 doi:10.20944/preprints202108.0178.v1

Article

Detection of Influential Observations in Spatial Regression Model Based on Outliers and Bad Leverage Classification

Ali Mohammed Baba [1,2], Habshah Midi [1,3,*], Mohd Bakri Adam [1,3] and Nur Haizum Abd Rahman [1,3]

[1] Institute for Mathematical Research, Universiti Putra Malaysia 43400, Selangor Malaysia; ambabastats@gmail.com
[2] Department of Mathematical Sciences, Abubakar Tafawa Balewa University Bauchi, 0248 Bauchi Nigeria; mbali@atbu.edu.ng
[3] Department of Mathematics, Faculty of Science Universiti Putra Malaysia 43400, Selangor Malaysia
* Correspondence: habshahmidi@gmail.com.

Abstract: Influential Observations, which are outliers in x direction, y direction or both, remain a hitch in classical regression model fitting. Spatial regression model, with peculiar nature of outliers due to their local nature, is not free from the effect of such influential observations. Researchers have adapted some classical regression techniques to the spatial models and yielded satisfactory results. However, masking or/and swamping remain stumbling block to such methods. We obtained the spatial representation of the classical regression measures of diagnostic in general spatial model. Commonly used diagnostic measure in spatial diagnostic, the Cook's distance, is compared to some robust methods, H_i^2 (using robust and non-robust measures), and classification based on generalized residuals and diagnostic generalized potentials, $ISRs - P_{osi}$ and $ESRs - P_{osi}$, with the help of the obtained spatial prediction residuals and the spatial leverage term. Results of simulation and applications to real data have shown the advantage of the $ISRs - P_{osi}$ and $ESRs - P_{osi}$ due to classification of outliers over Cook's distance and non-robust H_{si1}^2, which suffer from masking, and robust H_{si2}^2 which suffer from swamping in general spatial model.

Keywords: Spatial regression model; Influential observation; Outlier; Leverage; prediction residual; Masking and swamping; Diagnostic.

1. Introduction

Belsley et al. [1] defined an influential observation (IO) as one which, either individually or together with several other observations has demonstrably large impact on the calculated values of various estimates. An influential observation could be an outlier in the X-space (leverage points) or outlier in the Y- space (vertical outlier). Leverage points can be classified into good (GLP) and bad leverage points (BLP). Unlike BLPs, GLPs follow the pattern of the majority of the data; hence they are not considered as IOs as they have little or no influence on the calculated values of numerous estimates [2,3]. In this connection, Rashid et al. **[2]** stated that IOs could be vertical outliers (VO) or BLPs. Thus, it is very crucial to identify IOs as they are responsible for the misleading conclusion about the fitted regression model and various estimates. Once the IOs are identified, we need to study their impact on the model and subsequent analyses. There are handful literatures on diagnostic of IOs in linear regression; some examples are **[1,3–12]**. Other articles in the literature deal with regressions with correlated residuals, e.g. **[13–16]**. However, only scarce articles deal with the detection of IOs in spatial regression models, some examples which include **[17–20]**. Christensen et al. **[17]** and Haining **[18]** adapted one of the diagnostic measures in **[3]**, in detecting influential observations in spatial error autoregression model. They achieved this by defining the correlated errors through the spatial weight matrix and coefficient of spatial autocorrelation in the error term.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 9 August 2021 doi:10.20944/preprints202108.0178.v1

Nonetheless, diagnostic works on models that have both spatial autocorrelation in dependent variable and residual terms are missing in the literature. The problem of masking and swamping is prevalent in spatial regression model diagnostics. This may be due to the presence of vertical outliers as well as leverage points as in the case of linear regression. This motivated us to adapt and extend some robust diagnostic measures of detection of outliers and IOs in linear regression such as Hadi's potential ($\boldsymbol{p_{oii}}$) , Cook's distance ($\boldsymbol{CD_i}$) **[3]**, overall potential influence ($\boldsymbol{H_i^2}$)**[10]**, external (ESR) and internal (ISR) studentized residuals **[1,9,10]**, to spatial regression models in order to minimize the problem of masking and swamping in spatial models.

In this article, we adapt some diagnostic measures in the linear regression model and their representations in the spatial regression model are obtained, with special emphasis on general spatial regression model (GSM), that has auto -regression on both dependent variable and the error terms. We extend the results of linear regression of identification of outliers and influential observations to the GSM model.

The main objective of this study are: (1) to represent the leverage values of hat matrix of linear regression to GSM model; (2) to extend the ISR of linear regression to GSM model; (3) to extend the ESR of linear regression to GSM model; (4) to extend the Cook's distance and the overall potential influence of linear regression to GSM model (5) to develop a method of identification of influential observations of GSM model by proposing a procedure of classification of observations into regular observations, vertical outliers, good and bad leverage points; (6) to evaluate the performances of the proposed methods by using simulation studies; (7) to apply the proposed methods on gasoline price data for retail sites in Sheffield, UK and Covid-19 data at Georgia, USA, and the Life expectancy data in USA counties. The significance of this study is that it can contribute to the development of method of identification of influential observation in spatial regression model.

2. Identification of Influential Observations in Linear Regression Model

Consider a k- variable regression model:

$$Y = X\beta + \varepsilon \tag{1}$$

Where Y is an $n \times 1$ vector of observations of dependent variables, X is an $n \times k$ matrix of independent variables, β is a $k \times 1$ vector of unknown regression parameters, ε is an $n \times 1$ vector of random errors with identically normal distribution as $\varepsilon \sim NID(0, \sigma^2)$.

The ordinary least squares (OLS) estimates in Equation 1 are given by

$$\hat{\beta} = (X^T X)^{-1} X^T Y \tag{2}$$

The vector of predicted values can be written as

$$\widehat{\boldsymbol{Y}} = \boldsymbol{X}\widehat{\boldsymbol{\beta}} = \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Y} = \boldsymbol{P}\boldsymbol{Y},$$

where $P = X(X^T X)^{-1} X^T$ is the hat/leverage matrix. The diagonal elements of leverage matrix are called the hat values and denoted as p_{ii}, given by

$$p_{ii} = x_i^T (X^T X)^{-1} x_i, \qquad i = 1,2,\cdots,n.$$

The hat matrix is often used as diagnostics to identify leverage points. Leverage is the amount of influence exerted by the observed response $\boldsymbol{y_i}$ on the predicted variable $\widehat{\boldsymbol{y}}_{\boldsymbol{i}}$. As a result, large leverage value indicates that the observed response has large effect in the predicted response.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 9 August 2021 doi:10.20944/preprints202108.0178.v1

Hoaglin and Welsh [3] suggested that observations which exceed $\frac{2k}{n}$,where $\frac{2k}{n}$ is the average value of p_{ii}, is considered as the leverage points, while Vellman and Welsch suggested $\frac{3k}{n}$ as a cut-off points for leverage points. Huber **[7]** suggested that the ranges $p_{ii} \leq 0.2$, $0.2 < p_{ii} \leq 0.5$ and $p_{ii} > 0.5$ are safe, risky and to be avoided respectively, for leverage values.

