SIMREL–M: A SIMULATION TOOL FOR MULTI-RESPONSE LINEAR MODEL DATA

Raju Rimal, PhD Candidate (Supervisors: Solve Sæbø, Trygve Almøy)

Why simrel

- Simulate data from multi-response random regression model
- Easy design generation for computer experiments
- Latent relevant spaces for both X and Y
- Handy tool for model and method comparison
- Relevant and irrelevant predictors
- Complex data structures controlled by few parameters

Input Parameters

- Number of responses, predictors and samples
- Dimensions and variances of relevant X and Y spaces
- Relevance of latent X space to latent Y space (Coefficient of Determination)
- Unconditional and conditional correlation between responses
- Degree of multicollinearity in X
- Number of relevant predictors

Process flow

Inputs → Construct Σ, i.e., $cov(Y, X)$

Construct Rotation Matrix for latent spaces of X and Y

Generate Latent Components

Simulate Data by random rotations of components

SIMPLE TOOL FOR COMPLEX STRUCTURE

SIMREL 2.0

Responses — Predictors

- Relevant for $Y1$
- Relevant for $Y2$
- Relevant for both $Y1$ and $Y2$
- Irrelevant
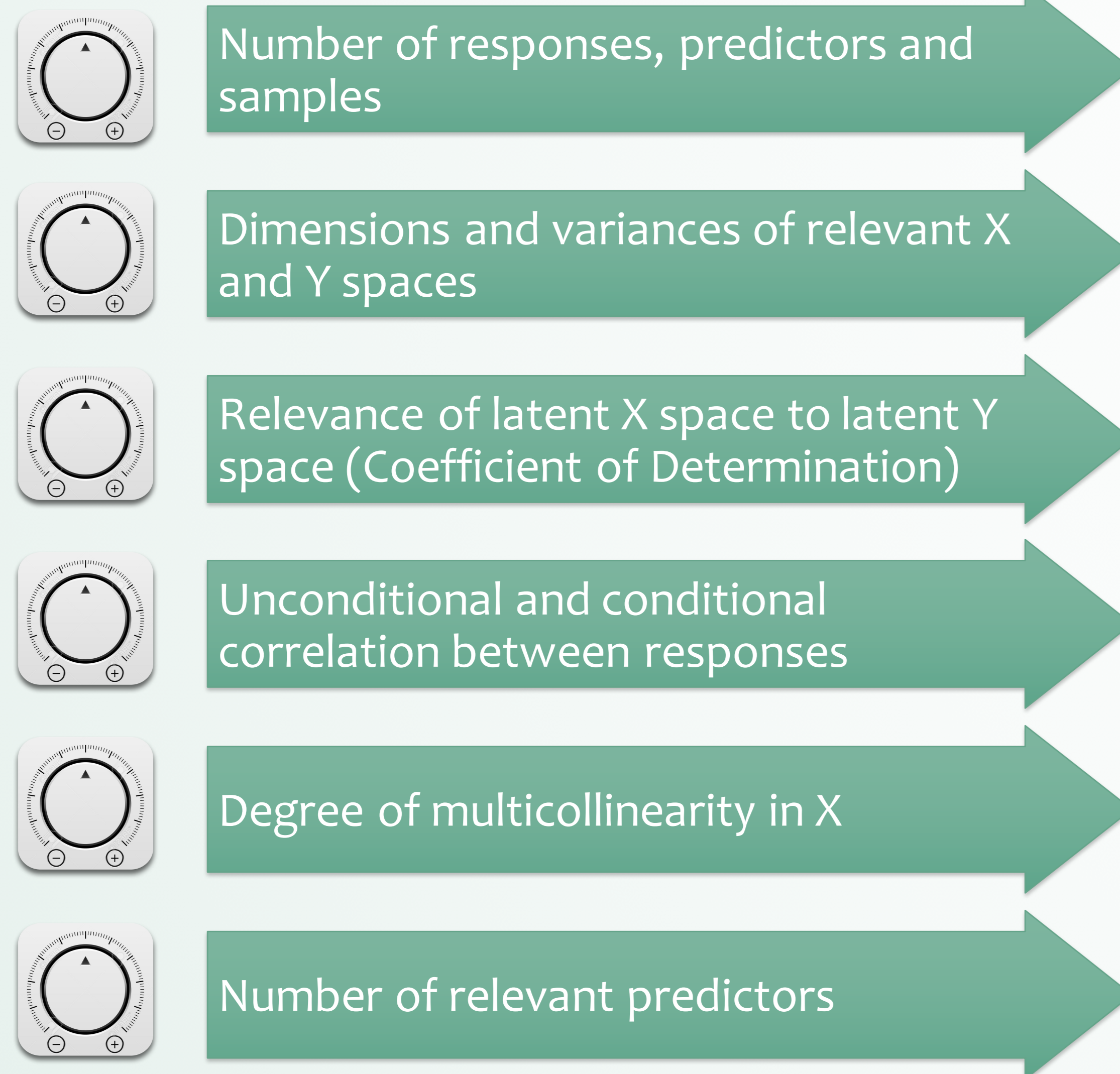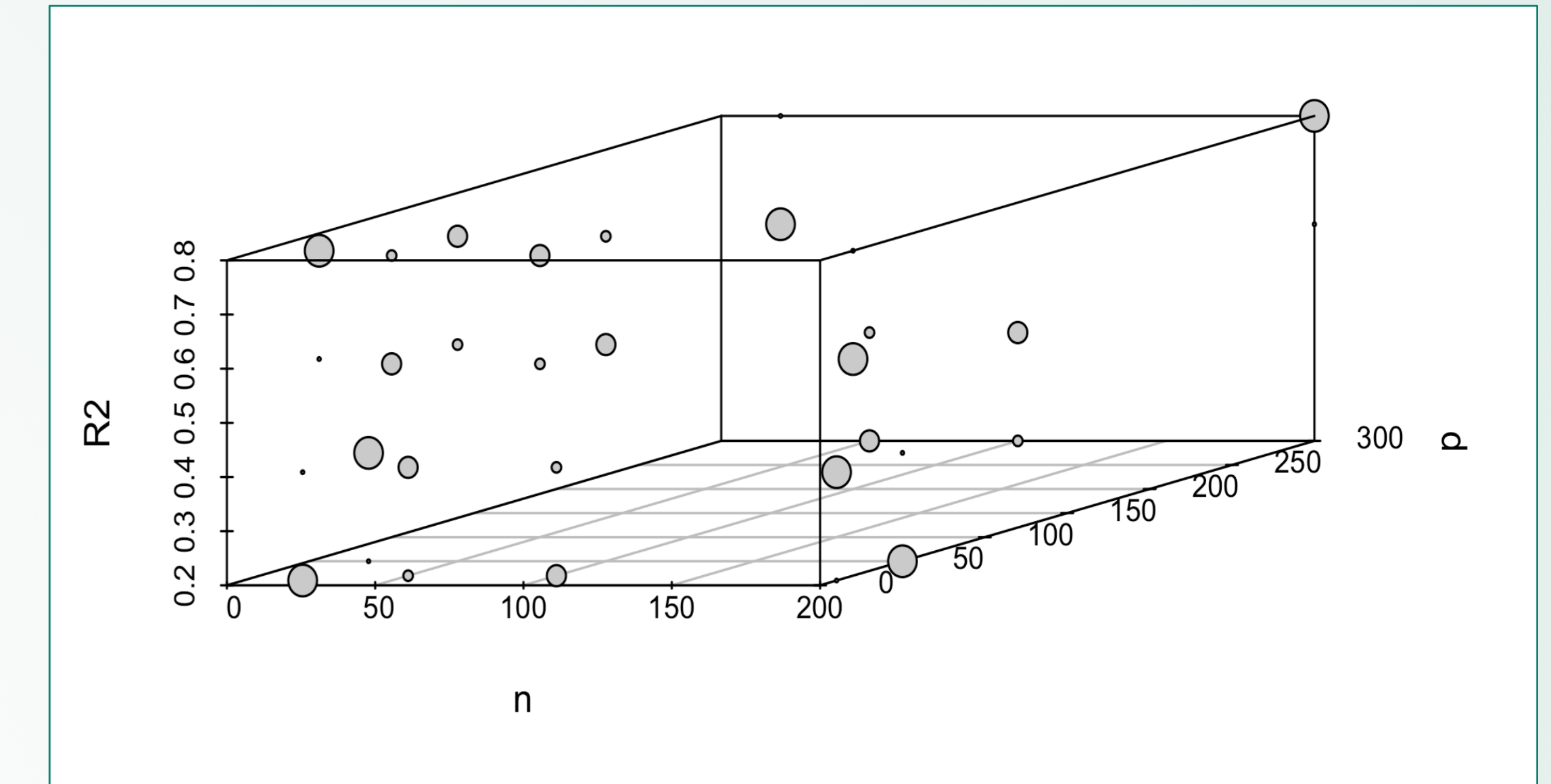
Examples

