

PhD Midway Seminar

Simulation Tool and its application

Raju Rimal

Supervisors

Solve Sæbø, Tryge Almøy

02 March, 2017



Norges miljø- og
biovitenskapelige
universitet

Introduction

- Make Simulation Tools for multi-response linear model data

- Make Simulation Tools for multi-response linear model data
- Using the tool, compare various estimation techniques and understand them

- Make Simulation Tools for multi-response linear model data
- Using the tool, compare various estimation techniques and understand them
- Extend the simulation tool incorporating model with background information

- Make Simulation Tools for multi-response linear model data
- Using the tool, compare various estimation techniques and understand them
- Extend the simulation tool incorporating model with background information
- Apply this extended tool to test multi-matrix extension of PLS models such as LPLS and UPLS

- Advanced Multivariate Model and technique to analyze it

- Advanced Multivariate Model and technique to analyze it
- Programming concept for developing statistical packages and applications for various statistical methods

- Advanced Multivariate Model and technique to analyze it
- Programming concept for developing statistical packages and applications for various statistical methods
- Extending and improving existing methods in statistics

- Advanced Multivariate Model and technique to analyze it
- Programming concept for developing statistical packages and applications for various statistical methods
- Extending and improving existing methods in statistics
- And, Obviously, to properly document what I have done

Today I will talk about:

- Simulation tool (`simulatr`) we are building

Today I will talk about:

- Simulation tool (`simulatr`) we are building
- A competitive study of various estimation techniques by simulating linear model data using `simulatr`

`simrel-m`: A versatile tool for simulating multi-response linear model data

- An extension of `simrel` [Sæbø et al., 2015] r-package

- An extension of `simrel` [Sæbø et al., 2015] r-package
- Uses the idea of reduction of random regression model by separating latent space of \mathbf{X} into subspaces that is relevant and irrelevant for predicting each response

- An extension of `simrel` [Sæbø et al., 2015] r-package
- Uses the idea of reduction of random regression model by separating latent space of \mathbf{X} into subspaces that is relevant and irrelevant for predicting each response
- The underlying concept is based on reparameterizing the population model,

$$\mathbf{Y} = \boldsymbol{\mu}_Y + \mathbf{B}^t (\mathbf{X} - \boldsymbol{\mu}_X) + \boldsymbol{\epsilon}$$

where, $\boldsymbol{\epsilon} \sim N(0, \boldsymbol{\Sigma}_{Y|X})$

- Collect population input parameter from users such as: number of variables, coefficient of determination and the position of relevant components

- Collect population input parameter from users such as: number of variables, coefficient of determination and the position of relevant components
- Make a covariance matrix satisfying input parameters

- Collect population input parameter from users such as: number of variables, coefficient of determination and the position of relevant components
- Make a covariance matrix satisfying input parameters
- Rotate the covariance matrix orthogonally

- Collect population input parameter from users such as: number of variables, coefficient of determination and the position of relevant components
- Make a covariance matrix satisfying input parameters
- Rotate the covariance matrix orthogonally
- Sample calibration and validation sets

A comparative study of different estimation methods using simulated data

Four estimation methods were considered

- Although unbiased, suffer highly from multicollinearity
- Widely used and can be used as reference for comparison
- Based on Maximum Likelihood but works better than OLS in p approaches n
- Our analysis is based on envelope of predictor or x envelope
- Well established and widely used method
- Based on Latent Structure and free of multicollinearity problem
- Bayesian Estimation of regression coefficient
- Promising performance was shown in previous studies [Helland et al., 2012]

From the possible combination of following parameter combination, 32 calibration sets were simulated with 5 replication of each.

Number of sample observation	50	50	50	50
Number of predictor variables	15	40	15	40
Coefficient of determination	0.5	0.9	0.5	0.9
Position of relevant component	1, 2	1, 3	2, 3	1, 2, 3
Decaying factor of eigenvalues	0.5	0.9	0.5	0.9

A Systematic Comparison



Conclusion

grow

shrink

fade-out

fade-up (also down, left and right!)

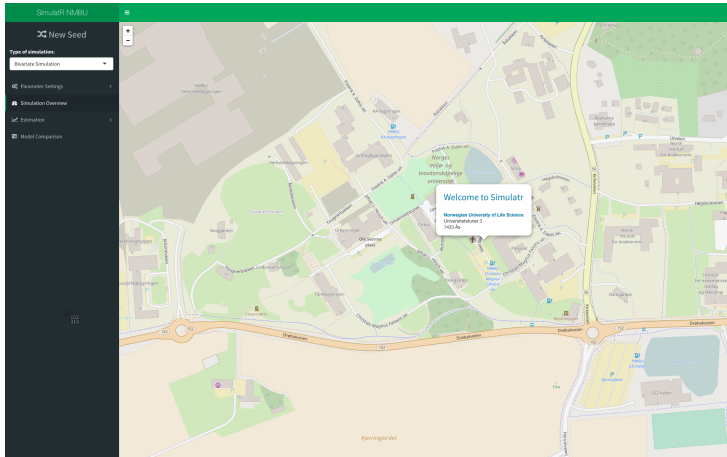
visible only once

blue only once

highlight-red

highlight-green

highlight-blue





References

Inge S Helland, Solve Sæbø, Ha Tjelmeland, et al. Near optimal prediction from relevant components. *Scandinavian Journal of Statistics*, 39(4):695–713, 2012.

Solve Sæbø, Trygve Almøy, and Inge S Helland. 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, 2015.