State Polls and National Forces: 
Forecasting Gubernatorial Election Outcomes

Jay A. DeSart
Utah Valley State

Abstract

This paper is a replication and extension of the DeSart and Holbrook presidential election forecast model to gubernatorial elections. It examines a simple model with three variables: September pre-election polls, prior election outcomes and presidential approval. The model generates reasonable forecasts, but falls quite short of its presidential election counterpart. However, it does show that presidential approval in the third quarter of the election year does indeed have a significant contribution to the model.


This manuscript is a draft of a work in progress. Comments, suggestions and questions are welcome and may be directed to the author via email at desartja@uvsc.edu
State Polls and National Forces:
Forecasting Gubernatorial Election Outcomes

Election forecasting has practically developed into a subfield in its own right within the general field of election studies. Election forecasters have been able to develop models that (usually) generate amazingly accurate predictions of the outcomes of presidential elections, months in advance of the election. (Rosenstone, 1983; Abramowitz, 1988; Lewis-Beck and Rice, 1992; Campbell, 1992; Holbrook, 1996; Lewis-Beck and Tien, 1996; Wlezien and Erikson, 1996; Campbell and Garand, 2000; Norpoth, 2000; Jones, 2001)

These forecast models have been largely based on a well-developed understanding of the determinants of such election outcomes. Thus, the enterprise of election forecasting is far more than a simple academic exercise for entertainment purposes (who will have the closest prediction?), but rather tests of alternative explanations of election outcomes.

However, this enterprise is largely limited to attempts to forecast national level elections: predicting presidential election outcomes or, occasionally, congressional seats gained/lost by the president’s party. Explanations of gubernatorial election outcomes abound in the literature examining such factors as incumbency and challenger quality (Squire, 1992; King, 2001), economic conditions (Chubb, 1988; Stein, 1990; Howell and Vanderleeuw, 1990; Leyden and Borrelli, 1995; Svoboda, 1995; Niemi, Stanley & Vogel, 1995; Atkeson and Partin, 1995; Carsey and Wright, 1998), issues (Kone and Winters, 1993; Cook, Jelen & Wilcox, 1994; Niemi, Stanley & Vogel, 1995; Lowery, Alt & Ferree, 1998), and the all-important question of the influence, or lack thereof, of national-
level forces (Holbrook, 1987; Chubb, 1988; Tompkins, 1988; Simon, 1989; Atkeson and Partin, 1995; Carsey and Wright, 1998).

This work is an attempt to cut through the forest of literature on the explanations of gubernatorial elections and develop a simple forecast model designed to generate predictions of those election outcomes. The basis of the model is a presidential election forecast model developed by Tom Holbrook and myself. (Holbrook and DeSart, 1999; DeSart and Holbrook, 2003). This model generates state- and national-level predictions of the presidential election outcome by relying upon state-level trial-heat polling data. We have shown that polls conducted in each state during the September preceding the election do a remarkable job of generating fairly accurate forecasts of the actual presidential vote in those states. If they do such a good job forecasting the outcomes of the presidential races at the state-level, it begs the question as to whether they will perform just as well in generating predictions of gubernatorial election outcomes.

However, there is a lingering question in the literature on gubernatorial elections as to whether, and how much, national-level forces matter. Some research seems to indicate that gubernatorial elections are relatively isolated events somewhat insulated from national-level factors like presidential approval and the condition of the national economy (Tompkins, 1988; Howell and Vanderleeuw, 1990; Atkeson and Partin, 1995), while others suggest just the opposite (Holbrook, 1987; Chubb, 1988; Simon, 1989, Carsey and Wright, 1998).

This paper offers a modest volley into that debate by including a national-level variable, presidential approval, into the model. If national forces matter in gubernatorial
elections, then we should see presidential approval adding some predictive power to the model.

