Comparison of Multivariate Prediction Methods

Raju Rimal[a,*], Trygve Almøy[a], Solve Sæbø[b]

[a]*Faculty of Chemistry and Bioinformatics, Norwegian University of Life Sciences, Ås, Norway*
[b]*Prorector, Norwegian University of Life Sciences, Ås, Norway*

Abstract

While Data science is battling to extract information from the enormous explosion of data, many estimators and algorithms are being developed for better prediction. Researchers and data scientists often introduce new methods and evaluate them based on various aspects of data. However, studies on the impact of/on model with multiple response model is limited. This study compares some newly-developed (envelope) and well-established (PLS, PCR) prediction methods based on simulated data specifically designed by varying properties such as multicollinearity, correlation between multiple responses and amount of information content in predictor variables. This study aims to give some insight on these methods and help researcher to understand and use them for further study.

Keywords: model-comparison,multi-response,simrel

1. Introduction

Prediction has been an essential components of modern data science, weather it is statistical analysis or machine learning. Modern technology has facilitated a massive explosion of data, however, such data often contain irrelevant information consequently making prediction difficult. Researchers are devising new methods and algorithms in order to extract information to create robust predictive models. Mostly such models contain predictor variables that are directly or indirectly correlated with other predictor variables. In

*Corresponding Author
Email addresses: raju.rimal@nmbu.no (Raju Rimal), trygve.almoy@nmbu.no (Trygve Almøy), solve.sabo@nmbu.no (Solve Sæbø)

addition studies often constitute of many response variables correlated with each other. These interlinked relationships influence any study, whether it is predictive modeling or inference.

Modern inter-disciplinary research fields such as chemometrics, econometrics and bioinformatics are handling multi-response models extensively. This paper attempts to compare some multivariate prediction methods based on their prediction performance on linear model data with specific properties. The properties includes correlation between response variables, correlation between predictor variables, number of predictor variables and the position of relevant predictor components. These properties are discussed more in the Experimental Design section. Sæbø et al. (2015) and Almøy (1996) have made a similar comparison in the single response setting. In addition, Rimal et al. (2018) has also made a basic comparison on some prediction methods and their interaction with the data properties of a multi-response model. The main aim of this paper is to present a comprehensive comparison of contemporary prediction methods such as simultaneous envelope estimation (Senv) (Cook and Zhang, 2015) and envelope estimation in predictor space (Xenv) (Cook et al., 2010) with customary prediction methods such as Principal Component Regression (PCR), Partial Least Squares Regression (PLS) using simulated dataset with controlled properties. An experimental design and the methods under comparison are discussed further, followed by a brief discussion of the strategy behind the data simulation.

2. Simulation Model

Consider a model where the response vector $(\mathbf{y})$ with m elements and predictor vector $(\mathbf{x})$ with p elements follow a multivariate normal distribution as follows,

$$\begin{bmatrix} \mathbf{y} \\ \mathbf{x} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \boldsymbol{\mu}_y \\ \boldsymbol{\mu}_x \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_{yy} & \boldsymbol{\Sigma}_{yx} \\ \boldsymbol{\Sigma}_{xy} & \boldsymbol{\Sigma}_{xx} \end{bmatrix} \right) \tag{1}$$

where, $\boldsymbol{\Sigma}_{yy}$ and $\boldsymbol{\Sigma}_{xx}$ are the variance-covariance matrices of $\mathbf{y}$ and $\mathbf{x}$, respectively, $\boldsymbol{\Sigma}_{xy}$ is the covariance between $\mathbf{x}$ and $\mathbf{y}$ and $\boldsymbol{\mu}_y$ and $\boldsymbol{\mu}_x$ are mean vectors of $\mathbf{x}$ and $\mathbf{y}$, respectively.

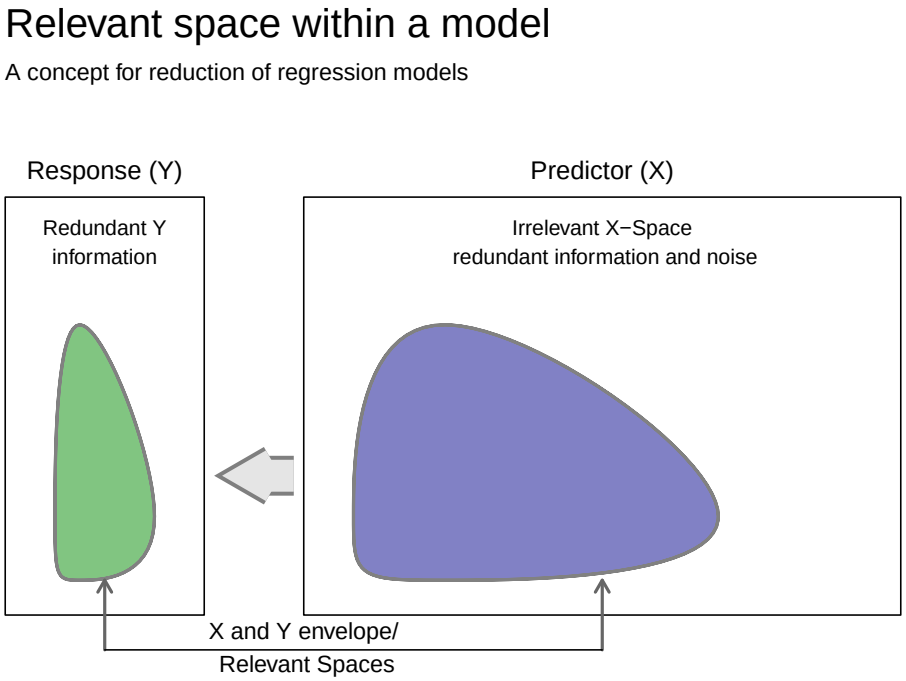

Figure 1: Relevant space in a regression model

A linear model based on (1) is,

$$\mathbf{y} = \boldsymbol{\mu}_y + \boldsymbol{\beta}^t(\mathbf{x} - \boldsymbol{\mu}_x) + \boldsymbol{\epsilon} \tag{2}$$

where, $\underset{m\times p}{\boldsymbol{\beta}^t}$ is a matrix of regression coefficients and $\boldsymbol{\epsilon}$ is an error term such that $\boldsymbol{\epsilon} \sim \mathcal{N}(0, \boldsymbol{\Sigma}_{y|x})$

In a model like (2), we assume that the variation in response $\mathbf{y}$ is partly explained by the predictor $\mathbf{x}$. However, in many situations, only a subspace of the predictor space is relevant for the variation in the response $\mathbf{y}$. This space can be referred to as the relevant space of $\mathbf{x}$ and the rest as irrelevant space. In the similar manner, we can assume that a subset of the response space contains the information that the predictors can explain for a given model (Figure-1). Cook et al. (2010) and Cook and Zhang (2015) have referred to the relevant space as material space, and the irrelevant space as immaterial space.

