Comparison of Multi-Response Estimation Methods

Reviewers' Comments

Here Blue colors are the comments and Green colors are the actual changes in the text.

Reviewer One

This manuscript compared the parameter estimation accuracy of some Multi-response linear regression model by simulated data with varying properties. This work gave some insights into these methods and helped the researchers to understand their performance. Based on this point, this work obtained some findings. I have the following comments and questions on this manuscript which the authors should consider.

- Some authors' name should be corrected, for example, Næs on Page 3. ☐

  Fixed.

- I think the writing about the scalar, vector, and matrix should be used according to the general rule. For example, italic lowercase is used for scalar, such as equations (9) and (10). ☐

  Old & new: pg-17 (last sentence)
  First paragraph in Analysis section
  Here the ... represented by cube notation.

- Please explain why the optimal components determined in these methods are different. I hope the authors could give further explanation about this. ☐

  Last paragraph of estimation methods
  Here, each methods uses ... smaller optimal dimension.

- Please show the relationships of the relevant space form the compared methods. I hope the authors could give further discussion about this. ☐

  Helland, I.S., et. al. (2018) have discussed this in details and the references are included in several places.

- For the real data, I think the manuscript has a significant flaw that don't use a real data to validate the parameter estimation accuracy of compared methods. It is of a prime importance to used real dataset to help readers of Chemo-lab to understand the method. Thus, it is strongly recommended the authors apply at least one real dataset to show their performance. ☐

  Here we are comparing estimation error for which true regression coefficients are required which is not possible with the real data.

- The GitHub repository is not available, an applicable version should be provided. ☐

It is working and is mentioned in the paper.

https://github.com/therimalaya/04-estimation-comparison

This paper gives an elaborate organisation that is very helpful for reading. This is commendable.

Thank You!

Reviewer Two

This paper is a followup to an earlier paper on multi-response prediction methods by the same authors. Accordingly, I had to read the prediction paper before turning to the current submission on estimation. I initially expected the estimation results to largely parallel those for prediction, but it turned out that there were a few surprises, particularly the behaviour of envelope estimation in the presence of highly correlated responses, as represented in Table 1. One possible explanation for this phenomenon comes to mind. The log likelihood objective functions for envelope methods are non convex. It seems possible, depending on the sample size, that highly correlated responses produce objective functions with multiple maxima of not substantially dissimilar magnitudes. If so, it may be possible for the algorithms to get caught at a local point relatively far from the true value. This might also explain the predictive performance of envelopes. Multiple local optimal of not dissimilar magnitudes means that the data support multiple values for β, which translates in terms of the likelihood into similar predictions. This might be tested by running the algorithms at different starting values, including one at the true value of the regression coefficients. Here Xenv might be sufficient since its performance in Table 1 is similar to that of Senv. Another possibility is to plot the objective functions along selected directions. This additional effort would seem necessary to (hopefully) explain one way in which the submission diverges from the results in the previous paper. Another of my general reactions is that the clarity of presentation needs attention. There seem to be simulation details missing and some statements are not sufficiently clear.

Old: pg-22, New: pg-24 (line-5)
Second paragraph of Discussion and Conclusion
This indicates that the ... function with multiple maxima (Cook and Zhang, 2016).

Thank you for the suggestion. It would be interesting to further study the behaviour of envelope and why we have this observed differences.

Old: pg-15 (top), new: pg-14
Exploration Section:
In both of these designs and in prediction and estimation error, Xenv ... more than one true response dimension.

See page 6 of this document.

A quibble: I prefer indented paragraphs, as the current style sometimes makes it unnecessarily hard to tell the start of a paragraph, which is annoying.

Fixed now.

Top of page 6: PCR cannot be equivalent to PLS and Xenv under the definition of PCR given on the bottom of page 4. Helland must have been using a different notion of PCR.□

Old: pg-6, New: pg-6 (line-3)
... have discussed when and under which condition the population models of PCR, PLS and Xenv are equivalent.

Page 6, sample size: The envelope methods, being based on the likelihood, are asymptotically efficient and so, with n sufficiently large, must do better than the other methods in both estimation

and prediction. This might be mentioned for completeness. □

Section 4. Experimental Design: The description of the experimental design should enable the reader to reconstruct the model as given in (1) and (2), but that does not now seem possible. How were Σ_{yy}, Σ_{yx} and Σ_{xx} constructed, as the eigenvalues are only part of the story? It's my understanding that the columns of β can be written as linear combinations of either eigenvectors 1,2,3,4 or 5,6,7,8 of Σ_{xx}. How was this accomplished? If β is full column rank then it seems that the true number of components must be 4. □

It is stated on page 7 that simulations were restricted to regressions with only one response component. How was this component constructed? Was the number of predictor components also restricted by the simulation? If so, how? This seems important because of the different numbers of components selected by the various methods as represented in Figure 7. Naturally one wonders about the true number of components set in the simulation.

