

The performance of robust-diagnostic F in the identification of multiple high leverage points

ABSTRACT

High leverage points have undue effects on the Least Square estimates. They are responsible for misleading conclusions in regression and multicollinearity problems. Hence, it is imperative to detect high leverage points and use robust estimators to estimate the parameters of a regression model, so as to arrive at valid conclusions. Several well-known methods have failed to detect multiple high leverage points correctly because of the swamping and/or masking effects. The Diagnostic Robust Generalized Potential (DRGP), is an appealing alternative method that successfully detects high leverage points correctly. However, for small percentages of high leverage points, it has the tendency to identify few low leverage points to be points of high leverage. In this paper, an attempt is made to correctly identify real high leverage point by reducing swamping effects. We propose a method we call Robust Diagnostic-F (RDF), in which robust approach is employed to detect the suspected high leverage points. Then, F statistics that relates the change in data covariance structure is used to confirm the suspicion. The performance of RDF is evaluated through real data and simulations. Comparisons are also made with existing methods.

Keyword: Robust; Diagnostic; Outliers; High leverage points; MM estimator