[Ensemble Predictions of the 2012 US Presidential Elections]

Jacob Montgomery, Washington University St Louis   
Florian M. Hollenbach, Duke University   
Michael D. Ward, Duke University

Since at least 1996 political scientists have been comparing true out-of-sample predictions of Presidential elections, and since the 2004 Presidential election this journal has presented comparisons of them, published prior to the election. The spirit of these symposia has generally been to use the validation of correct predictions to additionally garner bragging rights about which model is best in the sense that it accurately captures a story—also known as a theory—about the contexts and determinants of electoral behavior. Alfred Cuzán focuses his model on the effects of fiscal expansion as a policy that is likely drive voters away from the incumbent and toward the challenger. The point of all of these models is to develop the best model of the underlying data generating process. One heuristic is how well they do in predicting the electoral results in upcoming elections. An added bonus of having the most accurate predictions is bragging rights about the quality, accuracy, and beauty of the model.

Different models are built with different insights. For example, developed by Douglas Hibbs, the 'Bread and Peace' model argues that the aggregate votes for the president in postwar U.S. elections are well explained by just two fundamental determinants: (1) weighted-average growth of per capita real disposable personal income over the term, and (2) cumulative US military fatalities owing to unprovoked, hostile deployments of American armed forces in foreign wars.

Our approach is entirely different. Rather than search for the best model, the best theory, the best insight, we instead are looking to create the best prediction by incorporating the insights from a number of predictive models. Our approach can be thought of as predictive analytics, but quite simply we want to have the most accurate out-of-sample predictions. To do this, without creating a new theory or introducing a new specification, we rely on the insights of the extant models. We believe that each of these models captures an important set of insights about US electoral behavior, and each has been rigorously tested not only statistically, but also via a predictive heuristic. Therefore, our approach will attempt to combine the insights of each model into a single predictive model. It doesn’t matter to our approach if one model ``substantively’’ refutes another. All that matters is that they provide electoral predictions in previous elections.

The approach we use is called “Ensemble Bayesian Model Averaging,” or EBMA. EBMA was developed recently in the field of weather forecasting as a way of improving predictions by aggregating across models. Some weather models might be better at predicting “normal” weather patterns, and others better for rapidly changing conditions, for example. By averaging over these the overall prediction will be more accurate, even without having chosen the “best model.”

EBMA uses the predictive performance of the included component models on some *training* period to generate a weight for each individual component model. The EBMA prediction is then a kind of weighted average of the predictions made by each of the component models. In particular the EBMA model will give more weight to models that have been more accurate in the *training* period, as well as those models that make more unique predictions.[[1]](#footnote-1) One strong advantage of EBMA is that it only requires each component’s predictions as input, but not the models covariates. It is thus possible to use forecasts of subject experts or agent based models.

More technically, EBMA works in the following way.[[2]](#footnote-2) Assume we have an outcome in the future that is to be predicted and k predictive models (). Each component model comes from a prior distribution and thus one can describe in terms of its probability density function (pdf) conditional on . With the help of simple math and Bayes’ rule it is then possible to derive the marginal predictive distribution of given the k predictive models.

This pdf can be interpreted as a weighted prediction, where the weights of each model are dependent on the predictive performance in the training period prior to t\*.

Applying the EBMA to the example of presidential election forecasting, we use the in-sample predictions of each component model to calibrate the EBMA model. Thus we have XX number of forecasting models and a training period of XX presidential elections (from 19XX to 2004 or 2008). Our test period or out-of-sample prediction is the 2012 election. Each forecasting model is associated with a pdf, which is in our case a normal density function, following Raftery et al. (2005) is centered at a linear transformation of the individual forecast (). Thus the EBMA pdf and EBMA predictive distribution are then weighted averages of K normal pdfs. The predictive distribution for observation y (or our forecast for 2012) can be represented as the following.

Each represents the weight associated with each individual component model. The weights are estimated using maximum likelihood methods. While the log-likelihood functions cannot be calculated analytically, we can estimate it using an expectation maximization (EM) algorithm. The EM algorithm is iterated until the improvement in the log-likelihood is smaller than some predefined value. Given the estimated model weights for the EBMA model based on the training period, we can then create an EBMA prediction using the component model predictions and the estimated model weights.

The basic insight of this approach is that each component model in the ensemble captures some insight that yields predictions that are selectively accurate. Combining them and weighting them by their predictive success creates a sort of hyper model that in principle should be at least as good (in terms of predictive accuracy) as any individual model and it is likely that the ensemble will dominate each of its members. In a wide variety of applications such as inflation, economic growth, exchange rates, industrial production, and the gold-standard of prediction: weather. Its theoretical underpinnings, as well as its success in a variety of arena, suggest it could be useful in predicting elections, as well. We turn now to our results of the Bayesian aggregation of the ensemble of predictive models, which are described elsewhere in this Symposium.

Table XX shows the EBMA model statistics for the in-sample period. It shows the weights for each individual model, as well as … several error statistics for the in-sample predictions.

# References (this will change as we add refs.

Jacob M. Montgomery, F. M. (2012, March 22). Improving Predictions Using Ensemble Bayesian Model Averaging. *Political Analysis, 20* (tbd), p. tbd.

Stevens, J. (2012, June 24). Political Scientists Are Lousy Forecasters. *New York Times Sunday Review*, p. SR6.

About the authors

Jacob Montgomery is an Assistant Professor in the Department of Political Science at Washington University in St. Louis and a Fellow at the Center for Political Economy.  His research is in the areas of political methodology and American politics, with a special interest in political parties.  He teaches courses on statistical methods, experimental methods, and American political parties. He graduated with a B.A. from Wake Forest University with majors in Political Science and Mathematical Economics.  He earned an M.S. in Statistical Science, as well as his Ph.D. (2011) in Political Science from Duke University.

Florian M. Hollenbach is a PhD student with a focus in Political Economy and Methods. His current interests are the political economy of taxation, redistribution as well as state capacity and revenue extraction in developing countries. Additional interests include applied statistics and economic development.

Michael D. Ward is Professor of Political Science at Duke University. His primary interests focus on developing predictive models of political phenomena. His Ph. D. (1977) is from the home of political science forecasting, Northwestern University.

1. This means that component models with highly correlated predictions will be penalized and receive less weight. [↑](#footnote-ref-1)
2. We will describe the mathematical detail behind the EBMA model in less detail here. For a more detailed description the reader should consult (CITATION TO PA Paper). [↑](#footnote-ref-2)