With an orthogonal transformation of $\mathbf{y}$ and $\mathbf{x}$ to latent variables $\mathbf{w}$ and $\mathbf{z}$, respectively, by $\mathbf{w} = \mathbf{Q}\mathbf{y}$ and $\mathbf{z} = \mathbf{R}\mathbf{x}$, where $\mathbf{Q}$ and $\mathbf{R}$ are orthogonal rotation matrices, an equivalent

model to (2) in terms of the latent variables can be written as,

$$\mathbf{w} = \boldsymbol{\mu}_w + \boldsymbol{\alpha}^t(\mathbf{z} - \boldsymbol{\mu}_z) + \boldsymbol{\tau} \tag{3}$$

where, $\underset{m\times p}{\boldsymbol{\alpha}^t}$ is a matrix of regression coefficients and $\boldsymbol{\tau}$ is an error term such that $\boldsymbol{\tau} \sim \mathcal{N}(0, \boldsymbol{\Sigma}_{w|z})$. Model (3) follows the distribution,

$$\begin{bmatrix} \mathbf{w} \\ \mathbf{z} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \boldsymbol{\mu}_w \\ \boldsymbol{\mu}_z \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_{ww} & \boldsymbol{\Sigma}_{wz} \\ \boldsymbol{\Sigma}_{zw} & \boldsymbol{\Sigma}_{zz} \end{bmatrix} \right) \tag{4}$$

where, $\boldsymbol{\Sigma}_{ww}$ and $\boldsymbol{\Sigma}_{zz}$ are the variance-covariance matrices of $\mathbf{w}$ and $\mathbf{z}$, respectively. $\boldsymbol{\Sigma}_{zw}$ is the covariance between $\mathbf{z}$ and $\mathbf{w}$. $\boldsymbol{\mu}_w$ and $\boldsymbol{\mu}_z$ are mean vector of $\mathbf{z}$ and $\mathbf{w}$ respectively.

Here, the elements of $\mathbf{w}$ and $\mathbf{z}$ are the principal components of responses and predictors, which will respectively be referred as "response components" and "predictor components". The column vectors of respective rotation matrices $\mathbf{Q}$ and $\mathbf{R}$ are the eigenvectors corresponding to these principal components.

Following the concept of relevant space, a subset of predictor components can be imagined to span the predictor space. These components can be regarded as relevant predictor components. Naes and Martens (1985) introduced the concept of relevant components which was explored further by Helland (1990), Næs and Helland (1993), Helland and Almøy (1994) and Helland (2000). The corresponding eigenvectors were referred to as relevant eigenvectors. A similar logic is introduced by Cook et al. (2010) and later by Cook et al. (2013) as an envelope which is the space spanned by the relevant eigenvectors (Cook, 2018, pp. 101).

In addition, various simulation studies have been performed with the model based on the concept of relevant subspace. A simulation study by Almøy (1996) has used a single response simulation model based on reduced regression and has compared some contemporary multivariate estimators. In the recent years Helland et al. (2012), Sæbø et al. (2015), Helland et al. (2018) and Rimal et al. (2018) implemented similar simulation examples as we are discussing in this study. This paper, however, presents an extensive

simulation study based on multi-response data simulated with experimental design and compares relatively new methods such as simultaneous envelopes with well established methods such as partial least squares and principal components regression. Rimal et al. (2018) has a detail discussion about the simulation model that we have opted here. The following section presents the estimators under comparison in more detail.

3. Prediction Methods

Partial least squares regression (PLS) and Principal component regression (PCR) has been used in many disciplines such as chemometrics, econometrics, bioinformatics and machine learning, where wide predictor matrices, i.e. p (number or predictors) > n (number of observation) is common. These methods are popular in multivariate analysis, especially for exploratory studies and prediction.

In recent years, a concept of envelope introduced by Cook et al. (2007) based on reduction in regression model has been implemented for the development of envelope estimation in the subsequent papers.

In this study, we will follow estimation methods based on their prediction performance on data simulated with different controlled properties.

Principal Components Regression (PCR): Principal components are the linear combinations of predictor variables such that the transformation makes the new variables uncorrelated and the variation of the original dataset captured by them are ordered. In other words, each successive components captures maximum variation left by the preceding components in predictor variables (Jolliffe, 2002). Principal components regression uses these principal components to explain the variation in the response.

Partial Least Squares (PLS): Two variants of PLS: PLS1 and PLS2 will be used for comparison. The first one considers individual response variables separately, i.e. each response is predicted with a single response model, while the latter considers all response variables together. In PLS regression the components are determined such

as to maximize a covariance between response and predictors (de Jong, 1993).

Envelopes: The envelope, introduced by Cook et al. (2007), was first used as a response envelope (Cook et al., 2010) as a smallest subspace $\mathcal{E}$ in the response space such that the span of regression coefficients lies in that space. Since a multivariate linear regression model contains relevant (material) and irrelevant (immaterial) variation in both response and predictor, the relevant part provides information, while irrelevant part increases the estimative variation. The concept of envelope uses the relevant part for estimation while excluding the irrelevant part consequently increasing the efficiency of the model (Cook and Zhang, 2016).

The concept was later extended to the predictor space, where the predictor envelope was defined (Cook et al., 2013). Further Cook and Zhang (2015) uses envelopes for joint reduction of the responses and predictors and argued to produce efficiency gains greater than using individual envelops either of the response and predictors. All the variants of envelope estimations are based on maximum likelihood estimation. Here in this study we will also use predictor envelope (Xenv) and simultaneous envelope (Senv) for the comparison.

3.1. Modification in envelope estimation

Since envelope estimators (Xenv and Senv) are based on maximum likelihood estimation (MLE), it fails to estimate in case of wide matrices, i.e. $p > n$. In order to incorporate these methods in our comparison, we have used the principal components $(\mathbf{z})$ of the predictor variables $(\mathbf{x})$ as predictors, using the required number of components for capturing 97.5% of the variation in $\mathbf{x}$. The new set of variables, $\mathbf{z}$, were used for envelope estimation. The regression coefficients $(\hat{\boldsymbol{\alpha}})$ corresponding to these new variables $\mathbf{z}$ were transformed back to obtain coefficients for each predictor variable as,

$$\hat{\boldsymbol{\beta}} = \mathbf{e}_k \hat{\boldsymbol{\alpha}_k}$$

where, $\mathbf{e}_k$ is the eigenvectors with k number of components.

