Robust Weighted Partial Least Squares Regression

A. Alin[1*] and C. Agostinelli[2]

[1] *Department of Statistics, Dokuz Eylul University, Turkey ; aylin.alin@deu.edu.tr*
[2] *Dipartimento di Matematica, Università di Trento, Italy; claudio.agostinelli@unitn.it*
[*] *Presenting author*

Keywords. *Outliers; SIMPLS; Weighted Likelihood*

As technology develops, it gets much easier to reach and collect data. In (multivariate) linear regression model setting, large data sets may be in the form of $n \gg k$ or $n \ll k$ where n is the sample size and k is the number of independent variables. One common problem in these two scenarios is the multicollinearity causing large standard errors for the least squares parameter estimates. The result is the unreliable model as well as harmed hypothesis testing and estimation. Partial least squares regression is a well known multivariate technique for modelling data sets with multicollinearity problem. The idea is to solve multicollinearity and build regression model simultaneously. It is based on calculating components for both explanatory variables and response variables which will have maximum covariance. These components are calculated using well known algorithms. In this study we focus on the SIMPLS algorithm [DeJong, 1993] and, despite its popularity, it is vulnerable against the outlying observations. This vulnerability is the result of the least squares estimation method or using nonrobust covariance matrix to obtain the components. In the literature, there are proposals that use robust estimators for both covariance matrix and parameter estimates and hence they are more resistant to outlying observations. Here, we propose a robust weighted SIMPLS algorithm which is based on iteratively reweighting approach with robust weights calculated using weighted likelihood methodology Markatou et al. [1998]. Another improvement with respect to the original SIMPLS algorithm is in the calculation of loadings. The performance of the proposed method, the ordinary SIMPLS and Partial robust M-regression [Serneels et al., 2005] methods are compared with an extensive simulation study.

References

De Jong, S. (1993). SIMPLS:an alternative approach to partial least squares regression. *Chemometrics and Intelligent Laboratory Systems*, **18**, 251–263.

Markatou, M., Basu, A. and Lindsay, B.G. (1998). Weighted likelihood equations with bootstrap root search. *Journal of the American Statistical Association*, **93**, 740–750.

Serneels, S., Croux, C., Filzmoser, P. & Van Espen P.J. (2005). Partial robust M-regression. *Chemometrics and Intelligent Laboratory Systems*, **79**, 55–64.