Fig1: Multilevel fractional Simulation Design created using simrel (size corresponds to gamma – multicollinearity factor)
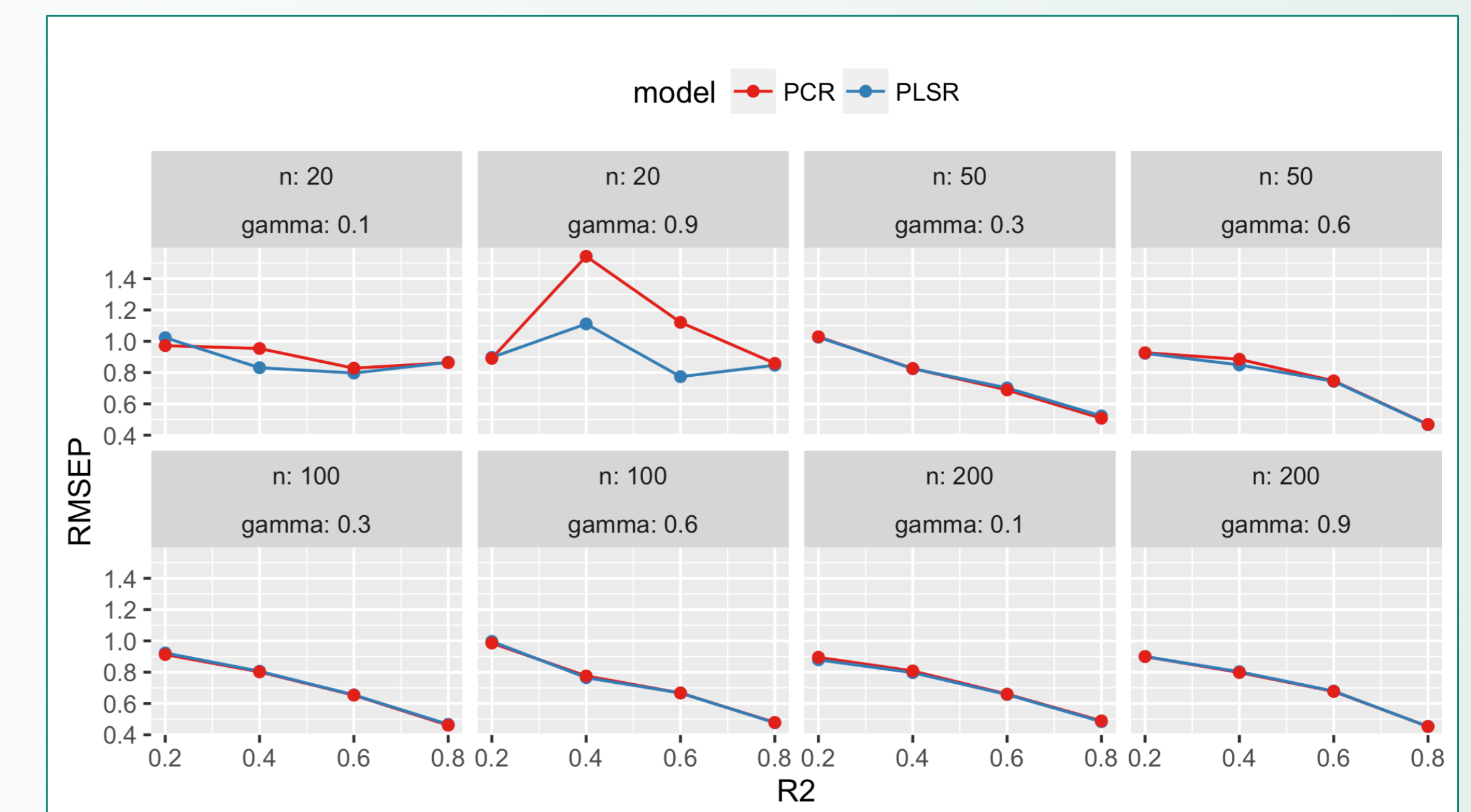

