

Our simulation study shows that:

- The number of selected components by MICE with Q²-LOO criterion is much closer to the true number of components when the proportion of missing data is small (< 30%) and n > p, followed by NIPALS-PLSR, KNNimpute and SVDimpute for both the MCAR and the MAR assumptions.
- The number of selected components using the Q²-LOO shows a more consistent decreasing pattern of the true number of components for an increasing proportion of missing data and a decreasing sample size.
- The number of selected components by AIC, AIC-DoF and BIC are almost twice as large as the true number of components.
- The behaviors of BIC-DoF to selecting the true number of components criterion is not consistent. Their performances increase and then decrease with an increasing proportion of missing data.
- The MICE execution took a long time. For example when n= 100, the proportion of missing data= 10% and under MCAR assumption, the running time of MICE was about 11 times slower than NIPALS-PLSR
- Whatever the criterion used, the missingness mechanism is also to be considered since it influences the number of selected components.
- The true number of components of a PLS regression is difficult to determine, especially for small sample size and when the proportion of missing data is larger than 30%

REFERENCES

- Bertrand, F., Meyer, N., and Maumy-Bertrand, M. (2015). Package 'pisRgIm': Partial Least Squares Regression for Generalized Linear Models. URL https://cran.r-1. project.org/web/packages/ plsRqim/index.html. Krämer, N., and Sugiyama, N. (2011). The degrees of freedom of partial least squares regression, Journal of the American Statistical Association, 106(494): 697-2.
- 705 Little, R. J. A., and Rubin, D. B. (2002). Statistical analysis with missing data (2nd ed.). Wiley, New York
- 3. 4.
- Perry, P. O. (2015). Package 'bcv': Cross-Validation for the SVD (Bi-Cross-Validation). URL https://cran.r- project.org/web/packages/bcv/index.html. Templ, M., Alfons, A., Kowarik, A., and Prantner, B. (2017). Package 'VIM': Visualization and Imputation of Missing Values. URL https://github.com/statistikat/VIM. 5.
- 6. 7. Van Buuren, S., and Groothuis-Oudshoorn, K. (2011). mice: Multivariate Imputation by Chained Equations in R, Journal of Statistical Software, 45(3): 1-67. White, I. R., Royston, P., and Wood, A. M. (2011). Multiple imputation using chained equations: Issues and guidance for practice, Stat. Med., vol. 30, no. 4, pp
- 377_390