**The Model**

The presidential election forecast model developed by Tom Holbrook and myself (Holbrook and DeSart, 1999; DeSart and Holbrook, 2003) is very parsimonious with just two variables. The model generates the predicted Democratic share of the two-party presidential vote, \( \text{VOTE}_{it} \), in each state, \( i \), in each election year, \( t \), with the following equation:

\[
\text{VOTE}_{it} = \beta_1 \text{POLLS}_{it} + \beta_2 \text{PRIOR}_{it};
\]

where \( \text{POLLS}_{it} \) represents the average Democratic share of the two-party vote in each poll taken in September before the election, and \( \text{PRIOR}_{it} \) represents the average Democratic share of the two-party vote in each of the two preceding elections.

It is thus a simple matter of plugging in the corresponding equivalent variables for gubernatorial elections into the model. Its simplicity, in spite of the complexity of explanations of voter behavior and election outcomes, is grounded in the assumption that much of those factors are already accounted for in the primary variable in the model, \( \text{POLLS} \).

However, since we are dealing with subnational elections in the current enterprise, and in light of the prevailing debate about the influence of national-level forces on state elections, it might be helpful to include a national-level indicator to the model as well. Therefore, another model that I ultimately test in this paper is represented by the following equation:
\[ VOTE_{it} = \beta_1 POLLS_{it} + \beta_2 PRIOR_{it} + \beta_3 APPROVAL_{it}; \]

where APPROVAL represents the average of presidential approval ratings for the third quarter of the election year.

**The Data**

The main source for the key independent variable in the model, POLLS, is the same as it is for our presidential election model, NationalJournal.com’s PollTrack. Every gubernatorial election poll conducted during the month of September preceding the election were averaged for each state and each year. Since September polls were not conducted in every state in every year, those states are necessarily eliminated from the analysis. The resulting dataset includes 83 cases for the elections spanning the years 1998 through 2005.¹

The presidential approval ratings were obtained from PollingReport.com. All presidential approval polls reported for the third quarter of the year (July through September) were averaged to generate a value for APPROVAL for each year. However, since the White House has changed party control across the time frame of the analysis the approval variable needs to be modified somewhat. To make the variable consistent with the direction of the dependent variable, it is subtracted from 100 for the years 2001 through 2005 to reflect the Bush presidency (in effect turning it into a “disapproval” variable for a Republican presidency).

Results

Figure 1 presents the scatterplot of the September poll data with the actual outcomes in each of the sample states. It shows that while there is a fairly robust relationship between the two variables and the eventual outcome, there is also a fair amount of error as well. The Pearson’s r correlation between the two variables is .87. Compared to the data in our presidential election model (Pearson’s r = .96), this demonstrates that while there is a strong correlation between September polls and the eventual outcome in gubernatorial races, they may be slightly less predictive of the eventual outcome in those races than they are at the presidential level.

Turning to the actual models, the results of the analysis are presented in Table 1. The first three columns of Table 1 build the replication of our presidential election forecast model, which I refer to as the Basic Model. The first two columns present the univariate models of each of the variables in the basic model: POLLS and PRIOR.

These results show that there is an additional weakness to the gubernatorial forecast model in comparison to its presidential election counterpart. Not only is the predictive power of the September polls weaker in the gubernatorial election data, but the prior vote variable shows no predictive power at all, whether by itself or part of the multivariate basic model. While the prior vote variable is indeed the weaker of the two variables in the presidential election model, it at least provides some predictive ability.

Using only the prior vote variable give us only a slightly better than fifty-fifty chance of accurately predicting the winner (based on the predicted vote percent), and it actually drags the predictive power down when included in the Basic Model with the September polls variable. The standard error of the estimate increases slightly (3.933 to
3.956), and the percent of races correctly predicted declines modestly (85.5% to 83.1%) when PRIOR is included with POLLS in the Basic Model. This would seem to show that gubernatorial elections show a bit less stability in their results from one election to the next, and may be susceptible to the idiosyncracies (e.g. incumbency, etc) of each individual election campaign context more so than demonstrated in presidential elections. This would certainly seem to be the case given the importance that incumbency shows in various explanations of gubernatorial elections (Squire, 1992; King, 2001).

Ultimately, the analysis thus far shows that the Basic Model we developed for forecasting presidential election outcomes loses some of its predictive power when we move down to the gubernatorial level. On the other hand, this model still generates an accurate prediction of the outcome of the election in roughly 7 out of 8 gubernatorial races. That’s a nice record, but certainly one that can be improved upon.