Unfortunately, the hat matrix suffers from the masking effect. So, p_{ii} often fails to detect high leverage points. Hadi [10] suggested a single case deleted measure called potentials or Hadi's potentials. The diagonal elements of potential denoted as p_{oii}, is given by

$$p_{0ii} = x_i^T\left(X_{(i)}^T X_{(i)}\right)x_i, \quad i = 1,2,\cdots,n \tag{3}$$

Where $X_{(i)}$ is the matrix X with the i^{th} row deleted. We can rewrite p_{0ii} as a function of p_{ii} as

$$p_{0ii} = \frac{p_{ii}}{1-p_{ii}}, \qquad 1,2,\cdots,n.$$

The vector of the residuals, $\boldsymbol{r}$, can be written as

$$\boldsymbol{r} = \boldsymbol{Y} - \hat{\boldsymbol{Y}} = (\boldsymbol{I} - \boldsymbol{P})\boldsymbol{Y} = \boldsymbol{QY},$$

The Studentized residuals (internally Studentized residuals) denoted as *ISR* and R-Student residual (externally Studentized residuals) denoted as *ESR* are widely used measures for the identification of outliers (see [7]). The *ISR*, denoted as $\boldsymbol{t_i}$, is defined as

$$\boldsymbol{t_i} = \frac{\boldsymbol{r_i}}{\hat{\boldsymbol{\sigma}}\sqrt{\boldsymbol{1} - \boldsymbol{p_{ii}}}}$$

where $\hat{\boldsymbol{\sigma}}$ is the standard deviation of the residuals, $\boldsymbol{r_i}$ and $\boldsymbol{p_{ii}}$ are the $\boldsymbol{i^{th}}$ residual and diagonal element of the matrix $\boldsymbol{P}$, respectively (see **[9]** for details). Meanwhile, Chatterjee and Hadi **[9]** defined *ESR* denoted as $\boldsymbol{t_i^*}$, given by

$$\boldsymbol{t_i^*} = \frac{\boldsymbol{r_i}}{\hat{\boldsymbol{\sigma}}_{(i)}\sqrt{\boldsymbol{1} - \boldsymbol{p_{ii}}}}$$

where $\hat{\sigma}_{(i)}$ is the residuals mean square excluding the i^{th} case. The *ESR* follows a Student's t-distribution with $(n - k - 1)$ degrees of freedom [9].

One of the most employed measures of influence in linear regression is the Cook's distance **[3]**. It measures influence on the regression coefficient estimate, or the predicted values. The Cook's distance is given by

$$\widehat{\boldsymbol{CD}}_{\boldsymbol{i}}(\boldsymbol{X^TX}, \boldsymbol{k}\hat{\boldsymbol{\sigma}}^{\boldsymbol{2}}) = \frac{\left(\hat{\boldsymbol{\beta}}^{(-i)} - \hat{\boldsymbol{\beta}}\right)^{\boldsymbol{T}}(\boldsymbol{X^TX})\left(\hat{\boldsymbol{\beta}}^{(-i)} - \hat{\boldsymbol{\beta}}\right)}{\boldsymbol{k}\hat{\boldsymbol{\sigma}}^{\boldsymbol{2}}}, \tag{4}$$

where $\hat{\boldsymbol{\beta}}$ is the vector of estimates of $\boldsymbol{\beta}$ using the full data, $\hat{\boldsymbol{\beta}}^{(-i)}$ is the vector of estimates of $\boldsymbol{\beta}$ with the $\boldsymbol{i^{th}}$ observation of $\boldsymbol{Y_i}$ and $\boldsymbol{X_i}$ omitted, k is the number of parameters and $\hat{\boldsymbol{\sigma}}^{\boldsymbol{2}}$ is the estimate of variance. Any $\boldsymbol{i^{th}}$ observation is declared influential observation (IO) if $\widehat{\boldsymbol{CD}}_{\boldsymbol{i}} > \boldsymbol{F}[\boldsymbol{0.5}; \boldsymbol{k}, (\boldsymbol{n} - \boldsymbol{k})]$. Meloun **[12]** noted that any observations in which $\boldsymbol{CD_i} > \boldsymbol{1}$ is considered as influential observations. The Cook's distance can also be written as **[8,9]**

 doi:10.20944/preprints202108.0178.v1

$$\widehat{CD}_i(X^TX, k\hat{\sigma}^2) = \frac{(\hat{Y} - \hat{Y}_i)^T(\hat{Y} - \hat{Y}_i)}{k\hat{\sigma}^2} \quad (5)$$

.

Computing the $\widehat{CD}_i(X^TX, k\hat{\sigma}^2)$ does not require fitting a regression equation for each of the i^{th} observations and the full model, instead Equation 3 can further be simplified as ([3,8,9])

$$\widehat{CD}_i(X^TX, k\hat{\sigma}^2) = \frac{1}{k} t_i^2 \frac{p_{ii}}{q_{ii}} \quad (6)$$

where $t_i = \frac{e_i}{\hat{\sigma}\sqrt{q_{ii}}}$ is the *ISR* and $\frac{p_{ii}}{q_{ii}}$ $(q_{ii} = 1 - p_{ii})$ is referred to as potential [7–9]. Interestingly, the Cook's distance is a measure of influence based on the potential ($\frac{p_{ii}}{q_{ii}}$) and studentized residual (t_i).

Hadi **[10]** demonstrated the drawbacks of methods that are multiplicative of functions, such as the Cook's distance **[3]**, Andrews-Pregibon statistic **[5]**, Cook and Weisberg statistic **[8]**, etc.. (see **[10]** for details) and proposed a method that is additive of the functions. Though both the multiplicative and additive are functions of residuals and leverage values, the former diminishes towards zero for smaller value of any of the two functions or both, while in the latter case, the measure is large if one of the two functions or both are large. He proposed a measure of overall potential influence, denoted as H_i^2, and defined as follows

$$H_i^2 = \frac{k}{m} \frac{e_I^T(I_m - P_I)^{-1}e_I}{e^Te - e_I^Te_I} + \frac{1}{m} tr(P_I(I_m - P_I)^{-1}), \quad (7)$$

with *k*, the number of the parameters in the model, $I = \{i_1, i_2, \cdots, i_m\}$ is set of indices of observations of length *m*, and P_I is the leverage indexed by I.

