PhD Midway Seminar

Simulation Tool and its application

Raju Rimal

Supervisors:

Solve Sæbø Tryge Almøy

07 March, 2017

Introduction

My PhD Plan

PhD Program	
Phase 1	Make a simulation Tool
Phase 2	Apply it for comparing different estimation Methods
Phase 3	Extend the simulation tool for model with background information
Phase 4	Apply it to test multi-matrix extension of PLS models such as LPLS and UPLS

Why I am doing this

Important for:

- Research
- Education and
- Method Evaluation

What I learn

- 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's Special

Today I will talk about:

- A comparative study of various estimation techniques by simulating linear model data using simulatr in single response situation Demonstration
- Simulation tool (simulatr) we are building

A comparative study of different estimation methods using simulated data

Overview

Four estimtion methods were considered

Ordinary Least Squares (OLS)

- Although **unbiased**, suffer highly from **multicollinearity**
- Widely used and can be used as **reference for comparison**

Partial Least Squares (PLS)

- **Well established** and widely used method
- Based on Latent Structure and **free of multicollinearity problem**

Overview

Four estimation methods were considered

Envelope

- Relatively **new method** (Cook, Helland, & Su, 2013) and is also based on reduction of regression model
- Based on **Maximum Likelihood** but works better than OLS in p approaches n

Bayes PLS

- **Bayesian Estimation** of regression coefficient
- **Promising performance** was shown in previous studies (I. S. Helland, Sæbø, & Tjelmeland, 2012)

Simulation Design

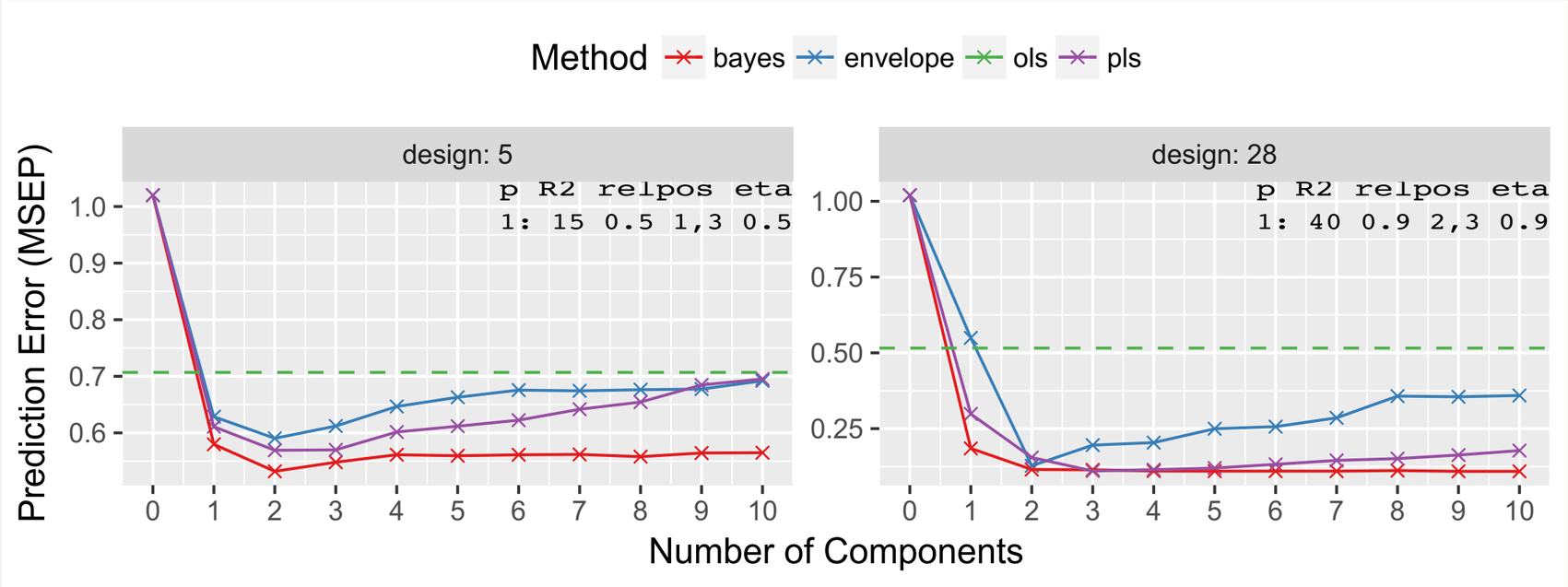
Population Parameters were set as follows:

- **Number of sample observations**: *50*
- **Number of predictor variables**: *15* and *40*
- **Coefficient of determination** (R^2): *0.5* and *0.9*
- **Level of multicollinearity**: *0.5* and *0.9*
- **Position of relevant components**: *1* and *2*; *1* and *3*; *2* and *3*; *1*, *2* and *3*

From the combination of above parameters, **32** datasets were simulated with **5** replication of each, i.e. **160 datasets** with 5 of them having similar population properties.

