

Robust Estimator to Deal with Regression Models Having both Continuous and Categorical Regressors: A Simulation Study

¹Bashar A. Talib and ^{1,2}Habshah Midi

*^{1,2}Department of Mathematics, Faculty of Science,
Universiti Putra Malaysia,*

43400 UPM Serdang, Selangor, Malaysia

²Laboratory of Applied and Computational Statistics

*Institute for Mathematical Research (INSPEM), Universiti Putra Malaysia
43400 Serdang, Selangor, Malaysia*

E-mail: bashar@math.upm.edu.my, habshah@putra.upm.edu.my

ABSTRACT

The Ordinary Least Squares (OLS) method has been the most popular technique for estimating the parameters of the multiple linear regression. However, in the presence of outliers and when the model includes both continuous and categorical (factor) variables, the OLS can result in poor estimates. In this paper we try to introduce an alternative robust method for such a model that is much less influenced by the presence of outliers. A numerical example is presented to compare the performance of the OLS, the Re-weighted Least Squares based on the Robust Distance Least Absolute Value (RLSRDL₁), and the Re-weighted Least Squares based on the Robust Distance S/M estimator (RLSRDSM). The latter is the modification of the RDL₁. The empirical evidence shows that the performance of the RLSRDSM is fairly close to the RLSRDL₁ up to 20% outliers. As the percentage of outliers increases to more than 20%, the RLSRDSM is slightly better than the RLSRDL₁. However, the Robust Distance Least Absolute Value (RDL₁) estimator posed certain computational problems such as degenerate non-unique solutions while the RLSRDSM do not have such problem.

Keywords: Outliers, Leverage points, Robust Distance, S/M-estimates, RLSRDL₁, RLSRDSM