For $m = 1$ and $I = i$, Equation 7 simplifies to

$$H_i^2 = \frac{k}{(1 - p_{ii})} \frac{e_i^2}{(e^Te - e_i^2)} + \frac{p_{ii}}{1 - p_{ii}} = \frac{k}{(1 - p_{ii})} \frac{d_i^2}{(1 - d_i^2)} + \frac{p_{ii}}{1 - p_{ii}}, \quad (8)$$

where $\sum p_{ii} = k$, $\sum d_i^2 = 1$, $d_i^2 = \frac{e_i^2}{e^Te}$ is the square of the i^{th} normalized residual.

Hadi **[10]** suggested a cutoff point for Hadi's potential (p_{oii}) and , H_i^2 denoted as (l_1) which is given as follows,

$$l_1 = mean(po_{ii}) + c\sqrt{Var(po_{ii})}$$

$$= \frac{k}{n} + c\sqrt{\frac{ns - k^2}{n(n-1)}},$$

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 9 August 2021 doi:10.20944/preprints202108.0178.v1

where, $\boldsymbol{c} = \mathbf{2}, \mathbf{3}, , \boldsymbol{s} = \sum \boldsymbol{p_{ii}}$. Since both the mean and the standard deviation are easily affected by outliers, suggested to employ such confidence bound type of cut-off points by replacing the mean and the standard deviation by robust estimators, namely median and normalized median absolute deviation, respectively. The resulting cut-off points is denoted as $\boldsymbol{l_2}$;

$$\boldsymbol{l_2} = \boldsymbol{Med}(\boldsymbol{po_{ii}}) + \boldsymbol{cMAD}(\boldsymbol{po_{ii}}),$$

3. Influential Observations in Spatial Regression Models

The general spatial autogressive model (GSM) includes the spatial lag term and spatially correlated error structure. The data generating process (DGP) of the general spatial model is given by:

$$\boldsymbol{y} = \boldsymbol{\rho W_1 y} + \boldsymbol{X\beta} + \boldsymbol{\xi}, \ \boldsymbol{\xi} = \boldsymbol{\lambda W_2 \xi} + \boldsymbol{\varepsilon}, \ \boldsymbol{\varepsilon} \sim \boldsymbol{N}(\mathbf{0}, \boldsymbol{\sigma^2 I_n}), \tag{9}$$

where y is an $\boldsymbol{n} \times \mathbf{1}$ vector of dependent variable. X is an $\boldsymbol{n} \times \boldsymbol{k}$ matrix of explanatory variables. $\boldsymbol{W_1}$ and $\boldsymbol{W_2}$ are $\boldsymbol{n} \times \boldsymbol{n}$ spatial weight matrices. $\boldsymbol{I_n}$ is an $\boldsymbol{n} \times \boldsymbol{n}$ identity matrix. $\boldsymbol{\xi}$ is the spatially correlated error term, $\boldsymbol{\varepsilon}$ is the random residual term. The parameter $\boldsymbol{\rho}$ is the coefficient of the spatially lagged dependent variable $\boldsymbol{W_1 y}$, and $\boldsymbol{\lambda}$ is the coefficient of the spatially correlated errors.

The general spatial autoregressive model in Equation 9 can be re-written as

$$\boldsymbol{Ay} = \boldsymbol{X\beta} + \boldsymbol{B^{-1}\varepsilon}, \tag{10}$$

where,

$\boldsymbol{A} = \boldsymbol{I_n} - \boldsymbol{\rho W_1}$, $\boldsymbol{\xi} = \boldsymbol{B^{-1}\varepsilon}$, $\boldsymbol{B} = \boldsymbol{I_n} - \boldsymbol{\lambda W_2}$, $\boldsymbol{\xi} \sim \boldsymbol{N}(\mathbf{0}, \boldsymbol{\sigma^2 V})$, and $\boldsymbol{V} = (\boldsymbol{B^T B})^{-\mathbf{1}}$. Estimation of the parameters is achieved using the maximum likelihood estimation methods.

The log-likelihood function $(\boldsymbol{L})$ is given by

$$\boldsymbol{L} = -\frac{\boldsymbol{n}}{\mathbf{2}}\boldsymbol{ln}(\boldsymbol{\sigma^2}) + \boldsymbol{ln}|\boldsymbol{A}| + \boldsymbol{ln}|\boldsymbol{B}| - \frac{\mathbf{1}}{\mathbf{2\sigma^2}}(\boldsymbol{Ay} - \boldsymbol{X\beta})^{\boldsymbol{T}}\boldsymbol{B^T B}(\boldsymbol{Ay} - \boldsymbol{X\beta}) \tag{11}$$

Let $\hat{\boldsymbol{\rho}}, \hat{\boldsymbol{\lambda}}, \hat{\boldsymbol{\sigma}}^{\mathbf{2}}, \hat{\boldsymbol{\beta}}$ be the maximum likelihood estimates (MLE) of $\boldsymbol{\rho}, \boldsymbol{\lambda}, \boldsymbol{\sigma^2}, \boldsymbol{\beta}$, respectively. The MLEs are obtained iteratively using numerical methods in the maximum likelihood estimation. Anselin **[22]** and LeSage **[23]** have discussed the maximum likelihood estimation procedure of the parameters.

3.1. *Leverage in spatial regression model*

Denote the vector of parameters in Equation 11 as $\boldsymbol{\beta_{ay}}$. The estimate of $\boldsymbol{\beta_{ay}}$, $\hat{\boldsymbol{\beta}}_{\boldsymbol{ay}}$ is given by

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 9 August 2021 doi:10.20944/preprints202108.0178.v1

$$\widehat{\boldsymbol{\beta}}_{\boldsymbol{ay}} = \left(\boldsymbol{X}^{\boldsymbol{T}}\widehat{\boldsymbol{V}}^{\boldsymbol{-1}}\boldsymbol{X}\right)^{\boldsymbol{-1}}\boldsymbol{X}^{\boldsymbol{T}}\widehat{\boldsymbol{V}}^{\boldsymbol{-1}}\widehat{\boldsymbol{A}}\boldsymbol{y}.$$

The model (11) is viewed as fitting a general linear model, Ay on X, that has correlated residual terms. Set $\boldsymbol{z} = \boldsymbol{Ay}$, where $\boldsymbol{var}(\boldsymbol{Ay}) = \boldsymbol{\sigma^2 V}$. Therefore,

$$\hat{\boldsymbol{z}} = \boldsymbol{X}\widehat{\boldsymbol{\beta}}_{\boldsymbol{ay}} = \boldsymbol{X}\left(\boldsymbol{X}^{\boldsymbol{T}}\widehat{\boldsymbol{V}}^{\boldsymbol{-1}}\boldsymbol{X}\right)^{\boldsymbol{-1}}\boldsymbol{X}^{\boldsymbol{T}}\widehat{\boldsymbol{V}}^{\boldsymbol{-1}}\boldsymbol{z} = \boldsymbol{P}_{\boldsymbol{ay}}\boldsymbol{z}$$