4. Experimental Design

This study compares prediction methods based on their prediction ability. Data with specific properties are simulated, some of which are easier to predict than others. These data are simulated using the R-package `simrel`, which is discussed in Sæbø et al. (2015) and Rimal et al. (2018). Here we will use four different factors to vary the property of the data: a) Number of predictors (`p`), b) Multicollinearity in predictor variables (`gamma`), c) Correlation in response variables (`eta`) and d) position of predictor components relevant for the response (`relpos`). Using two levels of `p`, `gamma` and `relpos` and four levels of `eta`, 32 set of distinct properties are designed for the simulation.

Number of predictors: In order to observe the performance of the methods on tall and wide predictor matrices, 20 and 250 predictor variables are simulated. Parameter `p` controls this properties in the `simrel` function.

Multicollinearity in predictor variables: Highly collinear predictors can be explained completely by few components. The parameter `gamma` (γ) in `simrel` controls decline in the eigenvalues of the predictor variables as (5).

$$\lambda_i = e^{-\gamma(i-1)}, \gamma > 0 \text{ and } i = 1, 2, \ldots, p \tag{5}$$

Here, $\lambda_i, i = 1, 2, \ldots p$ are eigenvalues of the predictor variables. Here we have used 0.2 and 0.9 as different levels of `gamma`. The higher the value of gamma, the higher will be the correlation between predictors and vice versa.

Correlation in response variables: Correlation among response variables is a less explored area. Here we have tried to explore that part with 4 levels of correlation in the response variables. We have used the `eta` (η) parameter of `simrel` for controlling the decline in eigenvalues corresponding to the response variables as (6).

$$\kappa_i = e^{-\eta(i-1)}, \eta > 0 \text{ and } j = 1, 2, \ldots, m \tag{6}$$

Here, $\kappa_i, i = 1, 2, \ldots m$ are the eigenvalues of the response variables and `m` is the number of response variables. Here we have used 0, 0.4, 0.8 and 1.2 as different levels of `eta`. The larger the value of eta, the larger will be the correlation between response variables and vice versa.

Position of predictor components relevant to the response: The principal components of the predictors are ordered. The first principal component captures most of the variation in the predictors. The second captures the most in the rest that is left by the first principal components and so on. In highly collinear predictors, the variation captured by the first few components is relatively high. However, if those components are not relevant for the response, prediction becomes difficult (Helland and Almøy, 1994). Here, two levels of the positions of these relevant components are used: 1, 2, 3, 4 and 5, 6, 7, 8.

Further, a complete factorial design from the levels of the above given parameters gave us 32 designs. Each design is associated with a dataset having unique properties. Figure~2, shows all the designs. For each design and prediction method, 50 datasets were simulated for replication. In total, there were $5 \times 32 \times 50$, i.e. 8000 dataset simulated.

Common parameters: Each dataset was simulated with $n = 100$ number of observation and $m = 4$ response variables. Further, the coefficient of determination corresponding to each response components in all the designs is set to and 0.8. In addition, we have assumed that there is only one informative response component. Hence, the informative response component is rotated orthogonally together with three uninformative response components to generate four response variables. This spread out the information in all simulated response variables. For further details on the simulation tool see (Rimal et al., 2018).

An example of simulation parameters for the first design is as follows:

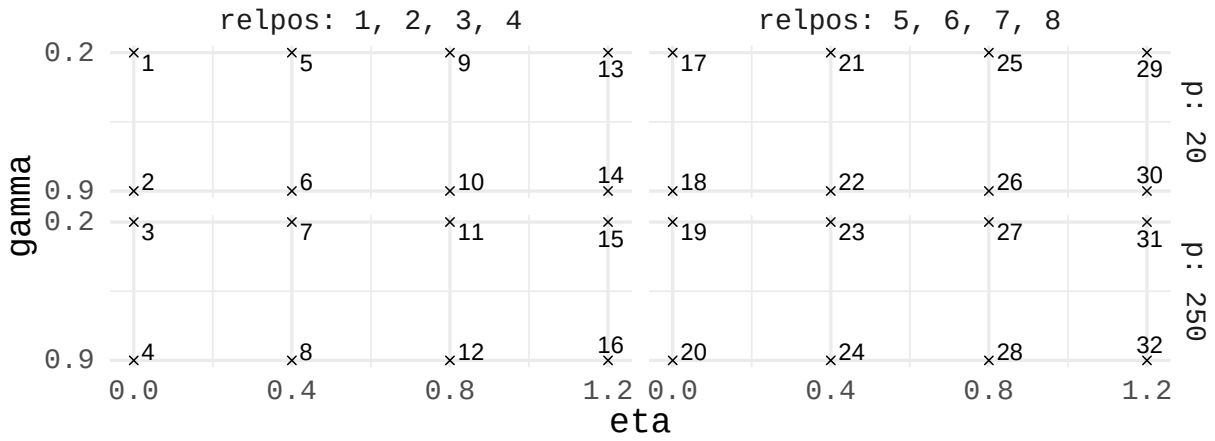

Figure 2: Experimental Design of simulation parameters. Each point represents an unique data property.

```
simrel(
    n       = 100,                  ## Training samples
    p       = 20,                   ## Predictors
    m       = 4,                    ## Responses
    q       = 20,                   ## Relevant predictors
    relpos  = list(c(1, 2, 3, 4)), ## Relevant predictor components index
    eta     = 0,                    ## Decay factor of response eigenvalues
    gamma   = 0.2,                  ## Decay factor of predictor eigenvalues
    R2      = 0.8,                  ## Coefficient of determination
    ypos    = list(c(1, 2, 3, 4)),
    type    = "multivariate"
)
```

Figure 3 shows the covariance structure of the data simulated with this design. The figure shows that the predictor components at position 1, 2, 3 and 4 are relevant for the first response component. After the rotation with orthogonal rotation matrix, all predictors

