the things that clearly benefited Barack Obama in 2008 was the undoubtedly historic nature of his candidacy. As the first African-American candidate to win the nomination from a major party, he most likely was able to ride a wave of enthusiasm as people flocked to the polls to cast a historic vote. So, too, it might be the case that Hillary Clinton is an historic candidacy and might similarly generate a groundswell of support and could lead the forecast to be wrong.

Finally, and certainly not least, one cannot discount the extremely unusual nature of the race that has developed since October. Clearly, the historical circumstances leading
into 2016 would have suggested that this would likely be a year that the Republicans would regain control of the White House. An implicit assumption from one election to the next in this, and most, forecast models is that there are some things that would likely be fairly constant across elections. One of those things are candidates who are fairly evenly matched in terms of their experience and ability.

But this year breaks from that pattern. A highly fractious Republican Party made for a highly unusual nomination process. Perhaps seeing a tremendous opportunity for victory, the race attracted an especially large number of candidates across the range of an ideologically divided party. This deeply divided field seemed to allow one highly unorthodox candidate to rise to the top, a candidate who does not seem to match previous nominees in terms of experience, ability, and discipline that the parties have typically nominated in the past.

What seems very clear at this point is that this should have been a fairly easy contest for the Republicans this year, according to this model as well as others, and given the pattern of history wherein a third consecutive term for a party has been a fairly rare occurrence since 1952. What increasingly appears to be the case, however, is that the Republicans very well could let it slip from their grasp because of the particular set of hands to which they entrusted it. Instead of listening to arguments about size of those hands, perhaps Republican primary voters should have paid more attention to the strength of the grip those particular hands.

There is still time for events, campaign or otherwise, to dramatically change the context and completely render the forecast meaningless. But at this point, it is difficult to
see much that could turn this around for the Republicans. We will just have to see if this new model is reliable, or whether the attempt to extend the lead-time in this manner was foolhardy.
References


### Table 1: Monthly Long-Range Model Performance

<table>
<thead>
<tr>
<th></th>
<th>October</th>
<th>November</th>
<th>December</th>
<th>January</th>
<th>February</th>
<th>March</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Prior Result</strong></td>
<td>1.040</td>
<td>1.040</td>
<td>1.040</td>
<td>1.041</td>
<td>1.042</td>
<td>1.042</td>
</tr>
<tr>
<td><strong>Home State</strong></td>
<td>3.135</td>
<td>3.089</td>
<td>3.118</td>
<td>3.176</td>
<td>2.981</td>
<td>2.986</td>
</tr>
<tr>
<td><strong>National Polls</strong></td>
<td>0.607</td>
<td>0.468</td>
<td>0.443</td>
<td>0.588</td>
<td>0.486</td>
<td>0.744</td>
</tr>
<tr>
<td><strong>Regime Age</strong></td>
<td>-1.033</td>
<td>-1.455</td>
<td>-1.630</td>
<td>-1.663</td>
<td>-1.884</td>
<td>-2.341</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>R²</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.893</td>
<td>0.893</td>
<td>0.894</td>
<td>0.883</td>
<td>0.882</td>
<td>0.885</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>SE_y</th>
<th>x</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
</table>

