president and his Republican predecessor, might it not do the same with fiscal policy? After all, during a president's first year in the White House spending is largely a function of the previous administration. It was in the fall of 2008 that the bank and auto bail-outs were enacted into law. So neither Mr. Obama nor his party is accountable for any but a relatively small portion of the enormous spike in outlays, from 20.8% of GPD in 2008 to 25.2% in 2009. Actually, as economists Paul Krugman and Mark Thoma pointed out, from 2010 to 2012 spending under Obama was relatively restrained. Indeed, after peaking in 2009, F drifted down a bit, settling at 24.3% this year.

To test this idea systematically, I proceeded as follows. In cases of a president in the first term of a new party reign, fiscal policy was measured not between presidential election years, but between the current year and his first year in the White House. Starting with the 1916 election which, following Fair, is the first in the series for estimating the Fiscal Model for forecasting purposes, this includes the first or only reelection bids of Wilson, Coolidge, F. D. Roosevelt, Eisenhower, Johnson, Nixon, Carter, Reagan, Clinton, G.W. Bush, and Obama. In the remaining 14 cases, the original fiscal policy value was retained. This resulted in two coding changes, both from an expansionary to a contractionary fiscal policy: Reagan and Obama.

The estimate of the model with the revised values for fiscal policy is similar to that published in the October issue (Cuzán 2012, 649). The coefficients undergo marginal changes. Also, in out-of-sample forecasting the mean absolute error of the model with either the traditional or the revised values of fiscal policy is less than 2%. But the difference in the forecast for 2012 is dramatic: whereas with the unrevised measure Obama was predicted to lose with 46.9% of the two-party vote, now it shows him winning with 51.1%.[1]

Conclusion

A preliminary assessment of this year's Fiscal Model failure suggests that the problem lay in the miscoding of Obama's fiscal policy, a measurement error caused by the spill-over in spending from the last year of the Bush administration into his own term. By recalculating the change in spending from the first to the fourth year of all presidents in the first term of a new party reign, including Obama, the Fiscal Model yields an accurate forecast for 2012.

NOTE

1. Space limitations do not allow display of model estimates, data, or a full list of references. By the time this article is published, all this will be available in an expanded version of this essay, available on my publications page (http://uwf.edu/govt/CuzanPublications.htm) and at ssrn.com.

REFERENCES

Campbell, James E. 2012. "Forecasting the 2012 American National Elections." *PS: Political Science and Politics* 45 (4): 610–13.

Cuzán, Alfred G. 2009. "Will Barack Obama Be a One-Term President?" *TCS Daily*, March 10. http://uwf.edu/govt/documents/Cuzan-2009-Will%20Barack%20Obama%20Be%20a%20One-Term%20President-tcsdaily.pdf.

Cuzán, Alfred G. 2012. "Forecasting the 2012 Presidential Election with the Fiscal Model." *PS: Political Science and Politics* 45 (4): 648–50.

Fair, Ray C. 2012. "Vote Share Equations: Predictions." http://fairmodel.econ.yale.edu/vote2012/index2.htm.

AGGREGATION AND ENSEMBLES: PRINCIPLED COMBINATIONS OF DATA

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

Perhaps *the* major story in forecasting the 2012 election is the growing awareness of the benefits of aggregating multiple sources of data for improving prediction. Most prominently, polling analysts including Simon Jackman, Drew Linzer, and Nate Silver made the strong case, ultimately validated by the election results, that combining information from multiple polls gives a better picture of the electorate than any poll analyzed in isolation.

Our approach, ensemble Bayesian model averaging (EBMA), draws on this same basic intuition, even if the specifics differ considerably from the poll aggregation methods referenced earlier. Instead of creating the single "best" forecasting model, EBMA aims to "combine the intuition, theories, and concepts implicit in all of the forecasting models presented in this symposium to make an accurate out-of-sample prediction" (Montgomery, Hollenbach, and Ward 2012a, 651).

Although no single election can validate the ensemble approach, the performance of EBMA in 2012 was encouraging. The final outcome, 51.8% of the vote for President Obama, was well within our 95% predictive credible interval [46.4%, 52.5%] . Our point prediction, the median of the predictive posterior, was 50.3%. Thus, our prediction was off by about 1.5%. Of the ten forecasts included in the ensemble, only the Abramowitz, Erikson and Wlezien, and Campbell (Trial-Heat) models offered a more accurate prediction with absolute errors of 1.2%, 0.8%, and 0.2% respectively. Indeed, we especially give credit to these authors, whose accurate forecasts of the 2012 outcome contributed so much to the ensemble's performance. The Trial-Heat Model and especially Abramowitz's Time for Change Model were among the most heavily weighted in the ensemble. EBMA's reliance on them was based, in part, on their previous accuracy, which they repeated this cycle.

EBMA's strong performance provides additional support for the notion that more information is better. Yet, note that EBMA does not simply aggregate forecasts without respect to their past performance. For the purposes of this symposium, we weighted forecasts based on the accuracy of each model's pseudo-forecasts (their in-sample predictive ability). Elsewhere, we took the more stringent approach of evaluating forecasting teams based on their true out-of-sample forecast made in advance of elections (Montgomery, Hollenbach, and Ward 2012b). In either case, models that provided less accurate forecasts received less weight. This discrimination is why EBMA, in general, outperforms more naïve approaches to aggregating forecasts. The mean prediction of the models included in our ensemble was 50.0 and the median was 49.4; the EBMA aggregation performed considerably better than

https://doi.org/10.1017/S1049096512001552 Published online by Cambridge University Press

either of these. At the same time it preserves information about the uncertainty of the ensemble average.