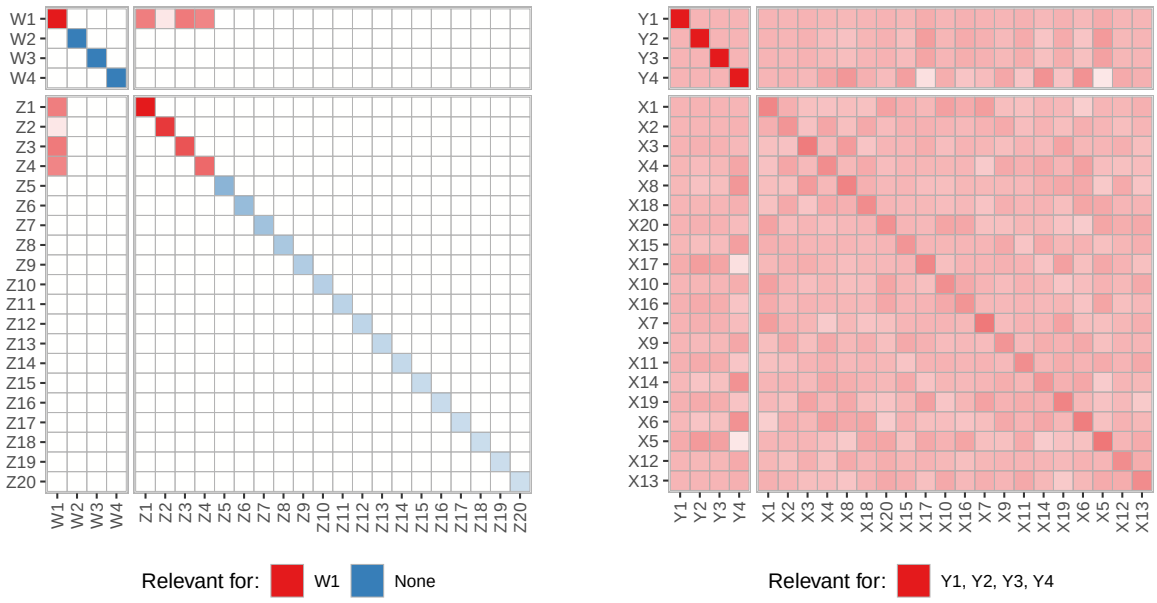

Figure 3: (left) Covariance structure of latent components. (right) Covariance structure of predictor and response

are somewhat relevant for all response variables, fulfilling other desired properties like multicollinearity and coefficient of determination. For this same design, Figure 4(top left) shows that the predictor components 1, 2, 3 and 4 are relevant for the first response component. All other predictor components are irrelevant and all other response components are uninformative. However, due to orthogonal rotation of the informative response component together with uninformative response components, all response variables in the population have similar covariance with the relevant predictor components (Figure 4(top right)). The sample covariances between the predictors components and predictor variables with response variables are in Figure 4 (bottom left) and (bottom right) respectively. The discussion here is made on the first design. A similar discussion can be made on all 32 designs where each of the design holds the properties of the data they simulate. These data are used by the prediction methods discussed in previous section. Each prediction methods are given independent dataset simulated in order to give them equal opportunity to understand the dynamics in the data.

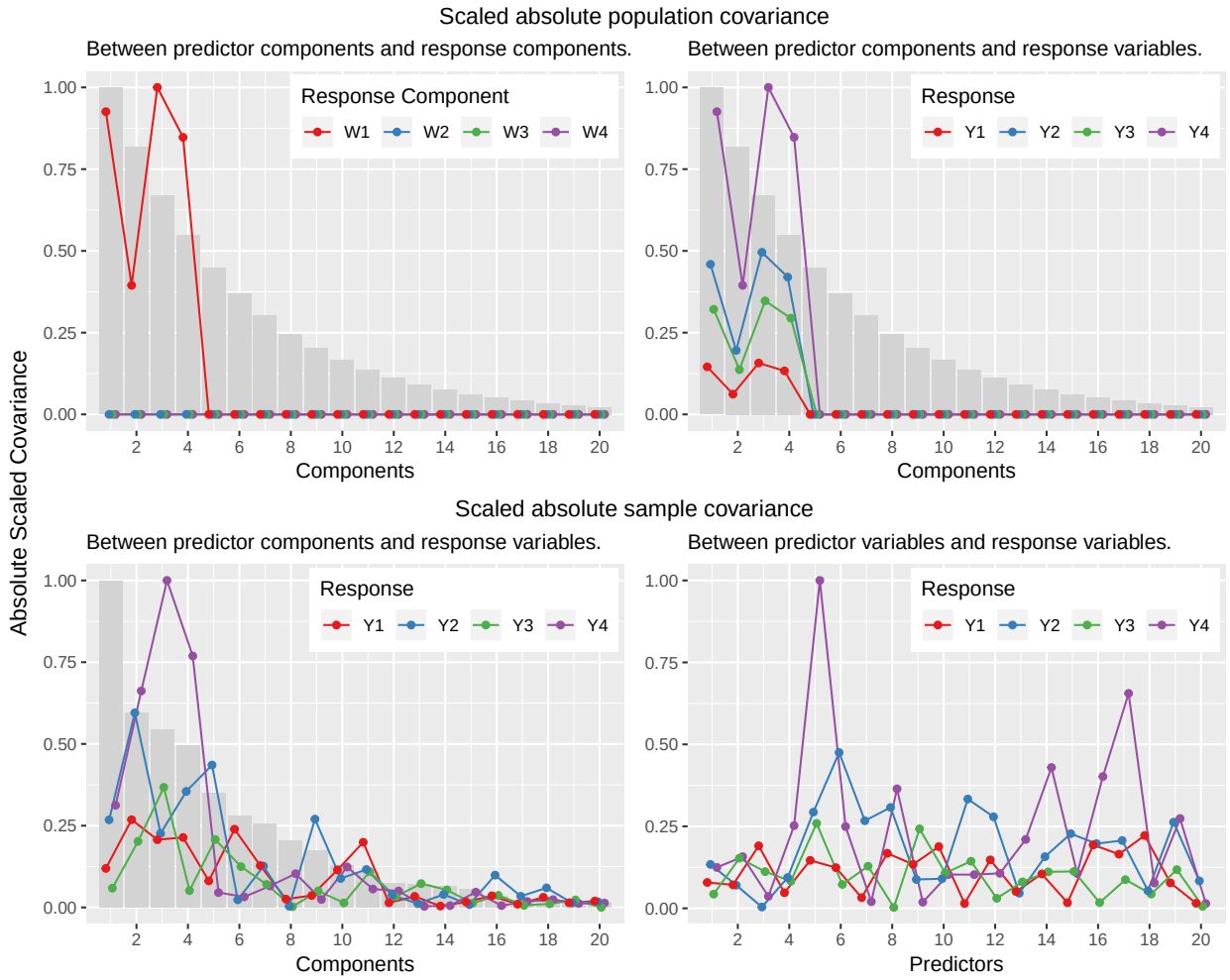

Figure 4: Expected Scaled absolute covariance between predictor components and response components (top left). Expected Scaled absolute covariance between predictor components and response variables (top right). Sample scaled absolute covariance between predictor components and response variables (bottom left). Sample scaled absolute covariance between predictor variables and response variables (bottom right). The bar in the background are eigenvalues corresponding to each components in population (top plots) and in sample (bottom plots).