Fig2: Comparing PLS and PCR model on the basis of test prediction error for different combination of parameters
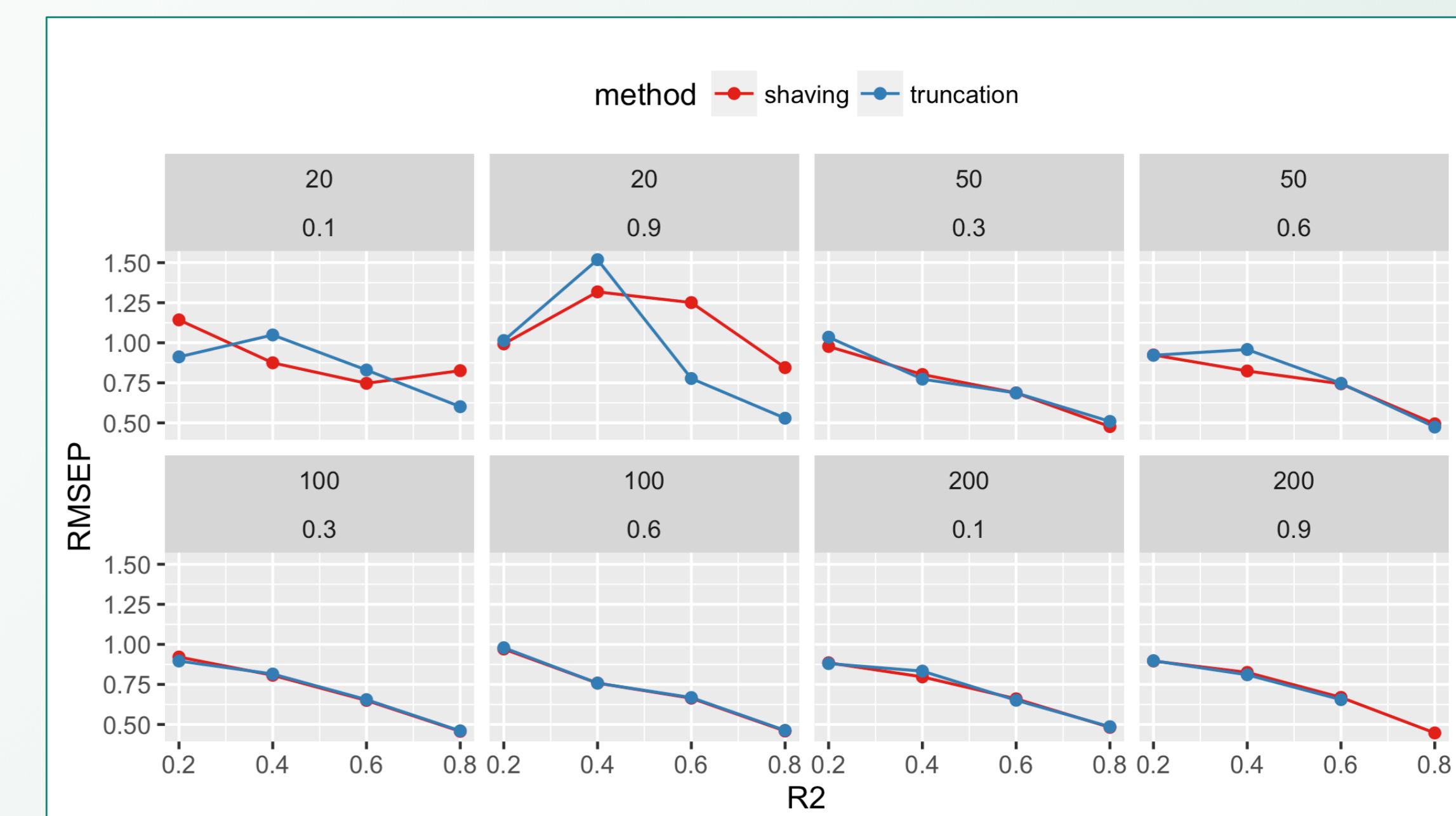

Fig2: Comparing Variable selection methods (shaving vs truncation) on the basis of test prediction error for different combination of parameters

References

- 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.
- Helland, I. S., & Almøy, T. (1994). Comparison of prediction methods when only a few components are relevant. *Journal of the American Statistical Association, 89*(426), 583-591.