

# Revision 1 of “Comparison of Multi-response estimation methods”

The authors have responded adequately to most of my comments, but I think a little additional revision may be in order.

## Section 4: $\eta$ or $\kappa$

Predictor collinearity is describe from (3) in terms of the exponential parameter  $\gamma$ , but response collinearity is described in terms of  $\eta$ , which is notation for eigenvalues from (4). It seems to me that  $\eta$  and  $\kappa$  have been confused throughout the paper. Here are some instances,

- The first sentence following (4) refers to four levels of eta. Then the following sentence gives further discussion in terms of  $\kappa$ .
- Figures 2, 4 & 5 have eta as a label, which makes no sense in terms of (4). I think these labels should be kappa.

In the two sentences following (5), shouldn't the  $\eta$ 's be  $\kappa$ ?

## Section 4: Construction of the $\Sigma$ 's following (5)

In the fourth line following (5), shouldn't  $p \times m$  be  $m \times p$ . Assuming that is so, here is how I understand the constructions from the description. Let  $\sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} & \sigma_{14} \end{pmatrix}^T$ ,

We thank the reviewer for making us aware of this error. There was a non-consistent use of  $\kappa$  and  $\eta$  in the previous version. Now all  $\eta$  refers to the decay factor of eigenvalues corresponding to responses and  $\kappa$  refers to the actual eigenvalues. Here  $\kappa$  is mentioned in only one place while  $\eta$  has been used in several places and are also used as eta in the r-code. Now, these should be fixed.

This has now been fixed.

let  $\mathbf{1}_4$  denote a  $4 \times 1$  vector of ones, and partition the  $p \times p$  orthogonal matrix  $\mathbf{Q} = \begin{pmatrix} \mathbf{Q}_1 & \mathbf{Q}_2 \end{pmatrix}$ , where  $\mathbf{Q}_1$  is  $p \times 4$ . Then,  $\square$

$$\Sigma_{xx} = \mathbf{Q}\Sigma_{zz}\mathbf{Q}^t$$

$$\Sigma_{yy} = \mathbf{R}\Sigma_{ww}\mathbf{R}^t$$

$$\Sigma_{x,y} = \mathbf{Q} \begin{pmatrix} \sigma \\ \mathbf{0} \end{pmatrix}_{p \times 1} \mathbf{1}_4^T \mathbf{R}^t = \mathbf{Q}_1 \sigma \mathbf{1}_4^T \mathbf{R}^T$$

$$\text{rank}(\Sigma_{x,y}) = 1$$

If my interpretation is correct then there are 4 true predictor components provided  $\gamma \neq 0$ , and 4 true response components provided  $\kappa \neq 0$ . The simulations never used  $\gamma = 0$  so there are always 4 predictor components. However,  $\kappa = 0$  was used (assuming that's what  $\eta = 0$  means) and then there is only one response component. If all of this is correct it should be stated for clarity in the paper.

## Response components

In the middle of page 8 it is stated that “Here we have assumed that there is only one response component.” Does this mean that in all estimation methods where relevant a single response component was used? If so, then in all settings where  $\kappa \neq 0$  the estimators were based on the wrong scenario since there are really 4 true response components. Importantly, no estimator had the ability to adapt to the number of response components. Additionally, all subsequent mentions of the “number of components”, like around (8), must be references to the number of *predictor* components. I think this need to be made clearer, if my guesses are correct. And I would like to see some discussion of the possible effects of forcing the wrong number of response components.

Additionally, designs 9 and 29, which are contrasted in Table 1, differ on the number of true response components, design 9 having 1 component and design 29 having 4, although data from both designs were evidently fitted with

We believe we are at the core of what is the uncertainty here. In our simulations we have used one single response component throughout, which means that the unit vector  $\mathbf{1}_4$  in your equations must be replaced by the vector  $\begin{pmatrix} 1 & 0 & 0 & 0 \end{pmatrix}^T$ . In which case, the covariance matrix of  $\mathbf{z}$  and  $\mathbf{w}$  becomes,

$$\Sigma_{z,w} = \begin{pmatrix} \sigma & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

This is then rotated by  $\mathbf{R}$  and  $\mathbf{Q}$  to get the covariance matrix of  $x$  and  $y$  yielding one response component and four predictor components. We refer to Rimal, Almøy, and Sæbø (2018) paper for more details on this (Note that in that paper  $\mathbf{R}$  and  $\mathbf{Q}$  were used oppositely from your use above.)

The number of true response components will still be 1 even in  $\kappa \neq 0$ , only all 4 response variables will have some correlation between them since they are rotated together.

**Page 8 (Second last paragraph):** In the discussion onwards, *number of components* refers to the number of predictor components unless otherwise stated.

one response component. Does this matter for interpretation?□

## Section 5: Basis of Comparison

I'm still unconvinced by the choice of the inner product matrix in (6). If the variances of the elements of  $\hat{\beta}$  differ greatly, then the metric (6) will emphasize the most variable element rather than the whole of  $\hat{\beta}$ . □

## References

Rimal, Raju, Trygve Almøy, and Solve Sæbø. 2018. "A tool for simulating multi-response linear model data." *Chemo-metrics and Intelligent Laboratory Systems* 176 (May): 1–10. <https://doi.org/10.1016/j.chemolab.2018.02.009>.

A small misunderstanding here as well. Design 9 and 29 differ not in the number of true response components but they differ on the correlation between *response variables*. Design 9 have uncorrelated response variables while Design 29 have highly correlated response variables. This is stated in **Page 7 final paragraph**.

If we use  $(X^t X)$  as a weight matrix, the result will be highly similar to the prediction error which we have already explored in our previous paper. In many designs with high multicollinearity, models may perform very differently with regard to estimation and prediction. As is known, a model may predict very well in spite of poor estimation, potentially arising from high multicollinearity, if new observations lie in the same space as the training set. This is why we in this paper want to explore estimation performance using the expected squared difference between the estimated and true regression coefficient to get a clear understanding of pure estimation error in different settings for the various methods.