The hat matrix, in this case, is given by $\boldsymbol{P}_{\boldsymbol{ay}}$,

$$\boldsymbol{P}_{\boldsymbol{ay}} = \boldsymbol{X}\left(\boldsymbol{X}^{\boldsymbol{T}}\widehat{\boldsymbol{V}}^{\boldsymbol{-1}}\boldsymbol{X}\right)^{\boldsymbol{-1}}\boldsymbol{X}^{\boldsymbol{T}}\widehat{\boldsymbol{V}}^{\boldsymbol{-1}}.$$

Let $\boldsymbol{Q}_{\boldsymbol{ay}} = \boldsymbol{I}_{\boldsymbol{n}} - \boldsymbol{P}_{\boldsymbol{ay}}$. Though $\boldsymbol{P}_{\boldsymbol{ay}}$ and $\boldsymbol{Q}_{\boldsymbol{ay}}$ have satisfied the idempotence property and sum of diagonal elements equal *k* and $\boldsymbol{n} - \boldsymbol{k}$, respectively, both are not symmetric. As a result, they are not positive semi-definite, and as such, the diagonal elements of $\boldsymbol{P}_{\boldsymbol{ay}}$ would have negative values. The hat matrices $\boldsymbol{P}_{\boldsymbol{ay}}$ and $\boldsymbol{Q}_{\boldsymbol{ay}}$ are not symmetric and their diagonal values do not lie between 0 and 1 (inclusive).

Martins **[15]** proposed a measure of leverage that is orthogonal in the models with correlated residuals, whose diagonal values lie in the interval [0, 1], which we denote by $\boldsymbol{P}^{*}_{\boldsymbol{ay}}$, such that:

$$\boldsymbol{P}^{*}_{\boldsymbol{ay}} = \widehat{\boldsymbol{V}}^{\boldsymbol{-1}}\boldsymbol{P}_{\boldsymbol{ay}} = \widehat{\boldsymbol{V}}^{\boldsymbol{-1}}\boldsymbol{X}\left(\boldsymbol{X}^{\boldsymbol{T}}\widehat{\boldsymbol{V}}^{\boldsymbol{-1}}\boldsymbol{X}\right)^{\boldsymbol{-1}}\boldsymbol{X}^{\boldsymbol{T}}\widehat{\boldsymbol{V}}^{\boldsymbol{-1}}$$

Let $\boldsymbol{Q}^{*}_{\boldsymbol{ay}} = \boldsymbol{I}_{\boldsymbol{n}} - \boldsymbol{P}^{*}_{\boldsymbol{ay}}$. $\boldsymbol{P}^{*}_{\boldsymbol{ay}}$ and $\boldsymbol{Q}^{*}_{\boldsymbol{ay}}$ are idempotent, symmetric and orthogonal with respect to $\boldsymbol{V}$. i.e.

1. $\boldsymbol{P}^{*}_{\boldsymbol{ay}}\widehat{\boldsymbol{V}}\boldsymbol{P}^{*}_{\boldsymbol{ay}} = \boldsymbol{P}^{*}_{\boldsymbol{ay}}$
2. $\boldsymbol{Q}^{*}_{\boldsymbol{ay}}\widehat{\boldsymbol{V}}\boldsymbol{Q}^{*}_{\boldsymbol{ay}} = \boldsymbol{Q}^{*}_{\boldsymbol{ay}}$
3. $\boldsymbol{P}^{*}_{\boldsymbol{ay}}\widehat{\boldsymbol{V}}\boldsymbol{Q}^{*}_{\boldsymbol{ay}} = \boldsymbol{0}$

Note that the sum of diagonal elements of $\boldsymbol{P}^{*}_{\boldsymbol{ay}}$ and $\boldsymbol{Q}^{*}_{\boldsymbol{ay}}$, the leverage, do not sum to $\boldsymbol{k}$ and $\boldsymbol{n} - \boldsymbol{k}$.

Again, consider a new set of dependent variables obtained by pre-multiplying Equation 11 by the matrix $\boldsymbol{B}$ ($\boldsymbol{B}$ as defined in Equation 10) so that $\boldsymbol{z}^{*} = \boldsymbol{BAy}$. Schall and Dunne **[14]** defined the matrix $\boldsymbol{V}^{-1}$ as a singular value decomposition such that $\boldsymbol{V}^{-1} = \boldsymbol{B\Delta B}^{\boldsymbol{T}}$; where $\boldsymbol{B}$ is of the same order as $\boldsymbol{V}^{-1}$ and $\boldsymbol{\Delta}$ is diagonal matrix. The transformation $\boldsymbol{z}^{*}$ are the principal component scores. Puterman **[13]** and Haining **[18]** defined it as canonical variates such that $\boldsymbol{BX}(\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{V}^{-1}\boldsymbol{X})\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{B}^{\boldsymbol{T}}$ is positive semi definite. By setting $\boldsymbol{z}^{*} = \boldsymbol{BAy}$, Equation 9 is re-written in Generalized Least squared (GLS) form as

$$\boldsymbol{z}^{*} = \boldsymbol{X}^{*}\boldsymbol{\beta}_{\boldsymbol{s}} + \boldsymbol{\varepsilon}, \boldsymbol{\varepsilon} \sim \boldsymbol{N}(\boldsymbol{0}, \boldsymbol{\sigma^2 I}_{\boldsymbol{n}}) \quad (12)$$

where $\boldsymbol{X}^{*} = \boldsymbol{BX}$.

The estimate $\widehat{\boldsymbol{\beta}}_{\boldsymbol{s}}$ of $\boldsymbol{\beta}_{\boldsymbol{s}}$ is now given by

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 9 August 2021 doi:10.20944/preprints202108.0178.v1

$$\hat{\boldsymbol{\beta}}_{\boldsymbol{s}} = \left(\boldsymbol{X}^{*\boldsymbol{T}}\boldsymbol{X}^{*}\right)^{-1}\boldsymbol{X}^{*\boldsymbol{T}}\boldsymbol{z}^{*}$$