4.1. *Basis of comparison*

This study focuses mainly on the prediction performance of the methods and emphasis specifically on the interaction between the properties of the data, controlled by the simulation parameters, and the prediction methods. The prediction performance is measured by the prediction error for each response as in (7). The prediction is the theoretically computed expected prediction when the model is applied to unseen observations corresponding to each response variable.

$$\text{prediction error}_j = \frac{1}{\sigma^2_{y_j|x}} \left[\left(\boldsymbol{\beta}_j - \hat{\boldsymbol{\beta}}_j \right)^t \boldsymbol{\Sigma}_{xx} \left(\boldsymbol{\beta}_j - \hat{\boldsymbol{\beta}}_j \right) \right] + 1 \tag{7}$$

where, $\boldsymbol{\Sigma}_{xx}$ is the true covariance matrix of predictor and $\sigma_{y_j|x}$ is the true model error obtained from simulation of response $j = 1, \ldots m$. Here prediction error is scaled by the true model error to remove its effect on prediction error. (Need better reason.) The prediction error in (7) is computed for all replications of 32 designs.

5. Exploration

The structure of final data for further analysis contains five factors, prediction methods, number of components, replications and prediction error for four responses. The prediction error is computed for four responses using 0 to 10 predictor components for each 50 replicates using (7). Thus there are 32 (design) $\times$ 5 (methods) $\times$ 11 (number of components) $\times$ 50 (replications), i.e. 88000 observations. Here the variables `Y1` to `Y4` corresponds to prediction error of respective response variables.

Here we will not only focus on the minimum prediction error that a method can obtain but also the number of components they use to get that error. So, the dataset discussed above is summarized to construct following two smaller datasets. Let us call them *Error Dataset* and *Component Dataset*.

- *Error Dataset*: One with minimum prediction error that a method can give using arbitrary number of components

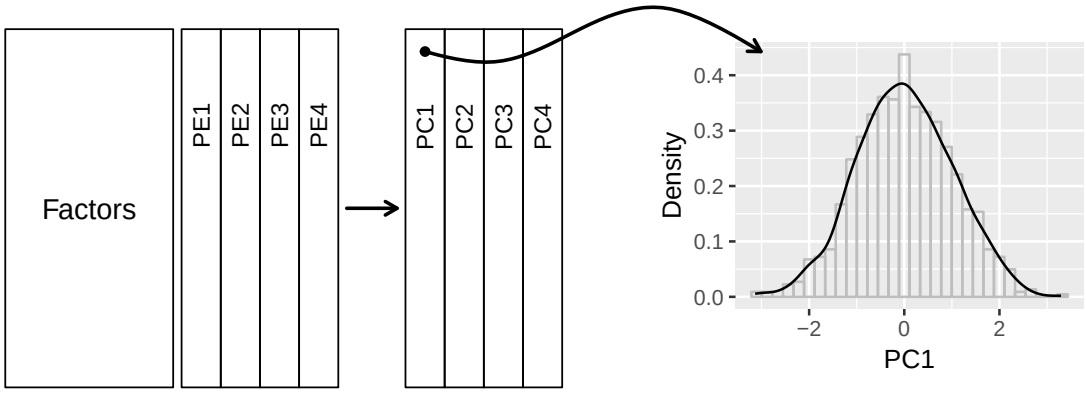

Figure 5: Analysis of density of principal components of prediction erorr ror each response.

- *Component Dataset*: Another with the number of components that the method has used to give that minimum prediction error

The Statistical Analysis section will perform multivariate analysis using MANOVA on these datasets however, this section focus more on exploring the variation in the prediction error and number of components due to our design parameters and prediction methods through plots.

Here we have a) four vectors of minimum prediction error and b) four vectors of corresponding number of components. The following exploration is based on the scores of principal components of these two matrices. The analysis will progress as in the Figure 5 in the case of both datasets.

Figure 6 plots the score density from first principal component of minimum prediction error. Since higher prediction error results in high scores the plot shows that the PCR, PLS1 and PLS2 methods are influenced by two levels of position of relevant predictor components. When the position of relevant predictors are at positon 5, 6, 7, 8, the eigenvalues corresponding to them becomes smaller making those designs difficult to model. However, the envelope methods have less influence of `relpos` in this regard.

In addition, the plot also shows that the effect of `gamma`, the level of multicollinearity, has

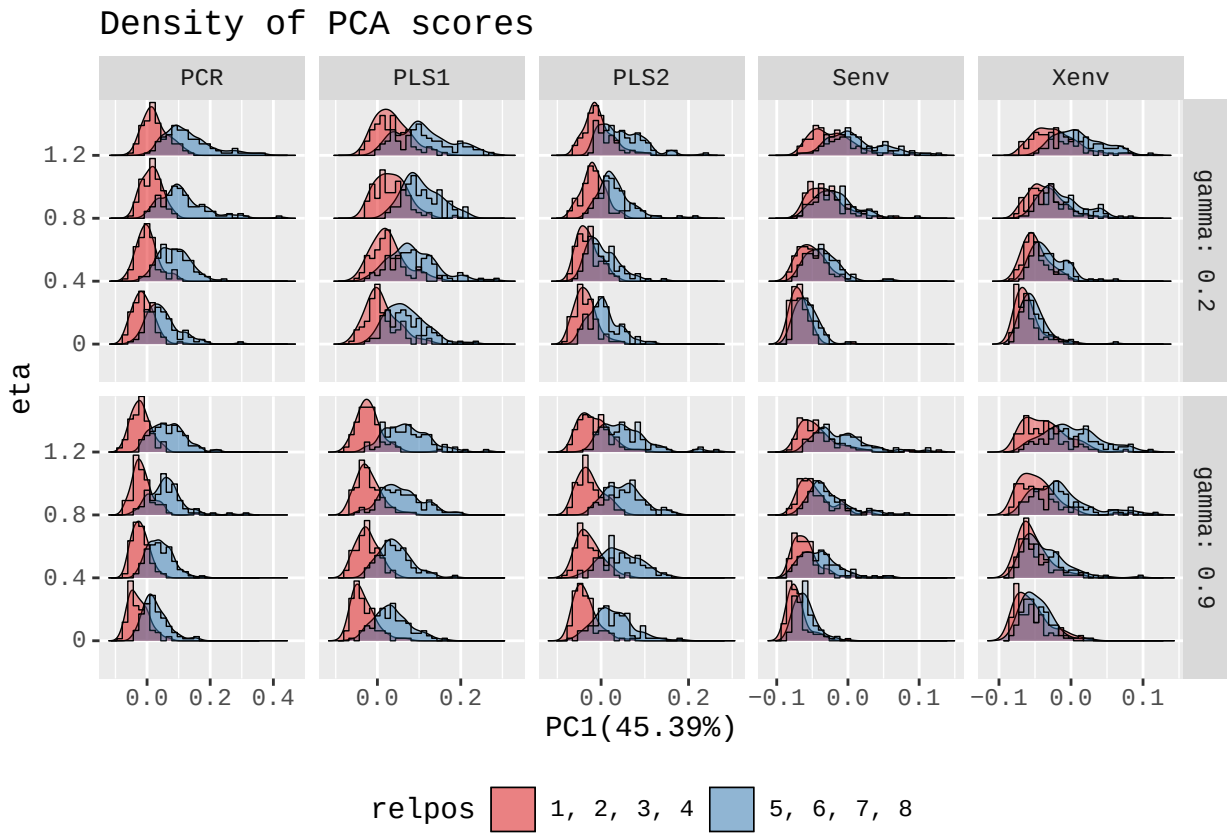

Figure 6: Density of first principal component of prediction error matrix subdivided by methods, gamma and eta and grouped by relpos.

smaller effect in all cases. This indicates that the methods are somewhat robust to handle collinear predictors.

