

Comment on Article by Monni and Tadesse

Chris Fraley*

The problem of variable selection in a multiresponse setting is a difficult one that deserves attention. With regard to this paper, I see two main areas for discussion: proof of concept in terms of the scope of the examples, and the establishment of context in terms of the relationship to and departure from existing work.

1 Simulated and Real Data Examples

The simulated and ‘real data’ examples leave a number of fundamental questions unanswered.

Simulated Data Examples. The design of the simulated examples raises two immediate issues:

- In the simulated examples, the outcomes were generated with equal variance throughout — formula (6) on page 10 — whereas in the model underlying the method described on page 3, the variance is allowed to vary among components. How well would the method work in simulations where the variance is allowed to vary?
- Were the outcomes and/or covariates scaled, standardized or normalized in any way? This could affect the resulting analysis and conclusions.

The subsection entitled **Performance in the presence of high collinearity** on page 15 starts with the sentence “*It is reasonable to assume that the presence of highly correlated covariates may complicate the identification of relevant predictors.*” Two new covariates, which are linear combinations of existing covariates, are then added to simulated data which had $N = 50$ samples and $p = 200$ covariates, with the conclusion that “*... the algorithm is reasonably resistant to colinearity between predictors.*”

It’s not clear what’s intended here, since the dimensions of the data imply multicollinearity, and the method is supposed to address cases in which the number of regressors is much larger than the sample size (e.g. page 2).

Questions of interest with regard to multicollinearity would include:

- How would permutations of the covariates and/or outcomes affect the results in the simulated and ‘real data’ examples?
- Is identifiability an issue?
Because of the inherent multicollinearity, more than one set of predictors could fit a

*Department of Statistics, University of Washington and Insilicos LLC, Seattle WA, <http://www.stat.washington.edu/fraley/>