## Data, Measures and Methods

# Looking back on the forecasting of the French legislative election

Martial Foucault[a,b]
[a]Department of Political Science, University of Montreal, C.P. 6128, succursale Centre-ville, Montreal, QC H3C 3J7, Canada.
[b]Visiting Professor at Science Po, CEE, France.
E-mail: martial.foucault@umontreal.ca

**Abstract:** This note offers a review of a forecasting model implemented for the 2012 legislative election. Based on real outcomes, I highlight to what extent this model provides better forecasting than preceding ones and how it could be improved for the next national elections. Introducing local data to forecast national electoral outcomes is no doubt a promising research avenue in such a field.
*French Politics* (2012) **10,** 383–388. doi:10.1057/fp.2012.22

**Keywords:** legislative election; forecasting; France

Predicting vote choice for French legislative elections is a complicated exercise due to the fact that few studies have put forth a stable, robust and high-quality specification. In a 2012 special edition of *French Politics*, I gave not only two predictions (48.22 per cent and 49.33 per cent for the right), but also an original method of prediction based on both local- and national-level data. The predictive model performed rather well, as the right-wing majority received a score of 49.43 per cent in the second round. Let us return to this model by discussing both its characteristics (Lewis-Beck, 2005) and its limits for the next legislative elections.

## Quality of Prediction

The predictive electoral model that I proposed aimed to calculate the vote percentage in the second round by the incumbent majority (Foucault, 2012). On the basis of an economic voting perspective, this model makes use of two categories of variables: economic variables (*département*-level unemployment rate and national GDP growth rate) and political variables (vote during the last national elections, ideological polarization by *department*, popularity of the prime minister and period of cohabitation). Let us return to the estimated form of these two specifications:

 

(*Model 1*)

$$\begin{aligned}\text{Vote}^{2012} =& 30.1 + 0.32\,\text{GDP} - 0.86\,\text{UNEMP}^{**} + 5,61\,\text{POPU}^{***} \\ &+ 0.33\,\text{PRE}^{***} + 0.37\,\text{LEG}^{***} - 0.29\,\text{STRENGTH}^{***} \\ &+ 0,03\,\text{TRIANG}^{***} - 3.78\,\text{DIV}^{***} - 15.2\,\text{TIMING}^{***} \\ &+ 0.37\,\text{TIMING}\times\text{PRE}^{***} + \phi_i + \omega_t + \varepsilon_{i,t}\end{aligned}$$

(*Model 2*)

$$\begin{aligned}Vote^{2012} =& 29.5 - 0.88\,\text{UNEMP}^{**} + 6.84\,\text{POPU}^{***} + 0.34\,\text{PRE}^{***} \\ &+ 0.36\,\text{LEG}^{***} - 0.29\,\text{STRENGTH}^{***} + 0.03\,\text{TRIANG}^{***} \\ &- 3.99\,\text{DIV}^{***} - 15.4\,\text{TIMING}^{***} + 0.37\,\text{TIMING}\times\text{PRE}^{***} \\ &+ \phi_i + \omega_t + \varepsilon_{i,t}\end{aligned}$$

The complete form of Model 1 led to a prediction of 48.22 per cent for the incumbent majority. Model 2 (which does not take the GDP growth rate into account) gave a prediction of 49.33 per cent. The results of the legislative elections held on 17 June 2012, as reported by the Ministry of the Interior (www.interieur.gouv.fr/elections), gave a score of 49.43 per cent for the right (including extreme right).

The first model (with an error of 1.2 percentage points) conforms to the final result of the election in the sense that it predicted a minority score for the incumbent majority. The second model is even closer to the final score, with an error of only 0.1 percentage points. The differences between these two models lie in the non-significance of the GDP variable in the first model and a larger coefficient for the variable measuring the popularity of the prime minister in the second model, with the other coefficients being rather close. Given that the quality of an electoral prediction is judged by its ability to be the closest to the final result and within the confidence interval of the margin of error, it is preferable in the future to use a specification that gives priority to the unemployment rate and not the rate of GDP growth. Despite a very weak correlation between the rate of GDP growth and unemployment rate, it is easy to argue that unemployment is a variable that affects French voters more than the national rate of GDP growth. Furthermore, the unemployment rate is calculated at the *department* level, whereas the GDP growth rate is calculated only at the national level. As expected, the more that the unemployment rate rises, the more that the incumbent majority is punished. Thus, it is preferable in the future to prioritize an economic variable such as unemployment rate to characterize local-level economic conditions. The larger estimated coefficient for the popularity variable shows the ability of this variable to capture the reward/punish effect when the GDP variable is removed. As the other

estimated coefficients are very close and that the *department*-level fixed effects vary little from one model to another, it is easy to understand why Model 2 gives a higher predicted vote share than Model 1.

