Supplemental material for "Compositional covariance shrinkage and regularised partial correlations"

Suzanne Jin[1,2], Cédric Notredame[1,2] and Ionas Erb[1,3]

December 2023

The material contained herein is supplementary to the article named in the title and published in SORT-Statistics and Operations Research Transactions Volume 47(2).

[1] Centre for Genomic Regulation (CRG), Dr Aiguader, 88, 08003 Barcelona, Spain.

[2] Universitat Pompeu Fabra, Barcelona, Spain.

[3] Corresponding author: ionas.erb@crg.eu

S.1. Supplementary figures

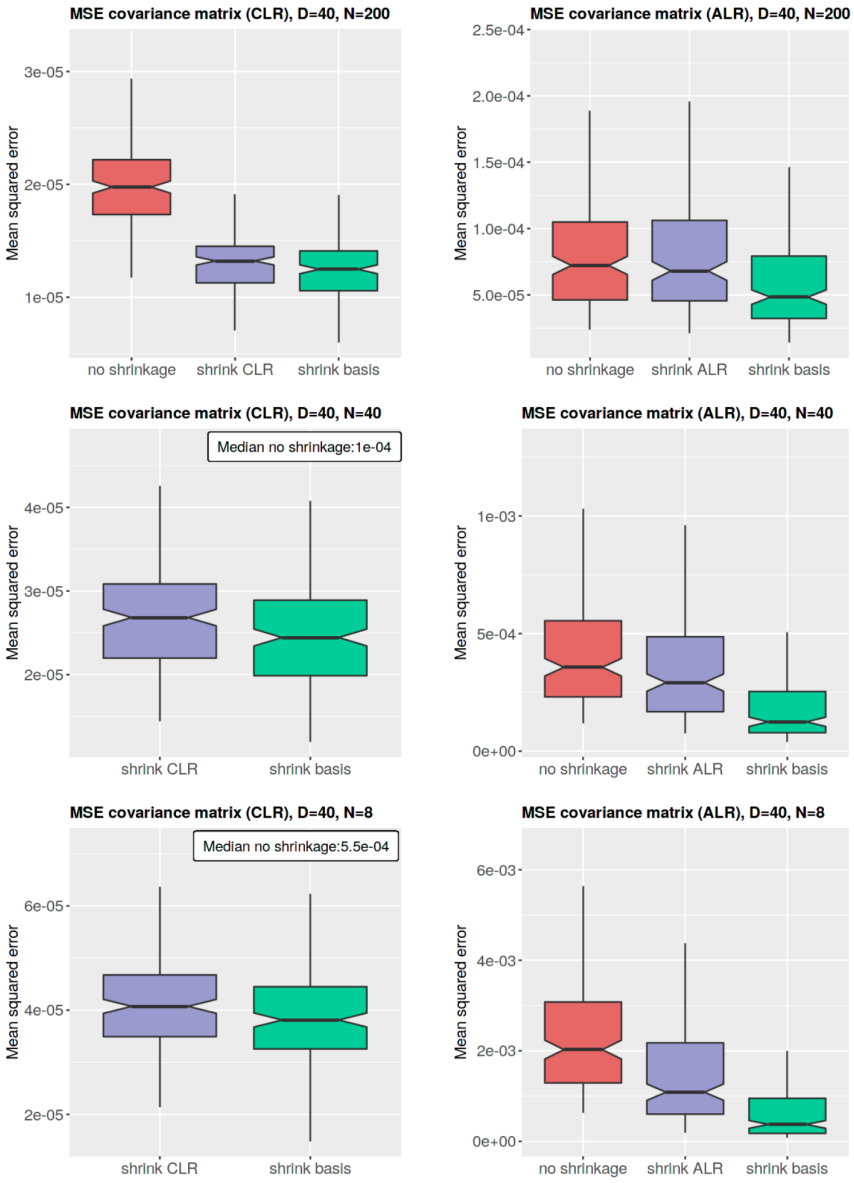

Figure S.1. *Mean squared error of CLR and ALR covariance matrices for different sample sizes (N=200, 40, 8) and estimation procedures (no shrinkage, naive shrinkage of CLR/ALR covariance matrix, and basis covariance shrinkage) computed on synthetic data. Each boxplot contains the results of 200 simulations. Whenever estimates without shrinkage are not shown, their median value is given in the legend.*

