Detection of influential observations in spatial regression model based on outliers and bad leverage classification

Influential observations (IOs), which are outliers in the x direction, y direction or both, remain a problem in the classical regression model fitting. Spatial regression models have a peculiar kind of outliers because they are local in nature. Spatial regression models are also not free from the effect of influential observations. Researchers have adapted some classical regression techniques to spatial models and obtained satisfactory results. However, masking or/and swamping remains a stumbling block for such methods. In this article, we obtain a measure of spatial Studentized prediction residuals that incorporate spatial information on the dependent variable and the residuals. We propose a robust spatial diagnostic plot to classify observations into regular observations, vertical outliers, good and bad leverage points using a classification based on spatial Studentized prediction residuals and spatial diagnostic potentials, which we refer to as ISRs−Posi and ESRs−Posi. Observations that fall into the vertical outliers and bad leverage points categories are referred to as IOs. Representations of some classical regression measures of diagnostic in general spatial models are presented. The commonly used diagnostic measure in spatial diagnostics, the Cook's distance, is compared to some robust methods, H2i (using robust and non-robust measures), and our proposed ISRs−Posi and ESRs−Posi plots. Results of our simulation study and applications to real data showed that the Cook's distance, non-robust H2si1 and robust H2si2 were not very successful in detecting IOs. The H2si1 suffered from the masking effect, and the robust H2si2 suffered from swamping in general spatial models. Interestingly, the results showed that the proposed ESRs−Posi plot, followed by the ISRs−Posi plot, was very successful in classifying observations into the correct groups, hence correctly detecting the real IOs.

Keyword: Spatial regression model; Influential observation; Outlier; Leverage; Prediction residual; Masking and swamping; Diagnostic