A Systematic Comparison

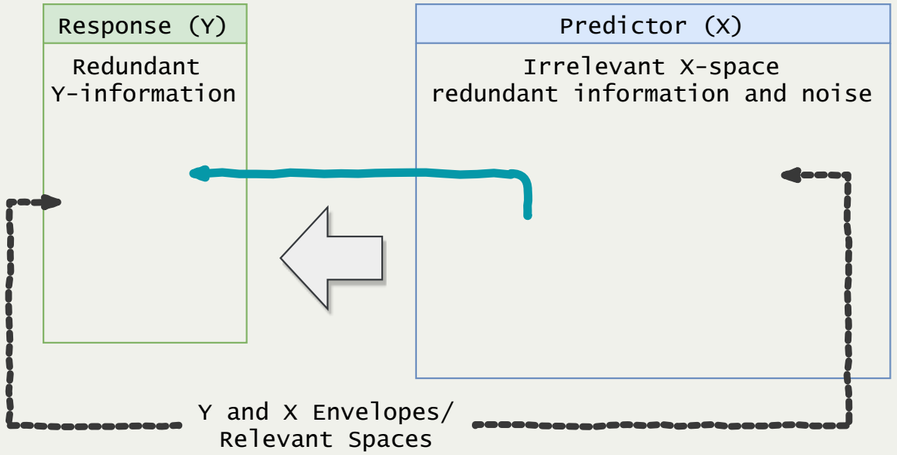
A Systematic Comparison



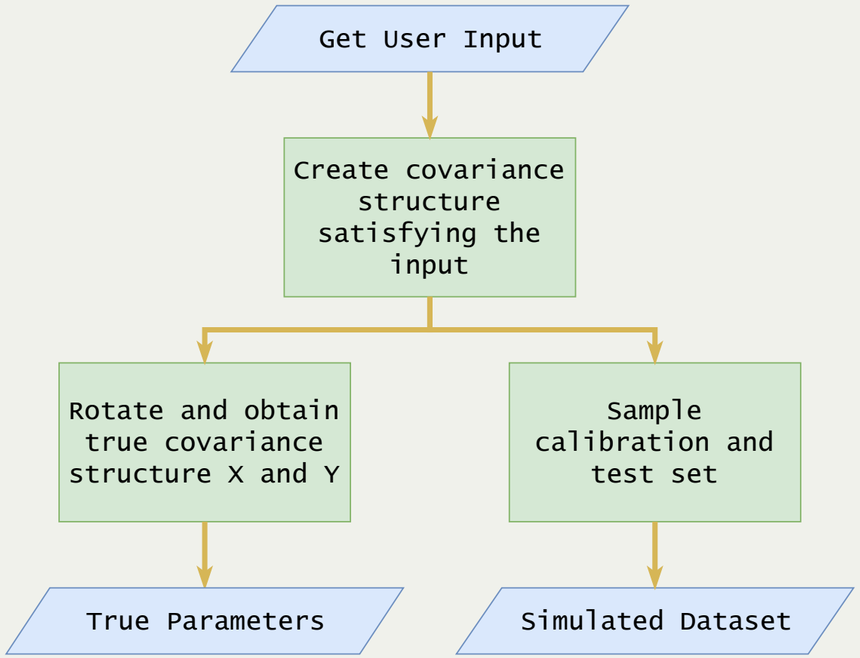
simrel-m: A versatile tool for simulating multi-response linear model data

simrel-m

It is an extension of **simrel** *(Sæbø, Almøy, & Helland, 2015) r-package for simulating* ***multi-response data***