It seems easy to predict that ensemble averaging will continue to be a part of election forecasting, given Drew Linzer and Simon Jackman's success this year, as well as the accuracy and popularity of Nate Silver's efforts. Doubtless it will be advantageous—and fruitful—to continue to broaden the range of models included in the ensemble in the next symposium.

REFERENCES

Montgomery, J.M., F.M. Hollenbach, and M.D. Ward. 2012a. "Ensemble Predictions of the 2012 US Presidential Election," *PS: Political Science and Politics* 45 (4): 651–54.

Montgomery, J.M., F.M. Hollenbach, and M.D. Ward. 2012b. "Say Yes to the Guess: Tailoring Elegant Ensembles on a Tight (Data) Budget," Annual Meeting of the American Political Science Association, New Orleans, LA.

2012 PRESIDENTIAL, US HOUSE, AND US SENATE FORECASTS

Carl E. Klarner, *Indiana State University*

For the second presidential election in a row the prediction of the "Klarner Model" was within half a percent of the actual result. This model forecast that Obama would receive 51.3% of the two-party vote, while he actually received 51.8%.[1] The model also called all states correctly, with the exception of Florida, which was predicted to be narrowly lost by Obama with 49.7% of the vote. These forecasts were made on July 15, 2012. This success does not indicate the model is better than the other models in the symposium as luck is a major determinant of which model gets closest to the mark. Although the Klarner Model correctly called 49 out of 50 states, that means little in an election where merely calling states on the basis of 2008 results would have called all but two correctly.

Two other state-level presidential forecasting models were presented in the October *PS* election forecast symposium: that of Jerome and Jerome-Speziari (J&J), which called two states incorrectly, and that of Berry and Bickers (B&B), which called nine incorrectly. The accuracy of the J&J Model for the national popular vote was essentially tied with the Klarner Model, while B&B were 4.3% off. If we stopped here, we could conclude that the J&J Model did as well as the Klarner Model whereas the B&B Model did worse. However, all three models were equally unimpressive in one way, and the B&B Model brought information to the table that was not in a variety of other sources, as explained next.

Making accurate forecasts at the state level is a function of two things: correctly calling the national tide and correctly ordering Democratic success across states. One standard of how good a model is at ordering the states is determined by its ability to add prediction success to a variable measuring the percent of the vote obtained by the Democrat in the last election. Table 1 reports the results of five regressions, with the dependent variable in all being the 2012 Democratic percent of the two-party vote. All variables are centered around 50% to facilitate assessment of bias.

The first regression reported in table 1, in row two, indicates that the 2008 vote accurately ordered the states from most to least Democratic. When using the predictions from the three state-level forecasts as independent variables (rows three, four, and five), none perform better than lagged vote. The fifth regression in the table uses Nate Silver's November 5, 2012, state predictions as an independent variable as a useful summary of forecasts from polls immediately before Election Day: it only marginally improves on 2008 vote share in its ability to order states, but this indicates that it is possible to improve over prior vote share. When the four forecasting variables are added to a model with lagged vote share in turn, only those of B&B and Silver attain statistical significance (analyses not shown). Perhaps the care that B&B took in modeling state economic conditions yielded predictive capacity not found in other sources. When pitted against Silver's Election Eve forecasts, their forecasts still attained statistical significance ($p < .05$, analyses not shown).

Figure 1 reports the unstandardized regression coefficient from bivariate regressions of the state Democratic two-party vote on lagged vote share for each of the 17 elections from 1948 to 2012, but excluding the "one party" states not used in the prediction model (Klarner 2012, 655). The figure indicates that the impact has gone up over time, while prediction error (measured by the standard error of the estimate from those 17 regressions, not shown) has gone down, with lagged vote doing better at predicting current vote in 2012 than any other postwar election. This trend is what one would expect with the increased polarization of the electorate over time. These findings suggest that the longer time period used in the Klarner Model may have decreased its ability to order states correctly

Table 1

Five Regressions Assessing the Relationship between 2012 State Democratic Vote, Forecasts, and Lagged Vote

INDEPENDENT VARIABLE	INTERCEPT	INDEPENDENT VARIABLE COEFFICIENT	STANDARD ERROR OF THE ESTIMATE	R-SQUARED
Lagged Vote	−2.408*	1.049*	2.241	.966
	(.320)	(.029)		
Klarner Prediction	−.190	1.229*	2.468	.958
	(.346)	(.037)		
J&J Prediction	−2.163*	1.239*	5.449	.797
	(.777)	(.089)		
B&B Prediction	4.711*	1.059*	2.267	.965
	(.344)	(.029)		
Nate Silver Prediction	.151	1.049*	1.882	.976
	(.264)	(.024)		

Note: standard error in parentheses. *= $p < .05$. $N = 51$

https://doi.org/10.1017/S1049096512001552 Published online by Cambridge University Press