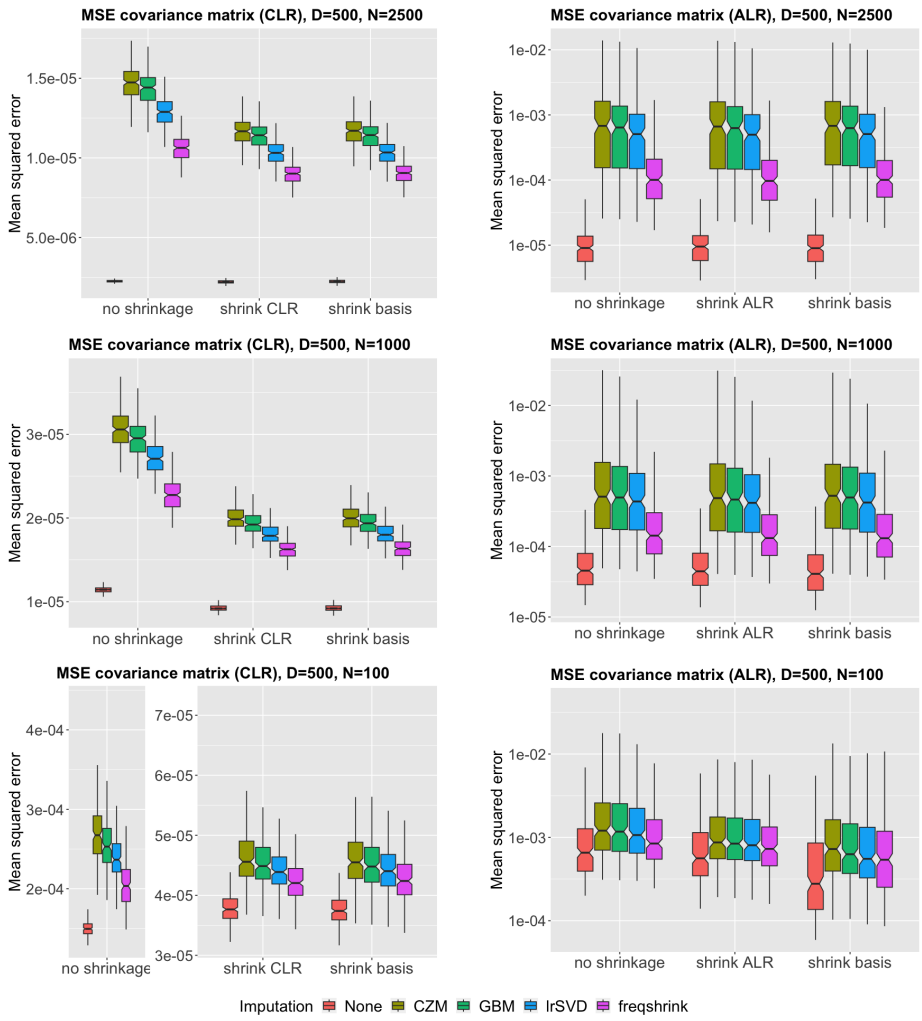

Figure S.2. *Mean squared error of CLR and ALR covariance matrices for different sample sizes (N=2500, 1000, 100) and estimation procedures (no shrinkage, naive shrinkage of CLR/ALR covariance matrix, and basis covariance shrinkage) computed on single-cell gene expression data. Each boxplot contains the results of 200 resamplings from data. Colours indicate the type of zero imputation used.*

S.2. Code and data availability

All the code used to perform the benchmark and reproduce the results of this paper, as well as the subset of single-cell gene expression data, are available on GitHub under https://github.com/suzannejin/pcor.bshrink.git.

The R package propr (Quinn et al., 2017) enables compositional data analysis on relative gene expression data. Originally designed for efficient calculations of proportionality indices, it has been updated several times, e.g., to include differential proportionality

across groups (Erb et al., 2017). We have now implemented partial correlations with basis shrinkage, which can be called using the following command line: propr(count, metric=”pcor.bshrink”, ivar=”clr”,...).