

Comparison of estimation methods for cellwise robust regression

I. Wilms^{1*} and C. Croux¹

¹ *Faculty of Economics and Business; ines.wilms@kuleuven.be, christophe.croux@kuleuven.be.*

**Presenting author*

Keywords. *Cellwise outliers; componentwise contamination; multiple regression.*

In multiple regression analysis, a response variable is predicted based on a set of p predictor variables. We collect all available information in an n by $(p + 1)$ matrix where the n observations are contained in the rows and the response and predictor variables are contained in the columns of the data set. Robust statistics has mostly focused on developing ‘rowwise robust’ estimation methods, methods that remain reliable in the presence of outlying rows in the data set. Such rowwise robust estimation methods flag either a whole row in the data set as outlying or not. Only recently, ‘cellwise robust’ estimation methods have been developed that flag a cell of the data set as outlying or not. Such cellwise robust estimation methods are likely to be more suited for situations where a large number of observations suffer from contamination in only a small number of variables. We perform an extensive simulation study where we compare the performance of several cellwise robust regression methods. We consider both low-dimensional (‘thin’) data sets, data sets with a large number of observations (rows) relative to the number of variables (columns) and high-dimensional (‘fat’) data sets, data sets with a large number of variables (columns) relative to the number of observations (rows).

References

- Leung, A., Zhang, H. & Zamar, R. (2016). Robust regression estimation and inference in the presence of cellwise and casewise contamination, *Computational Statistics & Data Analysis*, 99, 1-11.
- Öllerer, V., Alfons, A. & Croux, C. (2015). The shooting S -estimator for robust regression. *Computational Statistics*, DOI 10.1007/s00180-015-0593-7.