The success of this predictive model should be seen in relation to criteria set forth for evaluating any form of electoral prediction. Lewis-Beck (2005) defined four criteria for evaluation: accuracy, lead, parsimony and reproducibility. With a score of 49.33 per cent and a prediction given in December 2011, my predictive model satisfies the first two criteria. The third criterion, parsimony, needs some explanation. Normally, electoral prediction models rely on time-series data such that one election corresponds to one observation. The idea is to estimate the score at election $t$ as a function of past parameters and as a consequence, these models rely on many past elections to come up with a large enough N to give robust estimates. Faced with this problem, many authors are forced to choose a specification with a small number of explanatory variables in order to not reduce the number of degrees of freedom (see Lebo and Norpoth, 2007 for a robust method of estimation). This is why the criterion of parsimony is indispensable in these models; otherwise, the number of degrees of freedom would be too small. In my model, the number of observations is higher due to the fact that the unit of observation is not the election ($n$), but the number of departments for an election ($n \times 96$); this is an original contribution to the literature on French elections. As a consequence, the criterion of parsimony is not applied *stricto sensu*. Conversely, the addition of independent variables not related to a well-established theoretical expectation could harm the model's performance by introducing a risk of correlation between these variables and the error term. With nine explanatory variables, this model offers the possibility of capturing both space and time effects. Finally, the criterion of reproducibility is entirely met in the sense that all the statistical data used are publically available from economic (INSEE) and political bodies (Ministry of the Interior, SOFRES).

## Limits of the Model

Despite the good performance overall of the predictive model, there are still some improvements in order. Among them, there are the value of fixed effects and the unemployment variable. The particularity of the specification used is based on the estimation of a spatial fixed effect (that is, the *department*) that allows us to capture all the non-observed effects by our independent variables. Even if this method has the advantage of controlling for characteristics of each *department*, it is difficult to demonstrate that these fixed effects are absolutely not correlated with the regressors. For example, a more industrially oriented *department* is more likely to confront negative variations in the unemployment

**Table 1:** Distribution of forecasting errors by *départements*

| *2012 vote* | *Forecasts* | *Error* | *Absolute error* | *Département* |
|---|---|---|---|---|
| 66.05 | 50.64 | −15.41 | 15.41 | VAUCLUSE |
| 59.26 | 46.21 | −13.05 | 13.05 | EURE |
| 58.29 | 48.77 | −9.52 | 9.52 | HAUTE LOIRE |
| 44.04 | 35.26 | −8.78 | 8.78 | PAS DE CALAIS |
| 56.16 | 47.83 | −8.33 | 8.33 | MARNE |
| 56.02 | 63.73 | 7.71 | 7.71 | HAUTE SAVOIE |
| 34.18 | 41.93 | 7.75 | 7.75 | HAUTES PYRENEES |
| 43.19 | 52.10 | 8.91 | 8.91 | INDRE |
| 38.21 | 49.67 | 11.46 | 11.46 | PARIS |
| 46.76 | 63.70 | 16.94 | 16.94 | MAYENNE |
| 52.44 | 52.17 | −0.27 | 0.27 | MOSELLE |
| 52.2 | 52.04 | −0.16 | 0.16 | EURE ET LOIR |
| 48.57 | 48.54 | −0.03 | 0.03 | MAINE ET LOIRE |
| 52.32 | 52.60 | 0.28 | 0.28 | YONNE |
| 51.43 | 51.75 | 0.32 | 0.32 | COTE D'OR |

rate than one in the Ile-de-France region. Thus, there is an inherent limit in this type of predictive model: the impossibility to capture fixed effects that have an interpretive value. As a consequence, the overall quality of the prediction includes prediction errors for each *department*, but these largely cancel each other out. An illustration of this situation is given by Table 1. The first five lines of the table indicate the *departments* for which the model underestimated the score of the incumbent majority, while the next five lines correspond to the *departments* where the model overestimated the right's actual vote share. Finally, at the end of the table, the model seems to have done well in predicting the vote share in five *departments* (Moselle, Eure-et-Loir, Maine-et-Loire, Côte d'Or and Yonne).

