Bayesian Analysis (2009) **4**, Number 3, pp. 437–438

Comment on Article by Monni and Tadesse

Hugh Chipman*, Edward George† and Robert McCulloch‡,

We congratulate the authors on an interesting and well-written article. We enjoyed reading it and found the approach and ideas presented stimulating. The problem of variable selection in regression with multiple responses has not received much attention, and this article opens up a number of interesting directions in this area. By grouping responses and assuming a homogeneous regression relationship with each group, the proposed model appears able to squeeze considerable information out of "small n, large p" datasets. Such data are increasingly available due to the large number of high-throughput studies conducted in many areas of science.

The challenge of simultaneously discovering groups of responses and subsets of predictors is significant, even if approached from a purely algorithmic view. The model developed is thus quite impressive, since it offers a full inferential framework and incorporates an effective search algorithm. In particular, we were pleased to see the authors report a successful and effective implementation of parallel tempering. The design decision to develop just a few effective moves, combined with parallel tempering, seems to have produced a MCMC algorithm capable of exploring a large and complex parameter space.

Assessing the success of such models is difficult due to the high dimension. In our recent BART work we have emphasized out-of-sample predictive performance on real and simulated data sets. In this paper the authors report success in correctly identifying the known structure in a simulated example, outperforming the method of Brown, Vannucci, and Fearn (BVF), and obtain interesting results in a real data set.

We feel that there are some difficulties in comparisons with BVF because the fundamental modeling assumptions are different. In BVF there is only one component or group. The meaning of a component is quite different: in BVF, all dependent variables are related to the same set of explanatory variables but the coefficients are not constrained to be the same for each dependent variable. In addition, BVF allow for dependent errors.

Monni and Tadesse emphasize that their model has the ability to capture "correlation among the outcomes in a component" (Section 2.2) but this dependence is of a very different nature than that captured by BVF. In this paper, dependence within a component comes from integrating out the common coefficients. Conditional on the parameters and the component, errors, and hence responses are independent.

*Department of Mathematics and Statistics, Acadia University, Wolfville, NS, Canada, mailto:hugh.chipman@acadiau.ca

†Department of Statistics, The Wharton School, University of Pennsylvania, Philadelphia, PA, mailto:edgeorge@wharton.upenn.edu

‡Department of Information, Risk & Operations Management, Red McCombs School of Business, University of Texas, Austin, TX, mailto:robert.mcculloch1@gmail.com

 DOI:10.1214/09-BA416A

We appreciate that a major goal of the paper is the discovery of the components, but conditional on the components, the strong assumptions made convert the problem into a "big n" situation. Given there is only one set of common slopes it is almost as if there are $n_k N$ observations. That is, each response in a group contributes N observations to the likelihood for a single common vector of regression coefficients. This is evident in the double sum in the likelihood equation given below (1). Consider the comparison with BVF on page 12. BVF has to do variable selection and estimate a different set of coeficients for each y as opposed to just one set of slopes. When there are many y's, BVF is estimating many more parameters. The conclusion on page 15 that BVF "is not as suited as ours for variable selection with many response variables" seems like an overstatment given that the data are simulated in accordance with the modeling assumptions made in the paper.

The assumption of a common set of slopes within a component could be viewed either as an enabling model that borrows strength across different responses, or a simplifying assumption that may be convenient when data are scarce. We are somewhat concerned that nonstatisticians might adopt such strong assumptions in order to obtain dramatic increases in sample size. In this light, we find reassuring the author's opening remark that they view this method as exploratory.

Another issue complicating the evaluation of the methodology is the difficulty in reporting the uncertainty. This issue is shared by many modern Bayesian papers. The methods are powerful in that they allow us to explore complex, high-dimensional structures, but it is then difficult to report the results fully. In the current paper much reliance is made on the MAP estimate. While we understand why this choice is made, we need to be aware of possible adverse consequences. Would the MAP estimate give the best out of sample performance? If we could convey the uncertainty, would we see that BVF is more uncertain rather than wrong when asked to estimate more parameters?

In noting that the model of this paper is not the same as BVF we do not mean to imply that we find the model uninteresting. We agree with the authors that basic problem is important and has been neglected. We very much like the approach taken in this paper and feel it is an important contribution.