

Outlier Detection and Robust Estimation in Linear Regression Models with Fixed Group Effects

Pérez B.¹, Molina I.², and Peña D.²

¹Department of Quantitative Methods, University of Pannonia, Hungary

²Department of Statistics, Universidad Carlos III de Madrid, Spain

Abstract

This work studies outlier detection and robust estimation with data that are naturally distributed into groups and which follow approximately a linear regression model with fixed group effects. For this, several methods are considered. First, the robust fitting method of Peña and Yohai [3], called principal sensitivity components method, is adapted to the grouped data structure and the mentioned model. The robust methods RDL_1 of Hubert and Rousseeuw [1] and M-S of Maronna and Yohai [2] are also considered. These three methods are compared in terms of their effectiveness in outlier detection and their robustness through simulations, considering several contamination scenarios and growing contamination levels. Results indicate that the adapted principal sensitivity components procedure is able to detect a high percentage of true outliers and a small number of false outliers. It is appropriate when the contamination is in the error term or in the covariates, detecting also possibly masked high leverage points. Moreover, in simulations the final robust regression estimator preserved good efficiency under Normality while keeping good robustness properties.

keywords: Linear models with fixed effects, Masking effect, Outlier detection, Principal sensitivity components, Robust estimation

References

- [1] Hubert M, Rousseeuw PJ (1997) Robust regression with both continuous and binary regressors. *Journal of Statistical Planning and Inference* 57: 153–163
- [2] Maronna RA, Yohai VJ (2000) Robust regression with both continuous and categorical predictors. *Journal of Statistical Planning and Inference* 89: 197–214
- [3] Peña D, Yohai VJ (1999) A fast procedure for outlier diagnostics in large regression problems. *Journal of the American Statistical Association* 94: 434–445