The second limit concerns the use of the unemployment variable whose inclusion is far the better proxy of economic conditions. In my model, I have selected the departmental unemployment rate between the first quarter and the fourth quarter before the election as a proxy for economic conditions. The choice of this variable could be revised in the future because the labor market dynamic could be affected by either the number of successive trimesters where the unemployment rate improved (or worsened) or by the absolute number of unemployed people in the market, which shows the psychological effects of making public unemployment statistics.

A final series of observations concerns the particularity of the French electoral system, that is, the two-ballot system. My model is only concerned with the second round of the election. Therefore, it is possible that the results

of the first round will give a very detailed measure of the dynamics of the second round. It would be of great help to be able to pick up on the 'amplifying effect' of the first round results in predicting the results of the second round. From a methodological point of view, the answer to this question seems complicated, as it would be necessary to construct a conditional model to estimate the vote function for the second round based on an estimation of votes in the first round. This is certainly a direction to keep in mind for the next legislative elections, especially if the number of political parties in France continues to be so high. Finally, the challenge of a legislative predictive model lies in determining which majority will govern. In building a vote function, my model focuses on an electoral score for the incumbent majority. What is missing is transferring this predicted vote percentage into a number of seats. Many studies have tried to estimate the number of seats as a function of votes (Whiteley, 2005; Campbell, 2010), one being conditional on the other. In the case of France, this research perspective could be useful because being able to accurately predict a majority of votes for the incumbent government does not necessarily mean that this same government will have a majority of seats. In fact, the geographic representation of the 577 electoral districts makes a district-level vote prediction impossible, due to the absence of socioeconomic variables for these districts over a long period of time. Thus, the *department* level could be more promising in determining how to translate number of votes into seats.

## Conclusion

Predicting elections remains a fragile art. The analytical logic necessitates a well-established theoretical framework. The model presented by Foucault (2012), based on a politico-econometric vote function where economic and political factors are clearly identified as predictors and whose value is suspected of advantaging or sanctioning the incumbent government. Contrary to a large number of predictive models developed in the United States, the United Kingdom or France, my specification proposes a new avenue of research to taking into account the local dimension for the purpose of not only raising the number of observations (and therefore the robustness of the prediction), but also capturing more of the heterogeneity of the areas involved. Such an approach has already provided encouraging results for predicting the 2012 presidential score in France (Foucault and Nadeau, 2012).

Finally, with an error between 0.1 (Model 2) and 1.21 (Model 1), my prediction turned out to be reliable in declaring the winner of the legislative election while reducing the level of estimated error ($<1.5$), compared with other models using time-series data. Finally, often neglected in predictive

models, taking into account the rate of participation in legislative elections should be studied with more attention, focusing on a form of demobilization among the electorate, due to the legislative election's closeness to the presidential election (which has a very high participation rate). But information on voter mobilization is something for which it is impossible to obtain data a few months before the election, except in the case of public opinion polls (of the rolling panel variety).

## References

Campbell, J. (2010) The seats in trouble forecast of the 2010 elections to the U.S. house. *PS: Political Science & Politics* 43(3): 667–670.

Foucault, M. (2012) Forecasting the 2012 French legislative election. *French Politics* 10(1): 68–83.

Foucault, M. and Nadeau, R. (2012) Forecasting the 2012 French Presidential Election. *PS: Political Science & Politics* 45(2): 218–222.

Lebo, M.J. and Norpoth, H. (2007) The PM and the pendulum: Dynamic forecasting of British Elections. *British Journal of Political Science* 37(1): 71–87.

Lewis-Beck, M.S. (2005) Election forecasting: Principles and practice. *British Journal of Politics and International Relations* 7(2): 145–164.

Whiteley, P.F. (2005) Forecasting seats from votes in British general elections. *The British Journal of Politics & International Relations* 7(2): 165–173.