Thus,

$$\hat{\boldsymbol{z}}^{*} = \boldsymbol{X}^{*}\left(\boldsymbol{X}^{*\boldsymbol{T}}\boldsymbol{X}^{*}\right)^{-1}\boldsymbol{X}^{*\boldsymbol{T}}\boldsymbol{z}^{*} \tag{13}$$

where, $\hat{\boldsymbol{A}} = \boldsymbol{I_n} - \hat{\boldsymbol{\rho}}\boldsymbol{W_1}$ and $\hat{\boldsymbol{B}} = \boldsymbol{I_n} - \hat{\boldsymbol{\lambda}}\boldsymbol{W_2}$. Note that $\hat{\boldsymbol{y}}$ is deduced from Equation 13 as follows:

$$\hat{\boldsymbol{B}}\hat{\boldsymbol{A}}\hat{\boldsymbol{y}} = \hat{\boldsymbol{B}}\boldsymbol{X}\left(\boldsymbol{X}^{\boldsymbol{T}}\hat{\boldsymbol{B}}^{\boldsymbol{T}}\hat{\boldsymbol{B}}\boldsymbol{X}\right)^{-1}\boldsymbol{X}^{\boldsymbol{T}}\hat{\boldsymbol{B}}^{\boldsymbol{T}}\hat{\boldsymbol{B}}\hat{\boldsymbol{A}}\boldsymbol{y}$$

$$\xrightarrow{\boldsymbol{yields}} \hat{\boldsymbol{y}} = \hat{\boldsymbol{A}}^{-1}\boldsymbol{X}\left(\boldsymbol{X}^{\boldsymbol{T}}\hat{\boldsymbol{V}}^{-1}\boldsymbol{X}\right)^{-1}\boldsymbol{X}^{\boldsymbol{T}}\hat{\boldsymbol{V}}^{-1}\hat{\boldsymbol{A}}\boldsymbol{y}$$

Denote the projection matrix in the transformed spatial regression model as $\boldsymbol{P_s}$,

$$\boldsymbol{P_s} = \boldsymbol{X}^{*}\left(\boldsymbol{X}^{*\boldsymbol{T}}\boldsymbol{X}^{*}\right)^{-1}\boldsymbol{X}^{*\boldsymbol{T}}$$

$$= \hat{\boldsymbol{B}}\boldsymbol{X}\left(\boldsymbol{X}^{\boldsymbol{T}}\hat{\boldsymbol{V}}^{-1}\boldsymbol{X}\right)^{-1}\boldsymbol{X}^{\boldsymbol{T}}\hat{\boldsymbol{B}}^{\boldsymbol{T}},\ \hat{\boldsymbol{V}} = \left(\hat{\boldsymbol{B}}^{\boldsymbol{T}}\hat{\boldsymbol{B}}\right)^{-1}$$

Properties of leverage in the transformed spatial model in Equation 13:

Property I: Idempotent and symmetric.

Property Ia: Idempotence

$$\boldsymbol{P_s^2} = \hat{\boldsymbol{B}}\boldsymbol{X}\left(\boldsymbol{X}^{\boldsymbol{T}}\hat{\boldsymbol{V}}^{-1}\boldsymbol{X}\right)^{-1}\boldsymbol{X}^{\boldsymbol{T}}\hat{\boldsymbol{B}}^{\boldsymbol{T}}\hat{\boldsymbol{B}}\boldsymbol{X}\left(\boldsymbol{X}^{\boldsymbol{T}}\hat{\boldsymbol{V}}^{-1}\boldsymbol{X}\right)^{-1}\boldsymbol{X}^{\boldsymbol{T}}\hat{\boldsymbol{B}}^{\boldsymbol{T}}$$

$$= \hat{\boldsymbol{B}}\boldsymbol{X}\left(\boldsymbol{X}^{\boldsymbol{T}}\hat{\boldsymbol{V}}\boldsymbol{X}\right)^{-1}\boldsymbol{X}^{\boldsymbol{T}}\hat{\boldsymbol{B}}^{\boldsymbol{T}}$$

$$= \boldsymbol{P_s}$$

Hence, $\boldsymbol{P_s}$ is idempotent.

Property Ib: Symmetric

$$\boldsymbol{P_s^T} = \left(\hat{\boldsymbol{B}}\boldsymbol{X}\left(\boldsymbol{X}^{\boldsymbol{T}}\hat{\boldsymbol{V}}^{-1}\boldsymbol{X}\right)^{-1}\boldsymbol{X}^{\boldsymbol{T}}\hat{\boldsymbol{B}}^{\boldsymbol{T}}\right)^{\boldsymbol{T}}$$

$$= \hat{\boldsymbol{B}}\boldsymbol{X}\left(\boldsymbol{X}^{\boldsymbol{T}}\hat{\boldsymbol{V}}^{-1}\boldsymbol{X}\right)^{-1}\boldsymbol{X}^{\boldsymbol{T}}\hat{\boldsymbol{B}}^{\boldsymbol{T}}$$

$$= \boldsymbol{P_s}$$

The matrix $\boldsymbol{P_s}$ is symmetric. Therefore, $\boldsymbol{P_s}$ in the transformation $\boldsymbol{z}^{*} = \hat{\boldsymbol{B}}\hat{\boldsymbol{A}}\boldsymbol{y}$ is both idempotent and symmetry.

Property II: The sum of diagonal term of projection matrix is $\boldsymbol{k}$, the number of parameters including the constant term.

$$\boldsymbol{trace}(\boldsymbol{P_s}) = \boldsymbol{trace}\left(\hat{\boldsymbol{B}}\boldsymbol{X}\left(\boldsymbol{X}^{\boldsymbol{T}}\hat{\boldsymbol{V}}^{-1}\boldsymbol{X}\right)^{-1}\boldsymbol{X}^{\boldsymbol{T}}\hat{\boldsymbol{B}}^{\boldsymbol{T}}\right)$$

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 9 August 2021 doi:10.20944/preprints202108.0178.v1

$$= \boldsymbol{trace}\left(\widehat{\boldsymbol{B}}^{\boldsymbol{T}}\widehat{\boldsymbol{B}}\boldsymbol{X}\left(\boldsymbol{X}^{\boldsymbol{T}}\widehat{\boldsymbol{V}}^{-\boldsymbol{1}}\boldsymbol{X}\right)^{-\boldsymbol{1}}\boldsymbol{X}^{\boldsymbol{T}}\right)(\boldsymbol{cyclic\ permutation\ of\ trace\ of\ matrix})$$

$$= \boldsymbol{trace}\left(\boldsymbol{X}^{\boldsymbol{T}}\widehat{\boldsymbol{V}}^{-\boldsymbol{1}}\boldsymbol{X}\left(\boldsymbol{X}^{\boldsymbol{T}}\widehat{\boldsymbol{V}}^{-\boldsymbol{1}}\boldsymbol{X}\right)^{-\boldsymbol{1}}\right)(\boldsymbol{cyclic\ permutation\ of\ trace\ of\ matrix})$$

$$= \boldsymbol{trace}(\boldsymbol{I_k})$$

$$= \boldsymbol{k},$$

where, $\boldsymbol{I_k}$ is an $\boldsymbol{k} \times \boldsymbol{k}$ identity matrix.

Therefore, $\sum_{i=1}^{k} \boldsymbol{ps_{ii}} = \boldsymbol{k}$. $\boldsymbol{ps_{ii}}$ is the $\boldsymbol{i^{th}}$ diagonal element of the leverage $\boldsymbol{P_s}$.