The page 7 sentence "In the final dataset all predictors together span the same space as the relevant predictor components . . ." makes no sense to me since all predictors together should span $\mathbb{R}^p$, unless Σ_{xx} is singular. □

Page 8, line 5 from the bottom: Figure ?? □

Section 5. Basis of Comparison: I'm a bit perplexed by the choice of the inner product matrix in (5). Since the eigenvalues of Σ_{xx} are different, the predictor variances differ, unless Σ_{xx} was constructed as a constant times a correlation matrix. And the predictors are of course correlated. The columns of β were scaled by the response variances but there was evidently no effort to correct for the different variances or correlations of the predictors. In view of what I know about the simulation setup, I would have thought that Σ_{xx}, instead of the identity, is a more appropriate inner product. A case can also be made for using Σ_{xx}^{-1} as the inner product, but I can't see any clear justification for using the identity. □

Section 6. Exploration:. Figure 4 is difficult to interpret because one needs to take careful note of the scales for the $\gamma = 0.2$ and $\gamma = 0.9$ plots. At $\gamma = 0.9$ I can't see any obvious differences

Old: pg-22, New: pg-24
In discussion section:
Since the envelope methods ... others usual MLE methods.

Old: pg-6, New: pg-6
Experimental Design: First paragraph
... In the simulation, number of response variables $m = 4$ and number of observations $n = 100$ are fixed, and the ...
Old: pg-7, New: pg-8
After the discussion of simulation parameters
Using these simulation parameters, a latent covariance matrix is constructed as in 5 ... mechanism in details.

Old: pg-7, New: pg-8 (paragraph-2)
For the predictors, there are 4 true relevant components, so the relevant space for predictor matrix has 4 dimension.

Fixed

Here we have scaled the estimation error by response variance just to keep them together in the same scale while plotting. Since these responses are analysed separately, also in MANOVA, using identity matrix in (5 on old, 6 in new), we believe, will give a proper measure of squared error of regression coefficients.

in the methods. The claim that "... envelope methods tend to have increased estimation error in cases of highly correlated responses (eta: 1.2), whereas ..." may be true, but I can't see it from the figure. In particular, I can't see that any method is "best" at $\gamma = 0.9$ and $\eta = 1.2$. The plot does clearly show the impact of relative position. □

We have removed the last sentence of the paragraph. I think the reviewer is correct. It is difficult to see the difference from the figure alone.

Might the relatively large number of components for PCR, PLS1 and PLS2, shown in Figure 5 partly explain their relatively poor performance in the previous paper? When reading the top of page 13, I again found myself wondering about the true number of components. □

Experimental design details has been added in New: pg-8.

The conclusion in the penultimate sentence on page 13 evidently comes from Table 1. But Figure 6 shows that envelopes do well in estimation with one component and no evident bias, while PLS2 needs 7 or 8 components to give the same visual impression. I would have thought that the estimation variation increases with the number of components, so PLS2 estimators would be noticeably more variable than envelope estimators. Is this reflected in the simulations somehow, or is my intuition wrong? □

Old: pg-22 (line: -5), New: pg-24
Third last paragraph:
... this can easily be controlled through a method like cross-validation.
At the end of mentioned paragraph:
Here, the variation ... included in the model.

In the first full sentence on page 15, to what does "Here the . . . " refer. All of section 6? Just the material following the paragraph? □

Replaced "Here" in:
Second last paragraph of exploration:
In the figure, PLS2 ...
First full paragraph of page 15(old)
In this study, the ...

Last sentence in third full paragraph on page 15 "However, one should ..." and sentence "However, non-optimal number ..." starting on line 6 from the bottom on page 22: My guess is that these statements come from the estimative performance of Senv in Figure 7 for 7–10 components. The conclusion is then based on comparing Senv at > 5 components above its optimum number of components to PLS2 at < 3 components above its optimum. To be on a equal footing, shouldn't these methods be compared at the same number of components above the optimum? □

Old: pg-22, New: pg-24
Last sentence of last complete paragraph
Although non-optimal number of ... like cross-validation.

Section 7. Analysis: Nice discussion.

Thank you

Minor Changes

7th line from the end of page 13 (old)

... using only one or two ...

Common Answers

There are nice ideas and suggestions to further explore the performance of envelope based methods. Since these methods have remarkable performance in prediction and also in estimation, we agree that exploration of the impact of initial values on optimising function is required. In addition, we have used principal components while using wide data matrix with envelope methods. The effect of this on the quality of estimation also need further exploration both theoretically and practically. Since this study tries to have the similar design, data and situation as the prediction paper discussed published before, these explorations and studies is beyond the scope of this paper.