Further, the density curve for PCR, PLS1 and PLS2 for different levels of `eta`, the factor controlling the correlation between responses, are similar. However, this is not true for envelope models. The envelope methods have shown to have significant interaction between position of relevant components and `eta`. Here higher levels of `eta` is giving larger scores and clear separation between two level of `relpos`. This behavior is expected in the simultaneous envelope as the method has claimed to model relevant response (material) response space.

Figure 7 plots the score density from first principal component of matrix with number of components used to get minimum prediction error. Here, the higher scores suggest

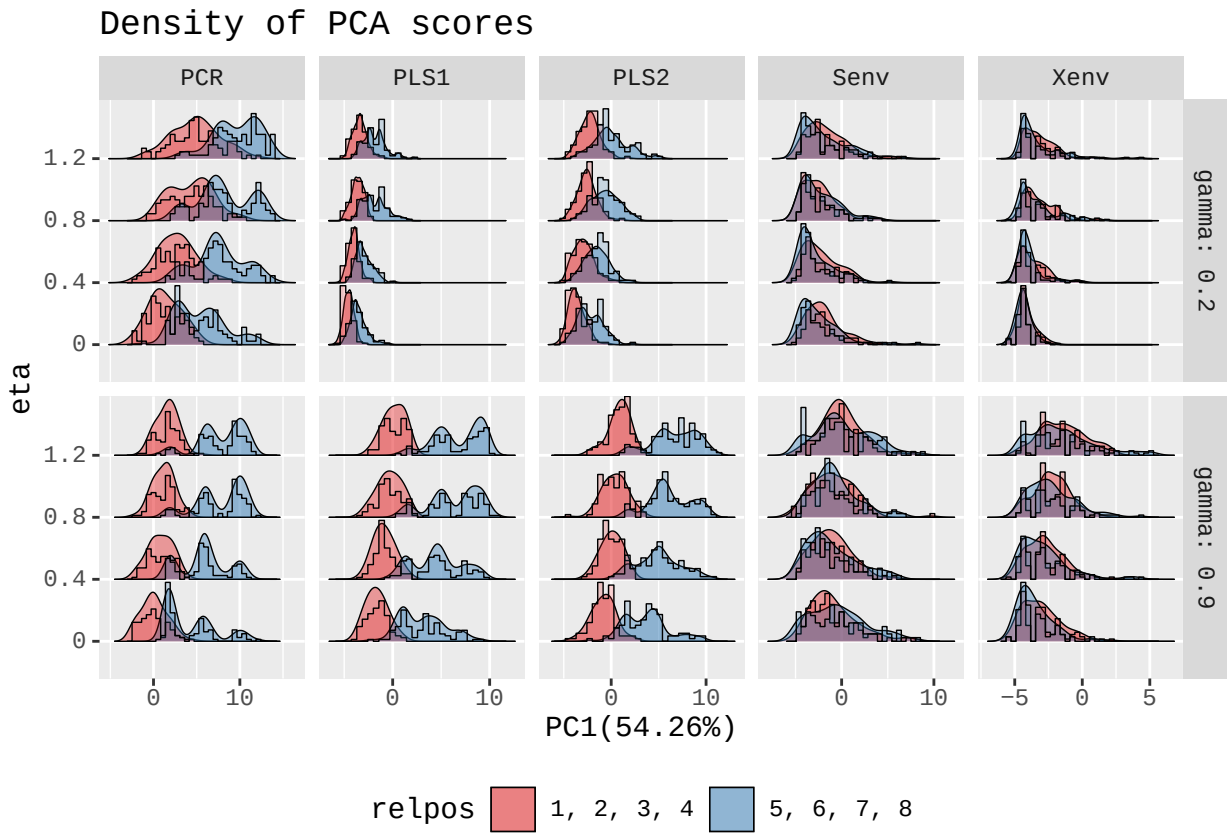

Figure 7: Density of first principal component of the matrix of number of principal components used to give minimum prediction error subdivided by methods, gamma and eta and grouped by relpos.

the use of more components to give minimum prediction error. The plot shows that the relevant predictor components at 5, 6, 7, 8 gives larger prediction error than that are at the position 1, 2, 3, 4. The pattern is more distinct in large multicollinearity case and PCR and PLS methods.

The plot also shows noticeable results in the case of envelope methods. The methods have shown equally better performance at both levels of `relpos` and `gamma`.

6. Statistical Analysis

The datasets discussed in previous section a) *error dataset* and b) *component dataset* are used on a multivariate analysis of variance. Let us call them *error model* and *component model* respectively. The models are fitted with the third order interaction of simulation

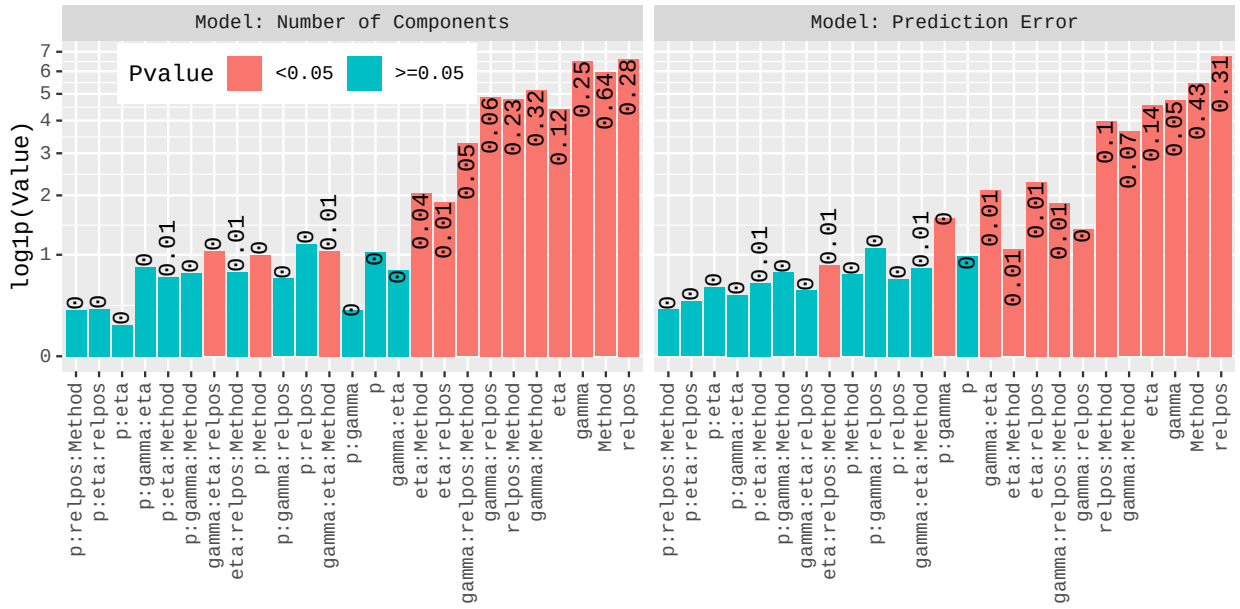

Figure 8: Pillai Statistic and F-value for the MANOVA model. The bar represents the F-value and the text labels are Pillai Statistic for corresponding factor.

