Calibrating Ensemble Forecasting Models with Sparse Data in the Social Sciences

Jacob M. Montgomery
Department of Political Science
Washington University in St. Louis
Campus Box 1063, One Brookings Drive
St. Louis, MO, USA, 63130-4899
(314) 935-9106
corresponding author: jacob.montgomery@wustl.edu

Florian M. Hollenbach
Department of Political Science
Duke University
Perkins Hall 326 Box 90204
Durham, NC, USA, 27707-4330

Michael D. Ward
Department of Political Science
Duke University
Perkins Hall 326 Box 90204
Durham, NC, USA, 27707-4330

August 21, 2014

Calibrating Ensemble Forecasting Models with Sparse Data in the Social Sciences

Jacob M. Montgomery, Florian M. Hollenbach and Michael D. Ward

Keywords: Bayesian methods, Election forecasting, Labour Market forecasting, Calibration, Ensembles

In preparing our final replication archive, we discovered some errors in our code that necessitate changing the results in Tables 5, 6, and Figure 4. The updated table results are below.

Table 1: Comparing adjusted EBMA models with Green Book, median, and mean forecasts of US unemployment (1981–2007)

	MAE	RMSE	MAD	RMSLE	MAPE	MEAPE	MRAE	PW
EBMA (c=0)	0.52	0.73	**0.34**	0.090	8.01	6.31	**0.71**	**27.36**
EBMA (c=0.05)	**0.52**	**0.73**	0.34	**0.090**	**7.98**	**6.20**	0.75	29.25
EBMA (c=0.1)	0.53	0.73	0.35	0.090	8.08	6.44	0.76	29.25
EBMA (c=1)	0.61	0.80	0.46	0.102	9.72	8.95	0.95	46.23
Green Book	0.57	0.73	0.43	0.093	9.37	8.81	1.00	45.28
Forecast mean	0.61	0.80	0.46	0.102	9.71	9.06	0.93	46.23
Forecast median	0.62	0.81	0.47	0.103	9.83	8.87	0.98	47.17

Note: Definitions of model fit statistics are provided in the Appendix. The model(s) with the lowest score for each metric are shown in bold. Differences between model performance may not be obvious due to rounding.

Table 2: Comparing predictive accuracy of EBMA and component models with eight metrics

Number of metrics on which EBMA performed better	Number of Predictions Made 4 –10	11-30	31–60	> 60
7 – 8	0.64	0.80	0.82	1.00
5 –6	0.15	0.11	0.14	0.00
2 –4	0.11	0.03	0.00	0.00
0 – 1	0.11	0.06	0.05	0.00
Number of components	74	66	22	9

The rows of the table show the number of metrics by which EBMA outperforms components, while columns show the number of forecasts made by these models. The values in each cell of the table are the proportion of component models falling into that category (columns will sum to unity). Note that EBMA performs very well against its components, especially those that make many predictions.

Figure 1: Observed and forecasted US unemployment (1981-2007)