Property III: Bounds on spatial leverage.

The bound on the leverage of the classical regression is $\boldsymbol{0} \leq \boldsymbol{p_{ii}} \leq \boldsymbol{1}$ due to the fact that the hat matrix $\boldsymbol{P}$ satisfies all the orthogonal properties, including symmetry. As such, it is positive semi-definite. However, the spatial leverage, $\boldsymbol{P_{ay}}$ is not symmetric because positive semi-definite matrix is symmetric **[24–26]**. The transformation in Equation 11 yielded the projection $\boldsymbol{P_s}$ that satisfies the symmetry condition.

From the idempotent property of $\boldsymbol{P_s}$,

$\boldsymbol{P_s} = \boldsymbol{P_s^2}$.

Equating diagonal terms of LHS and RHS, we have

$$\boldsymbol{ps_{ii}} = \boldsymbol{ps_{ii}^2} + \sum_{j \neq i} \boldsymbol{ps_{ij}}\,\boldsymbol{ps_{ji}},\ \ \sum_{j \neq i} \boldsymbol{ps_{ij}}\,\boldsymbol{ps_{ji}} \geq \boldsymbol{0}, \tag{14}$$

where $\boldsymbol{ps_{ij}}$ are the off diagonal terms. Equation 14 implies that $\boldsymbol{ps_{ii}} \geq \boldsymbol{0}$. Therefore,

$$\boldsymbol{ps_{ii}} \geq \boldsymbol{ps_{ii}^2}$$

$$\xrightarrow{\boldsymbol{yields}} \boldsymbol{ps_{ii}} \leq \boldsymbol{1}.$$

Note that $\boldsymbol{P_s}$ and $\boldsymbol{Q_s}$ are orthogonal:

1. $\boldsymbol{P_s P_s} = \boldsymbol{P_s}$
2. $\boldsymbol{Q_s Q_s} = \boldsymbol{Q_s}$
3. $\boldsymbol{P_s Q_s} = \boldsymbol{0}$

The model in Equation 9 gives rise to different special spatial regression in accordance with different restrictions. Such special spatial regression models are the spatial autoregressive-regressive model (SAR) and the Spatial Error model (SEM). While the former has spatial autoregression in the response variable, the latter has spatial autoregression in model residual; model 9, (GSM), combines both features.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 9 August 2021 doi:10.20944/preprints202108.0178.v1

The spatial autoregressive-regressive model is obtained when the coefficient of the lagged spatial autoregression in the residuals of Equation 9 is zero, i.e., $\boldsymbol{\lambda = 0}$. Thus, the SAR model is given by

$$\boldsymbol{y = \rho W_1 y + X\beta + \varepsilon,\ \ \varepsilon \sim N(0, \sigma^2 I_n)}. \tag{15}$$

The corresponding $\boldsymbol{P_s}$ to model in Equation 13 reduces to

$$\boldsymbol{P_s = X(X^T X)^{-1} X^T},$$

with the transformation in Equation 11 simplifying to $\boldsymbol{z^* = Ay}$, since, $\boldsymbol{V = (B^T B)^{-1}}$ and $\boldsymbol{B = I_n}$, when $\boldsymbol{\lambda = 0}$. Clearly, the hat matrix in the SAR model preserves the features of the hat matrix in the classical regression model.

In the spatial error model (SEM), the coefficient of the spatial autoregression on the lagged dependent variable is zero, i.e. $\boldsymbol{\rho = 0}$. This yields the model

$$\boldsymbol{y = X\beta + \xi,\ \ \xi = \lambda W_2 \xi + \varepsilon,\ \ \varepsilon \sim N(0, \sigma^2 I_n)}. \tag{16}$$

The transformation in Equation 11 simplifies to $\boldsymbol{z^* = By}$ and the projection matrix remains

$$\boldsymbol{P_s = BX(X^T V^{-1} X)^{-1} X^T B^T}.$$

It can be observed that leverage measure in spatial regression model is dominated by the autocorrelation in the residual term.

Works on spatial regression diagnostic in the literature mainly focused on autocorrelation in the residuals, mostly using time series analogy **[13–15]**. Some remarkable works in the spatial regression model can be found in **[17,18,20]**.

3.3 Influential Observations in spatial regression model

The leverages $\boldsymbol{P_s}$ and $\boldsymbol{Q_s}$ in Equation 11 satisfy all the properties of projection matrix, including sum of diagonal terms of $\boldsymbol{P_s}$ and $\boldsymbol{Q_s}$ equal $\boldsymbol{k}$ and $\boldsymbol{n - k}$ respectively. It also incorporates the autocorrelation in the dependent variable, $\boldsymbol{Wy}$. Hence, it can be used as a diagnostic measure of leverage points in spatial regression model.

By extending the results of linear regression to spatial regression with slight modification, the Cook's distance in the spatial regression of Equation 13, denoted as $\boldsymbol{CD_{si}}$, can be formulated as follows

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 9 August 2021 doi:10.20944/preprints202108.0178.v1

$$\widehat{CD}_{si} = \frac{\left(\hat{\beta}_s^{(-i)} - \hat{\beta}_s\right)^T (X^{*T}X^*)\left(\hat{\beta}_s^{(-i)} - \hat{\beta}_s\right)}{k\widehat{\sigma^2}}$$

$$= \frac{\left(\hat{\beta}_s^{(-i)} - \hat{\beta}_s\right)^T \left((B^{-1}X)^T (B^{-1}X)\right)\left(\hat{\beta}_s^{(-i)} - \hat{\beta}_s\right)}{k\widehat{\sigma^2}}$$

$$= \frac{\left(\hat{\beta}_s^{(-i)} - \hat{\beta}_s\right)^T \left(X^T (B^{-1})^T B^{-1}X\right)\left(\hat{\beta}_s^{(-i)} - \hat{\beta}_s\right)}{k\widehat{\sigma^2}}$$

$$= \frac{\left(\hat{\beta}_s^{(-i)} - \hat{\beta}_s\right)^T (X^T V^{-1}X)\left(\hat{\beta}_s^{(-i)} - \hat{\beta}_s\right)}{k\widehat{\sigma^2}},$$

where

$$\hat{\beta}_s^{(-i)} = X_{(i)}\left(X_{(i)}^T \hat{V}_{(i,i)}^{-1} X_{(i)}\right)^{-1} X_{(i)}^T \hat{V}_{(i,i)}^{-1} \hat{A}_{(i,i)} Y_{(i)}.$$

$\hat{V}_{(i,i)}$ and $\hat{A}_{(i,i)}$ denote $\hat{V}$ and $\hat{A}$ with the i^{th} row and the i^{th} column deleted.