parameters (p, gamma, eta and relpos) and Methods as in (8).

$$\mathbf{y}_{abcdef} = \boldsymbol{\mu} + (\mathtt{p}_a + \mathtt{gamma}_b + \mathtt{eta}_c + \mathtt{relpos}_d + \mathtt{Methods}_e)^3 + \boldsymbol{\varepsilon}_{abcdef} \tag{8}$$

where, $\mathbf{y}_{abcdef}$ is a vector of prediction error for factors,

- $\mathtt{p}_a = 20$ and 250
- $\mathtt{gamma}_b = 0.2$ and 0.9
- $\mathtt{eta}_c = 0$, 0.4, 0.8 and 1.2
- $\mathtt{relpos}_d = 1, 2, 3, 4$ and $5, 6, 7, 8$
- $\mathtt{Methods}_e =$ PCR, PLS1, PLS2, Xenv and Senv

Figure 8 (left) shows the result from *component model* and the figure 8 (right) shows the result from *error model*. The figure shows that all main effects except p are significant and has large effect on *component model* as well as *error model*.

In addition, the position of relevant components have largest effect on *error model*. In case of *component model*, multicollinearity also have large effect in addition to the position

of relevant components. However based on Pillai trace statistic, `Method` has the largest effect on both of the models. Further, the interactions `p:gamma` and `eta:relpos:Method` has significant effect on the *error model* but not in the *component model*. However, all of these interactions have small pillai statistic.

Further, in the case of *component model* interaction effects `gamma:eta:Method`, `p:Method` and `gamma:eta:relpos` are significant but not in the case of *error model*. Similar to previous point, they too have small pillai statistic. In the case of *error model*, the interaction effect of `eta:relpos:Method` which is particularly interesting to see that the methods performed differently for different cases of collinearity in response and collinearity in predictors.

Following section will continue on identifying the effect of different levels of factors in the case of these interactions.

6.1. Effect Analysis

Error Model:

Figure 9 (left) shows clear difference of `eta` for a given `relpos`. The plot also shows a clear difference in effect of methods on prediction error. Figure 9 (right) shows effect of `gamma` for a given `method` and `eta`. It shows that these methods gives low prediction error is high multicollinear situations.

Component Model:

Figure 10 (left) shows that `xenv` has used minimum number of components followed by `senv` methods than others in order to get their minimum prediction error. The same figure also suggest that the minimum number of components used by `PLS1`, `PLS2` and `PCR` vary and has high effect of `eta`. The `senv` model which consider both X and Y correlation structure while estimating regression coefficients has smallest variation of number of components used for different `eta` parameters.

Figure 10 (right) shows that in case of low multicollinearity in the model `PLS` methods are used less number of components than `PCR`. This is expected since, `PLS` methods consider the covariance structure of predictor and response which `PCR` does not.