The National Context

Clearly, each year is not created equal when it comes to its electoral context. In our work with the presidential election model, Holbrook and I found that including a dummy variable for each year added extra accuracy to the model (Holbrook and DeSart, 1999). The fourth column of Table 1 presents a similar application of that notion to the gubernatorial election model. By including dummy variables for the years 1999 through 2005 in the model we can account for the changing context across each year. In effect, these dummy variables tell us how the Democratic candidates fared in the gubernatorial elections in each year, compared to the baseline year of 1998.
The dummy variables all show a negative coefficient (showing that the yearly contexts all seemed to be less favorable to Democratic candidates, compared to 1998), however only two attain statistical significance, 2002 and 2004. In each of those, the negative impact is unmistakable. Those two years showed a clearly apparent pro-Republican influence, with Democratic candidates receiving less than 4% than the vote than they did in 1998. At the very least, this shows that something in the national context was exerting an influence across the state elections in those years.

As one would probably expect, this loading of variables into the model increases its accuracy. The $R^2$ increases from .76 for the Basic Model to .82 for the Year Specific Model. In addition, the adjusted $R^2$ climbs to .795 from .754. The standard error of the estimate and the mean absolute error in prediction both decrease slightly, and the percent of races correctly predicted increases modestly (83.1% to 84.3%) as two more races are correctly predicted with this model.

Of course, the fundamental weakness of the Year Specific Model is that we don’t know from one year to the next what the unique impact that year’s context will have upon the outcomes. Therefore it is of dubious usefulness as a true forecasting model. What we need is surrogate that we do have in advance that would work to try and capture the national political context.

The leading candidate for this variable would be presidential approval. To what extent does the president’s standing with the public provide an indicator of his party’s fate in the gubernatorial elections? The fifth column of Table 1 attempts to answer this question by replacing the year dummy variables with a single variable: average presidential approval in the third quarter (July through September) of the election year.
This has the nice feature of being readily available in advance of the election, thus giving the model the ability to generate true forecasts of the election.

This variable proves to perform quite nicely in place of the year dummies and demonstrates that the national context in general, and presidential approval specifically, appears to indeed influence the outcomes in gubernatorial elections. According to this finding, a seven to eight point shift in a president’s approval ratings results in a roughly one point gain or loss for his party’s gubernatorial candidates. Not a huge impact, to say the least, but it could be a key determinant in close races particularly in years where approval is particularly high (2002) or particularly low (2006).

The inclusion of this variable shows only a modest shift in the overall model performance compared to that of the Year Specific Model. The $R^2$ naturally declines with the swap out of one variable in place of seven, but adjusted $R^2$ shows no change. There is only a small change in both the standard error of the estimate and mean absolute error in prediction, and the percentage of races correctly predicted remains unchanged. So at the very least this model seems to do just as well as the Year Specific Model, but just does so more efficiently.

The final column of Table 1 addresses the continued failure of the prior vote variable to achieve significance in any of the models tested thus far. It begs the question: if it fails to achieve significance, why include it at all? To what extent does its inclusion in the model drag down its predictive power? The answer seems to suggest that, while it seems to fail to contribute much to the model, its exclusion decreases the model’s accuracy slightly. There are slight shifts in the diagnostics, but the direction of those changes are mixed. The adjusted $R^2$ and mean absolute error both suggest that the model
might be better off without PRIOR in the model. However, and perhaps more importantly, the percentage of races correctly predicted declines as well with the variable removed from the model.

This would seem to suggest that, rather than dropping the variable from the model entirely, tweaking it somewhat might make it more a more useful part of the model. This will be the approach in further analysis. Perhaps increasing the number of elections included in the average will smooth out some of the idiosyncrasies of individual prior elections and might give a more accurate reflection of the overall long-term partisanship of the state. That is, after all, the purpose it is intended fulfill in the presidential election model.

**Conclusion**

This paper has presented an application of the DeSart and Holbrook presidential election forecast model to gubernatorial elections. While each model generates predictions of election outcomes at the state-level, there are clear differences in its predictive abilities depending on whether we are forecasting outcomes in the elections for state executives versus that for the national executive.

