



Proposal Status | MAIN ►

Organization: Washington University

Review #1

Proposal Number: 1152472
NSF Program: Political Science
Principal Investigator: Montgomery, Jacob
Proposal Title: Collaborative Research on Ensemble Methods for the Prediction of Political Outcomes
Rating: Multiple Rating: (Good/Fair)

REVIEW:

What is the intellectual merit of the proposed activity?

The proposed research is to promote the use of ensemble Bayesian Model Averaging (EBMA) to social science applications and by doing so to encourage the use of forecast accuracy as an indicator of the utility of statistical models (in the social sciences). The proposed new research outlined in section 4 of the paper includes: 1) developing fully Bayesian EBMA estimation methods for censored-continuous, binary, and count data, 2) review and recommend a set of forecast comparison metrics, 3) assess the impact of the vintage of data on EBMA forecasts, 4) embed methods to handle missing data in EBMA, 5) develop weights that penalize model complexity and allow for non-constant variance, and 6) encode all of these methods in an R package.

The proposal is unbalanced in its presentation. Section 2 (pages 2-6) presents model details for BMA, EMBA, EBMA for binary data, and the EM algorithm that are already published elsewhere (and they provide citations; but this is not the work of the PIs). Section 3 provides several empirical applications that takes up far too much space if it is intended to provide a proof of concept. Finally, at the bottom of page 11 they provide about two pages on their research plans. Each element of the plan is only given a paragraph and there are too few details to evaluate how they expect to accomplish the tasks, in ways that are novel extensions beyond the existing literature.

While some of the proposed work is genuinely new and adapts the methods to political science applications (and social sciences more generally), other aspects appear to be no different than what is already available to researchers. The extension to censored-continuous outcomes is already published in Sloughter et al. (2007) [see their citations] and as they note in the proposal the extension to binary is "straightforward". Extending to counts is similarly trivial as they only need to change the functional form of the density to a Poisson. Similarly, a fully Bayesian implementation of EMBA is already published in Vrugt et al. (2008) [their citation]. In the same paper they accommodate non-constant variances. As for missing data, in a Bayesian setting any missing data is treated as missing parameters to be estimated. If this is their planned approach it is again following a well-worn path and is not particularly novel. Finally, in terms of software development, why are they proposing to develop new software when there is already a full package for EMBA in R (ensembleBMA), several packages for BMA, and Vrugt states that his Bayesian MCMC implementation (in Matlab) is freely available upon request (see <http://math.lanl.gov/~vrugt/software/>). It would make much better sense to work directly with Raftery and the other developers of the existing R package to port Vrugt's method to R. While the EMBA R package is tailored to weather forecasting, that is not something that should present much of a barrier to applied statisticians who are interested in using the method in any other field.

This means the intellectual merit rests primarily on 'technology-transfer' from statistics to a particular social science domain (the only examples they discuss are political science and the main venue they propose to exhibit is a political science meeting). I also suspect that economists are already moving to adopt EMBA given the widespread use of BMA in econometric forecasting (too little attention was given to this in the proposal).

The senior PI (Ward) is a highly accomplished senior scholar in political methodology and has acted as an agent of diffusion to bring spatial econometric methods and Bayesian methods into the political methodology mainstream. I fully expect he can complete all of the proposed work and his institutional setting is sufficient to support it.

What are the broader impacts of the proposed activity?

The broad impacts include the usual language about graduate training and publishing. The project would also produce an R package and vignette. I expect the impact to be limited by the fact that similar software packages already exist so the only benefit to their proposed new addition would be to individual researchers who are incapable of generalizing the existing software to their applications.

Summary Statement

The researchers are capable of delivering on the proposed research and the stature of the senior PI within political methodology means that any advances would reach a large audience. Even given those positives, it appears to me that the main contributions are only minor (sometimes trivial) advances over existing EMBA theory and in some cases highly duplicative. The software effort also appears to be highly duplicative. An improved proposal would include as co-PIs perhaps Raftery (a sociologists/statistician who would certainly be interested in generalizing his methods to suit another audience) and perhaps Vrugt. More attention should have also been given to details of the proposed extensions rather than restating published work that is already available to reviewers. Give the lack of novelty on the methods dimension, this makes the proposal a particularly poor fit for MMS.

 [Back to Proposal Status Detail](#)

Download [Adobe Acrobat Reader](#) for viewing PDF files

National Science Foundation
4201 Wilson Boulevard, Arlington, Virginia 22230, USA
Tel: 703-292-5111, FIRS: 800-877-8339 | TDD: 703-292-5090

[Privacy and
Security](#)