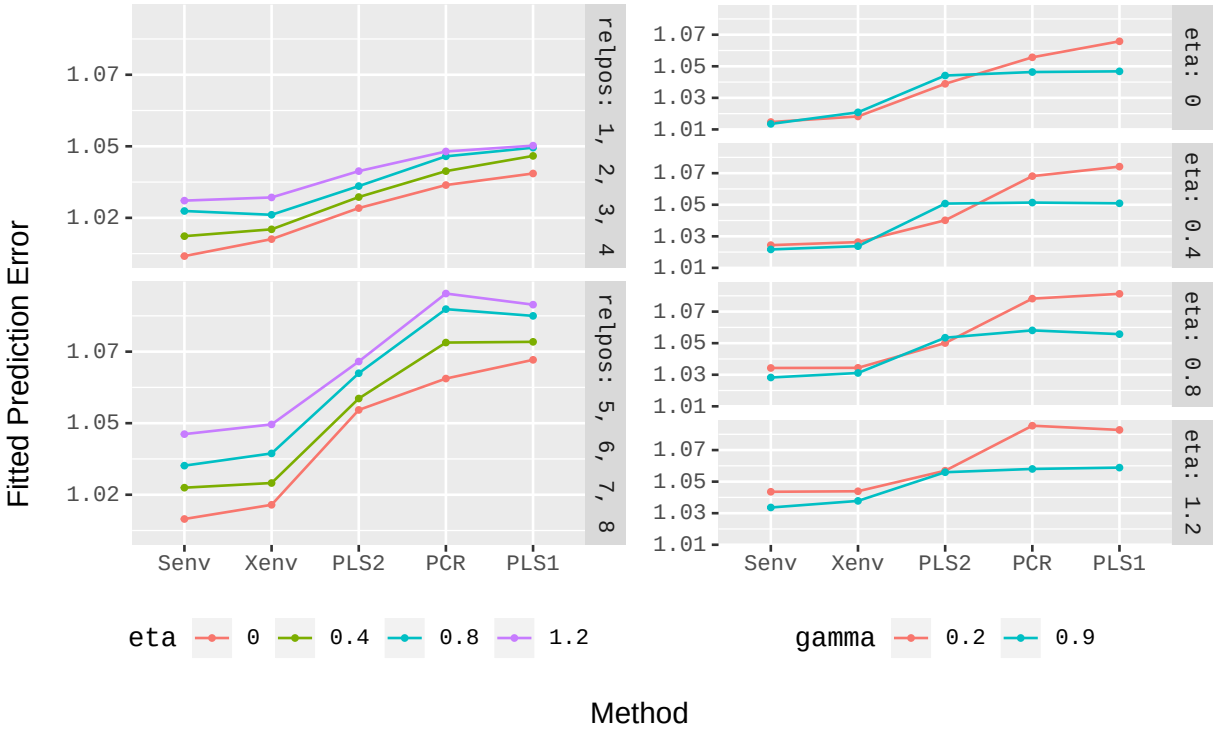

Figure 9: Effect plot of some interactions of the multivariate linear model

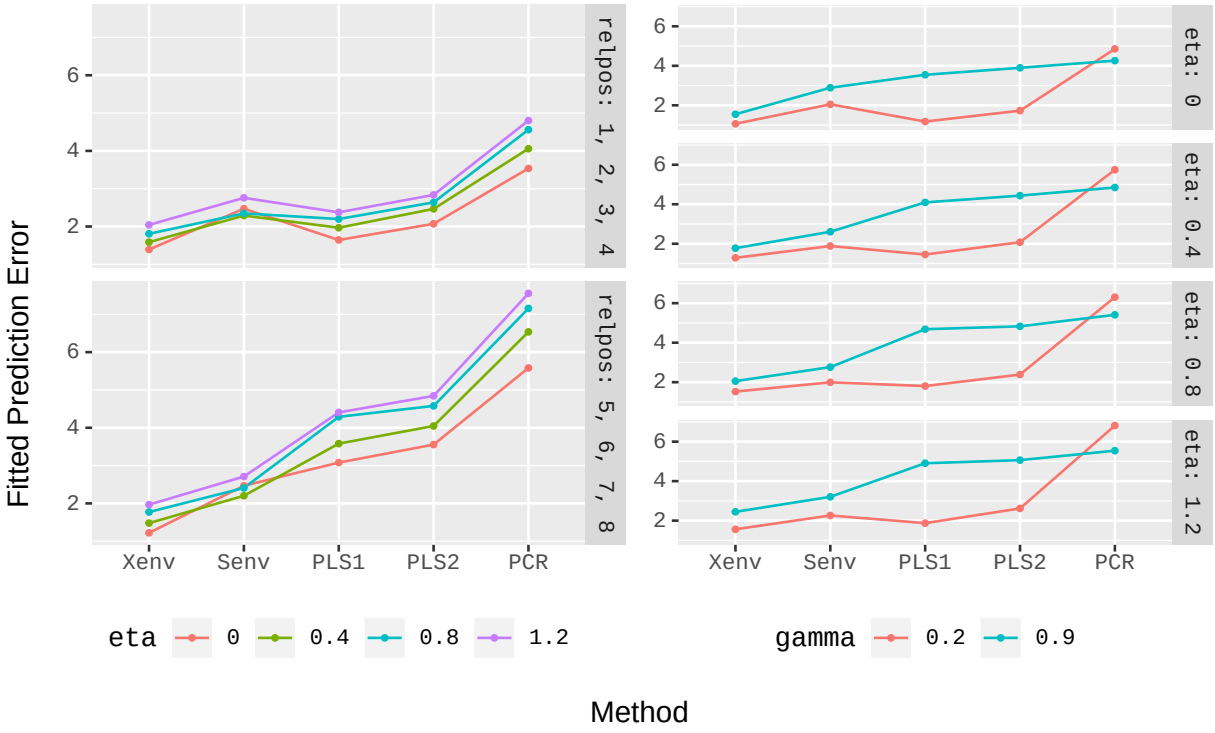

Figure 10: Effect plot of some interactions of the multivariate linear model

References

Almøy, T., 1996. A simulation study on comparison of prediction methods when only a few components are relevant. Computational Statistics & Data Analysis 21, 87–107. doi:doi:10.1016/0167-9473(95)00006-2.

Cook, R.D., 2018. An introduction to envelopes : dimension reduction for efficient estimation in multivariate statistics. 1 ed., Hoboken, NJ : John Wiley & Sons, 2018.

Cook, R.D., Helland, I.S., Su, Z., 2013. Envelopes and partial least squares regression. Journal of the Royal Statistical Society. Series B: Statistical Methodology 75, 851–877. doi:doi:10.1111/rssb.12018.

Cook, R.D., Li, B., Chiaromonte, F., 2007. Dimension reduction in regression without matrix inversion. Biometrika 94, 569–584. doi:doi:10.1093/biomet/asm038.

Cook, R.D., Li, B., Chiaromonte, F., 2010. Envelope Models for Parsimonious and Efficient Multivariate Linear Regression. Statistica Sinica 20, 927–1010.

Cook, R.D., Zhang, X., 2015. Simultaneous envelopes for multivariate linear regression. Technometrics 57, 11–25. doi:doi:10.1080/00401706.2013.872700.

Cook, R.D., Zhang, X., 2016. Algorithms for Envelope Estimation. Journal of Computational and Graphical Statistics 25, 284–300. doi:doi:10.1080/10618600.2015.1029577, arXiv:1403.4138.

Helland, I.S., 1990. Partial least squares regression and statistical models. Scandinavian Journal of Statistics 17, 97–114. doi:doi:10.2307/4616159.

Helland, I.S., 2000. Model Reduction for Prediction in Regression Models. Scandinavian Journal of Statistics 27, 1–20. doi:doi:10.1111/1467-9469.00174.

Helland, I.S., Almøy, T., 1994. Comparison of prediction methods when only a few components are relevant. Journal of the American Statistical Association 89, 583–591. doi:doi:10.1080/01621459.1994.10476783.

Helland, I.S., Saebø, S., Almøy, T., Rimal, R., Sæbø, S., Almøy, T., Rimal, R., 2018. Model and estimators for partial least squares regression. Journal of Chemometrics 32, e3044. doi:doi:10.1002/cem.3044.

Helland, I.S., Saebø, S., Tjelmeland, H.K., 2012. Near Optimal Prediction from Relevant Components. Scandinavian Journal of Statistics 39, 695–713. doi:doi:10.1111/j.1467-9469.2011.00770.x.

Jolliffe, I.T., 2002. Principal Component Analysis, Second Edition. doi:doi:10.2307/1270093, arXiv:arXiv:1011.1669v3.

de Jong, S., 1993. SIMPLS: An alternative approach to partial least squares regression. Chemometrics and Intelligent Laboratory Systems 18, 251–263. doi:doi:10.1016/0169-7439(93)85002-X.

Næs, T., Helland, I.S., 1993. Relevant components in regression. Scandinavian Journal of Statistics 20, 239–250.

Naes, T., Martens, H., 1985. Comparison of prediction methods for multicollinear data. Communications in Statistics - Simulation and Computation 14, 545–576. doi:doi:10.1080/03610918508812458.

Rimal, R., Almøy, T., Sæbø, S., 2018. A tool for simulating multi-response linear model data. Chemometrics and Intelligent Laboratory Systems 176, 1–10. doi:doi:10.1016/j.chemolab.2018.02.009.

Sæbø, S., Almøy, T., Helland, I.S., 2015. Simrel - A versatile tool for linear model data simulation based on the concept of a relevant subspace and relevant predictors. Chemometrics and Intelligent Laboratory Systems 146, 128–135. doi:doi:10.1016/j.chemolab.2015.05.012.