Clearly, the model is much less accurate in forecasting gubernatorial election outcomes than it is for presidential elections. Further development of the model is necessary to get it to the level of predictive ability demonstrated by its presidential election counterpart. Perhaps including more history in the prior vote variable, or even including a measure of incumbency, may very well improve this model, and this will be the focus of future iterations of the model.
There is, however, an even more troubling problem with the model that makes it fall short of the utility of our presidential election model: the unavailability of polling data in some states. One thing that limits this model is that, unlike in presidential elections, pre-election polls are not conducted as regularly in gubernatorial elections, at least not during the month of September. One thing that we are able to do with our presidential forecast model is extrapolate national-level outcomes (both popular vote and electoral vote) from the state-level forecasts (DeSart and Holbrook, 2003). It would be useful to be able to generate national-level results in terms of gains/losses for the president’s party. However, unless a September poll is conducted in each state holding a gubernatorial election in any given year, we will continue to fall short of that goal. The one glimmer of hope in that regard is that the trend is moving in our favor. The number of states for which we are missing polling data in 1998 was seven. For 2002 it was down to four. We can only hope that the trend will continue for 2006, just in time for the model’s first real test of its forecasting ability.

Finally, the test of this model offers one more modest viewpoint to the debate over the influence of national forces in gubernatorial elections. The significance of the presidential approval variable seems to place this model squarely down on the side of those advocating the position that national forces do indeed matter. Future iterations of the model will also include national economic indicators as well, in an effort to further improve its predictive power.
References


Figure 1

September Polls and Election Results
Gubernatorial Elections 1998-2005

Data are converted to margins for clarity of presentation. Margins are Democratic percent minus Republican percent.
Table 1 – Gubernatorial Election Forecast Models

<table>
<thead>
<tr>
<th></th>
<th>Sept Polls</th>
<th>Prior Vote</th>
<th>Basic Model</th>
<th>Year Specific Model</th>
<th>Basic Model + Approval</th>
<th>Polls + Approval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sept. Polls</td>
<td>0.646*</td>
<td>—</td>
<td>0.646*</td>
<td>0.703*</td>
<td>0.682*</td>
<td>0.680*</td>
</tr>
<tr>
<td>Prior Vote</td>
<td>—</td>
<td>0.017</td>
<td>0.016</td>
<td>-0.028</td>
<td>-0.036</td>
<td>—</td>
</tr>
<tr>
<td>Year Dummies</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>1999</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>-2.251</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>2000</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>-2.346</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>2001</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>-1.827</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>2002</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>-4.615*</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>2003</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>-3.427</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>2004</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>-4.180*</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>2005</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>-0.138</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Approval</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>0.130*</td>
<td>0.123*</td>
</tr>
<tr>
<td>Constant</td>
<td>17.700*</td>
<td>47.303*</td>
<td>16.937*</td>
<td>18.967*</td>
<td>11.480*</td>
<td>10.179*</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>R²</th>
<th>Adjusted R²</th>
<th>SE_y/x</th>
<th>Mean Absolute Error</th>
<th>% Correct</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>.759</td>
<td>.760</td>
<td>3.933</td>
<td>2.855</td>
<td>85.5%</td>
</tr>
<tr>
<td>Adj R²</td>
<td>.757</td>
<td>.754</td>
<td>8.020</td>
<td>6.020</td>
<td>51.8%</td>
</tr>
<tr>
<td>SE_y/x</td>
<td>3.933</td>
<td>3.956</td>
<td>3.612</td>
<td>2.855</td>
<td>83.1%</td>
</tr>
<tr>
<td>Mean Absolute Error</td>
<td>2.855</td>
<td>2.637</td>
<td>3.609</td>
<td>2.637</td>
<td>84.3%</td>
</tr>
<tr>
<td>% Correct</td>
<td>85.5%</td>
<td>84.3%</td>
<td>81.9%</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figures in the top half of the table represent unstandardized regression coefficients.
N = 83
* = p < .01