The spatial Cook distance, CD_{si}, is declared large if $CD_{si} > 0.70$ **[18]**. In its simplified form, the Cook's distance in spatial regression is written as

$$\widehat{CD}_{si}(X^T V^{-1}X, k\hat{\sigma}^2) = \frac{1}{k} t_{si}^2 \frac{p_{si}}{q_{si}}, \tag{17}$$

where t_{si} is the spatial studentized prediction residual (also called spatial internally studentized residual), p_{si} is the spatial leverage, which is the i^{th} diagonal element of P_s, and $q_{si} = 1 - p_{si}$. Let $r_{si} = y_i - \hat{y}_i$, then

$$t_{si} = \frac{b_i^T a_i r_{si}}{\hat{\sigma}\sqrt{q_{si}}}, \tag{18}$$

where b_i and a_i are the i^{th} columns of matrices B and A respectively. The spatial studentized residual has a cut off point of 2 to declare a point large **[18,27]**.

Similarly, the spatial externally studentized residual (ESRs), is defined as

$$t_{si}^* = \frac{r_{si}}{\hat{\sigma}_{(i)}\sqrt{1 - p_{si}}}$$

$$= t_{si}\sqrt{\frac{n-k-1}{n-k-t_{si}^2}}, \quad \hat{\sigma}_{(i)} = \hat{\sigma}\left(\frac{n-k-t_{si}}{n-k-1}\right).$$

where $\hat{\sigma}_{(i)}$ is the residuals mean square excluding the i^{th} case. The ESRs follows a Student's t-distribution with $(n-k-1)$ degrees of freedom. Thus, the spatial studentized prediction residuals contain the neighbourhood information on both the dependent variable and the residual of each r_{si}, and the leverage P_s, contains the residual autocorrelation effect. The spatial potential, which is analogous to the potential **[10]**, is defined in Equation 19 as,

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 9 August 2021 doi:10.20944/preprints202108.0178.v1

$$p_{osi} = \frac{p_{si}}{q_{si}}. \tag{19}$$

where $q_{si} = 1 - p_{si}$. Let $q_{osi} = 1 - p_{osi}$.

We define the spatial measure of overall potential influence as

$$\boldsymbol{H_{si}^2} = \frac{\boldsymbol{k}}{q_{osi}} \frac{\boldsymbol{d_i^2}}{(\mathbf{1} - \boldsymbol{d_i^2})} + \frac{\boldsymbol{p_{osi}}}{\boldsymbol{q_{osi}}}. \tag{20}$$

When measuring the influence of an observation in linear regression model by using the Cook's distance **[3]**, the observation in question is deleted and the model is then refitted. In a similar way, usually a group of suspected influential observations are deleted in the linear regression and admitted into the model if they are proven clean (BACON **[28]**, **[29]**, DGRP **[11]**). This is because IOs in linear regression are global in nature; however, in spatial regression model, IOs are local. Haining **[19]** noted that Spatial outliers are local in nature; their attribute values are outliers if they are extreme relative to the set of values in their neighbourhood on the map. IOs in spatio-temporal statistics usually carry vital information in applications. Kou et al. **[30]** further pointed out that detecting spatial outliers can help in locating extreme meteorological events such as tornadoes and hurricanes, identify aberrant genes or tumor cells, discover highway traffic congestion points, pinpoint military targets in satellite images, determine possible locations of oil reservoirs and detect water pollution incidents. Thus, measuring the influence of multiple spatial locations requires contiguous set of points to reveal the unusual features related to that neighbourhood.

Though, methods that detect multiple outliers in spatial regression work well (see **[20]**), the methods referred to above group observations as clean or suspect, irrespective of their positions (with reference to spatial data), and admitted into the model according to some conditions as clean observations.

According to Hadi**[10]**, examining each value of influence measure alone, such as $\boldsymbol{P_{si}}$, ISRs, ESRs, $\boldsymbol{CD_{si}}$, $\boldsymbol{H_{si}^2}$, might not be successful to indicate the IOs or source of influence. Imon **[31]** noted that one should consider both outliers and leverage points when identifying IOs. The easiest way to capture IOs is by using diagnostic plots. Following **[32,33]**, we adopt their rules for classification of observations into four categories, namely regular observations, vertical outliers, GLPs and BLPs. Once observations are classified accordingly, those observations that fall in the vertical outliers and BLPs categories are referred to as IOs. However, due to local nature of spatial IOs, we have to make some modifications on the classification scheme. In this paper, a new diagnostic plot is proposed by plotting the ISRs (or ESRs) in the *Y*-axis against the spatial potential, P_{osi}, in the *X*-axis. We consider the *ISRs* and *ESR*s because both measures contain spatial information. On the other hand, the potentials that are obtained from the transformed model in Equation 13 is considered in order to reflect spatial dependence. Hence, the proposed diagnostic plots are denoted as ISRs $-$ P_{osi} and ESRs $-$ P_{osi} plot and it is based on the following classification scheme:

a) ISRs $-$ P_{osi}

i) i^{th} observation is declared RO if $|ISRs| < 2.0$ and $p_{osi} < l_2$.

ii) i^{th} observation is GLP if $|ISRs| < 2.0$ and $p_{osi} > l_2$.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 9 August 2021 doi:10.20944/preprints202108.0178.v1

iii) i^{th} observation is BLP if $|ISRs| > 2.0$ and $p_{si} > l_2$.

iv) i^{th} observation is IO if $|ISRs| \geq 2.0$ and $p_{si} \leq l_2$.

Figure 1 and 2 show the classification of the observations as RO, GLP and IOs according to ISRs − P_{osi} and ESRs − P_{osi}, respectively.

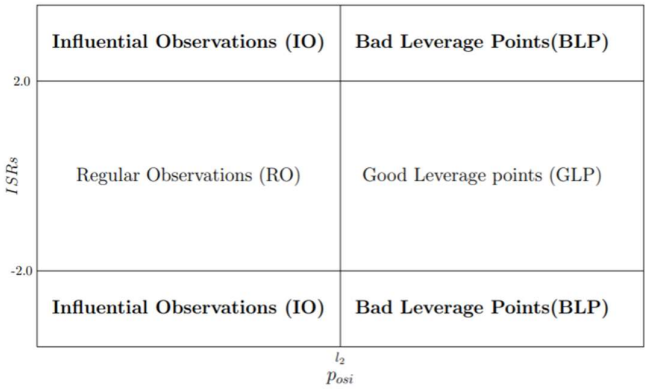


Figure 1. Classification of RO, GLP, and IO according ISRs − P_{osi}

b) ESRs − P_{osi}

i) $\boldsymbol{i^{th}}$ observation is declared RO if $|ESRs| < \boldsymbol{t_{n-k-1}}$ and $\boldsymbol{p_{osi} < l_2}$.

ii) $\boldsymbol{i^{th}}$ observation is GL if $|ESRs| < \boldsymbol{t_{n-k-1}}$ and $\boldsymbol{p_{osi} > l_2}$.

iii) $\boldsymbol{i^{th}}$ observation is IO if $|ESRs| > \boldsymbol{t_{n-k-1}}$ and $\boldsymbol{p_{si} > l_2}$.

iv) $\boldsymbol{i^{th}}$ observation is IO if $|ESRs| \geq \boldsymbol{t_{n-k-1}}$ and $\boldsymbol{p_{si} \leq l_2}$.