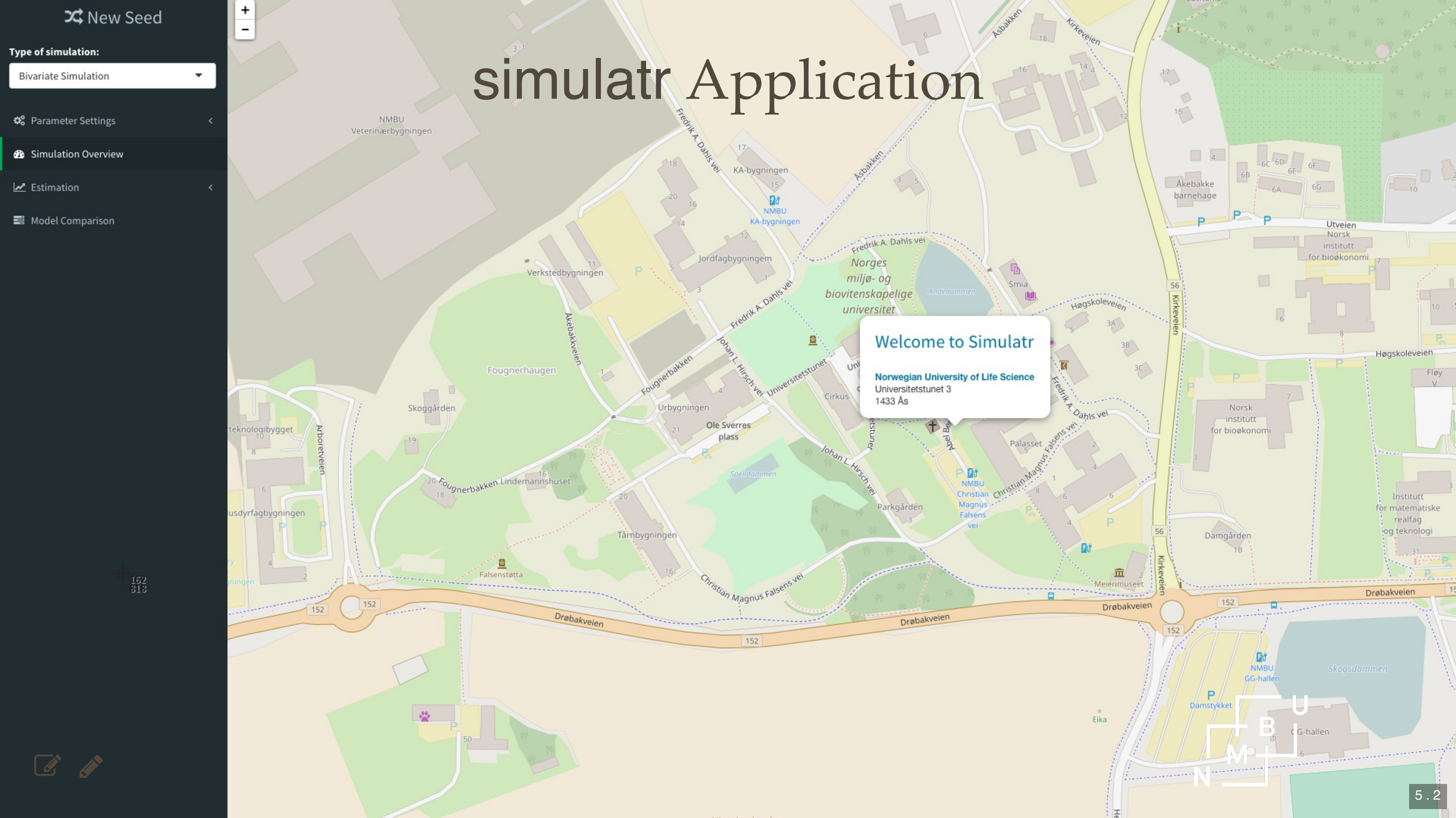


How it works



- Collect input parameters from user
- Make a covariance matrix satisfying those input parameters
- Computes true population properties such as regression coefficients
- Sample calibration and validation sets

Demonstration


New Seed
Type of simulation:
Bivariate Simulation
Parameter Settings
Simulation Overview
Estimation
Model Comparison
simulatr Application
Welcome to Simulatr
Norwegian University of Life Science
Universitetstunet 3
1433 Ås
Norges miljø- og biovitenskapelige universitet
Drøbakveien
Fougnerhaugen
Skogsdammen
5.2


Salamat
Dakujem
GRACIAS
ASANTE
Teşekkür ederim
SUWUN
TAK
hvala
Salamat
TAKK
HVALA
mersi
شكرا جزيلا
Ευχαριστώ
감사합니다
GRAZZii
DANKE
GRAZAS
kiitos
merci
Salamat
TAKK
MAHALO
Рахмет
Thank You
kiitos
arigato
ARIGATO
takk
suwun
धन्यवाद
HVALA
GRAZIE
DAKUJEM
MERCI
Teşekkür ederim
kiitos
hvala
mahalo
GRACIAS
GRAZAS
TAKK
ありがとう
DANKE
Salamat
SUWUN
Благодарам
ASANTE
多謝
SALAMAT
TAKK
gracias
спасибо

References

References

Cook, R., Helland, I., & Su, Z. (2013). Envelopes and partial least squares regression. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *75*(5), 851–877.

Helland, I. S., Sæbø, S., & Tjelmeland. (2012). Near optimal prediction from relevant components. *Scandinavian Journal of Statistics*, *39*(4), 695–713.

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.