| **Mean Absolute Error** | 2.403 | 2.398 | 2.385 | 2.506 | 2.519 | 2.491 |

<table>
<thead>
<tr>
<th><strong>States Correctly Predicted</strong></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Overall</strong></td>
<td>92%</td>
<td>92%</td>
<td>92%</td>
<td>92%</td>
<td>91%</td>
<td>91%</td>
</tr>
<tr>
<td>1996</td>
<td>86%</td>
<td>86%</td>
<td>86%</td>
<td>90%</td>
<td>86%</td>
<td>86%</td>
</tr>
<tr>
<td>2000</td>
<td>84%</td>
<td>84%</td>
<td>84%</td>
<td>84%</td>
<td>84%</td>
<td>84%</td>
</tr>
<tr>
<td>2004</td>
<td>94%</td>
<td>96%</td>
<td>94%</td>
<td>94%</td>
<td>94%</td>
<td>94%</td>
</tr>
<tr>
<td>2008</td>
<td>94%</td>
<td>94%</td>
<td>94%</td>
<td>94%</td>
<td>94%</td>
<td>92%</td>
</tr>
<tr>
<td>2012</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>98%</td>
<td>98%</td>
<td>98%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>National 2PPV Error</strong></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Mean (Abs)</strong></td>
<td>0.99</td>
<td>0.99</td>
<td>0.92</td>
<td>1.31</td>
<td>1.28</td>
<td>1.19</td>
</tr>
<tr>
<td>1996</td>
<td>-0.59</td>
<td>-0.24</td>
<td>-0.38</td>
<td>-1.47</td>
<td>-0.22</td>
<td>-0.09</td>
</tr>
<tr>
<td>2000</td>
<td>-1.95</td>
<td>-2.06</td>
<td>-1.97</td>
<td>-1.77</td>
<td>-1.99</td>
<td>-2.01</td>
</tr>
<tr>
<td>2004</td>
<td>1.12</td>
<td>1.44</td>
<td>1.14</td>
<td>1.37</td>
<td>2.67</td>
<td>2.50</td>
</tr>
<tr>
<td>2008</td>
<td>-0.33</td>
<td>-0.59</td>
<td>-0.35</td>
<td>-0.44</td>
<td>-1.38</td>
<td>-1.27</td>
</tr>
<tr>
<td>2012</td>
<td>0.94</td>
<td>0.99</td>
<td>0.92</td>
<td>1.51</td>
<td>1.28</td>
<td>0.07</td>
</tr>
</tbody>
</table>
### Table 1: Monthly Long-Range Model Performance

<table>
<thead>
<tr>
<th></th>
<th>April</th>
<th>May</th>
<th>June</th>
<th>July</th>
<th>August</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Prior Result</strong></td>
<td>1.042</td>
<td>1.040</td>
<td>1.040</td>
<td>1.042</td>
<td>1.042</td>
</tr>
<tr>
<td><strong>Home State</strong></td>
<td>2.958</td>
<td>3.089</td>
<td>3.039</td>
<td>2.943</td>
<td>2.953</td>
</tr>
<tr>
<td><strong>National Polls</strong></td>
<td>0.410</td>
<td>0.468</td>
<td>0.645</td>
<td>0.442</td>
<td>0.611</td>
</tr>
<tr>
<td><strong>Regime Age</strong></td>
<td>-2.120</td>
<td>-1.455</td>
<td>-1.702</td>
<td>-1.806</td>
<td>-1.837</td>
</tr>
<tr>
<td><strong>Constant</strong></td>
<td>-23.417</td>
<td>-25.539</td>
<td>-35.496</td>
<td>-25.306</td>
<td>-33.710</td>
</tr>
<tr>
<td>( R^2 )</td>
<td>0.884</td>
<td>0.891</td>
<td>0.894</td>
<td>0.884</td>
<td>0.885</td>
</tr>
<tr>
<td>( SE_{y</td>
<td>x} )</td>
<td>3.230</td>
<td>3.136</td>
<td>3.091</td>
<td>3.238</td>
</tr>
</tbody>
</table>

|                |      |      |      |      |        |
| **Mean Absolute Error** | 2.403 | 2.398| 2.385| 2.506 | 2.519  |

**States Correctly Predicted**

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>91%</td>
<td>86%</td>
<td>84%</td>
<td>94%</td>
<td>92%</td>
<td>100%</td>
</tr>
</tbody>
</table>

**National 2PPV Error**

<table>
<thead>
<tr>
<th>Year</th>
<th>Mean (Abs)</th>
<th>1996</th>
<th>2000</th>
<th>2004</th>
<th>2008</th>
<th>2012</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>1.21</td>
<td>1.03</td>
<td>1.26</td>
<td>1.22</td>
<td>1.32</td>
</tr>
</tbody>
</table>

Cell entries for each independent variable are unstandardized regression coefficients. For all coefficients \( p < .01 \)
Figure 1 – Continuity and Discontinuity in Statewide Presidential Election Outcomes, 1952-2012

Plots represent the correlations between the statewide Democratic share of the two-party popular vote of two successive elections.
Figure 2: Monthly Poll Averages and November Outcomes

- **Prior October** ($r = .77$)
- **Prior November** ($r = .85$)
- **December** ($r = .90$)
- **January** ($r = .82$)
Figure 3: Relative Importance of the Polls and Regime Age Across Monthly Models

Figures represent the absolute value of BETA coefficients for the NATIONAL POLLS and REGIME AGE coefficients in each monthly model.
Figure 4: Long-Range Model Monthly Projections