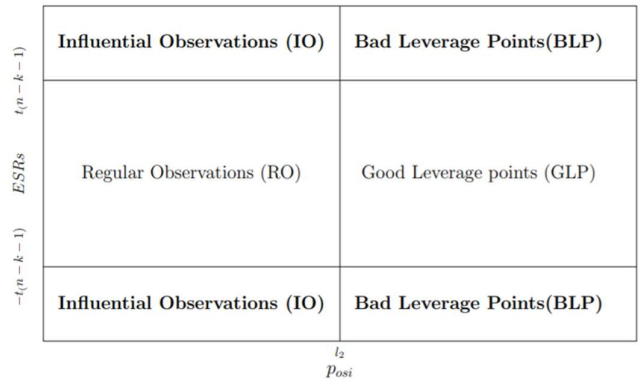


Figure 2. Classification of RO, GLP, and IO according ESRs − P_{osi}

4.0 Results and Discussions

In this section, the performance of all the proposed methods, i.e. the, Cook's Distance ($\widehat{\boldsymbol{CD}}_{\boldsymbol{si}}$), $\boldsymbol{H^2_{si}}$($\boldsymbol{H^2_{si1}}$(non-robust) and $\boldsymbol{H^2_{si2}}$(robust)), $ISRs - P_{osi}$ and $ESRs - P_{osi}$ are evaluated using simulation study, artificial data and real datasets of gasoline price data in the South-

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 9 August 2021 doi:10.20944/preprints202108.0178.v1

West area of Sheffield, UK, Covid-19 data in the counties of Georgia state, USA and Life expectancy data in counties of USA.

4.1. Simulated data

We simulated a spatial regression model in Equation 8 for a square spatial grid with sample size, $\boldsymbol{n = 400}$, $\boldsymbol{\rho = 0.4}$, $\boldsymbol{\lambda = 0.5}$ and $\boldsymbol{W_1 = W_2}$, using row standardized Queen's contiguity spatial weight. $\boldsymbol{x_0 = 1}$, $\boldsymbol{x_1 \sim N(0,1)}$, $\boldsymbol{\beta_0 = 0}$, $\boldsymbol{\beta_1 = 1}$ (Bold face 0 and 1 refer to column vectors of values zeros and ones respectively). The contamination is taken at two percent in each of X and y directions. The contamination in the $\boldsymbol{y}$ direction is taken from the Cauchy distribution due to its fat tails. Contamination in the $\boldsymbol{X}$ direction is taken from the following multivariate distribution,

$$\boldsymbol{X} \sim \left(\begin{bmatrix}\mathbf{0}\\ \mathbf{2}\end{bmatrix}, \begin{bmatrix}\mathbf{1} & \mathbf{0}\\ \mathbf{0} & \mathbf{1}\end{bmatrix}\right).$$

However, it is important to note that, during the contamination, some of the contaminations may have attributes similar to those in their neighbourhood, as noted by Dowd **[34]**, spatial simulation is conditioned to a real data set.

Figure 3 shows the graph of average attribute values in the neighbourhood of locations against their attribute values with added contamination. It can be observed that some of the added contamination, in black dots, are in the middle of clean data points while some stand out from the bulk of the data (i.e., away from their average neighbourhood values) which clearly indicates outlyingness.

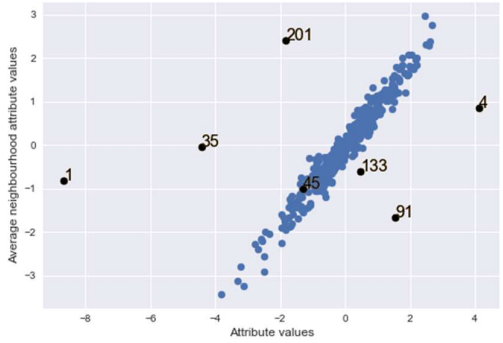


Figure 3. Graph of average attribute in neighbourhood of locations against the attribute in the locations with contamination (black points).

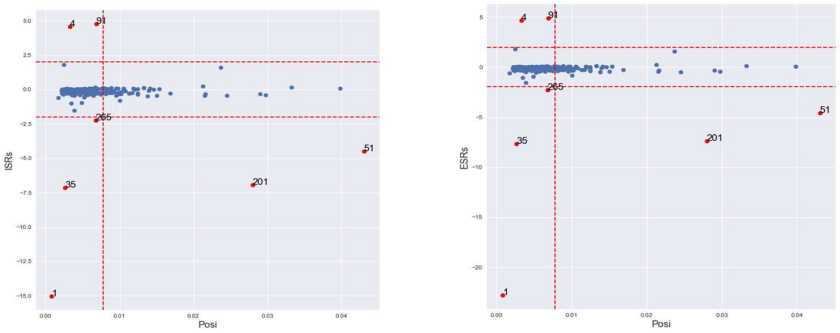
Table 1. ISRs, ESRs and $\boldsymbol{p_{osi}}$ of locations with large studentized residual in the simulated GSM model, with their cutoff points in parenthesis.

Location	$ISRs$ (2.00)	$ESRs$ (1.97)	p_{osi} (0.0078)
1	15.0378	22.8179	0.0008
4	4.5847	4.7046	0.0033
35	-7.1434	-7.6397	0.0026
51	-4.4695	-4.5801	0.0430

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 9 August 2021 doi:10.20944/preprints202108.0178.v1

91	4.7613	4.8965	0.0068
201	-6.9336	-7.3840	0.0280
265	-2.2644	-2.2762	0.0068

Table 1 presents the values of ISRs, ESRs and $\boldsymbol{p_{osi}}$ where values in parenthesis are their corresponding cut-off points. It shows seven locations with large studentized residuals according to ISRs and ESRs. There are 54 observations with large potentials (>0.0078). Two out of the 54 potentials correspond to studentized residual greater than the thresholds of ISRs and ESRs (locations 51 and 201).



(a) $\boldsymbol{ISRs - P_{osi}}$. (b) $\boldsymbol{ESRs - P_{osi}}$

Figure 4. Graph of IO classification according to GLP, BLP and vertical outlier in $\mathbf{ISRs - P_{osi}}$ and $\mathbf{ESRs - P_{osi}}$ for simulated data.

In order to confirm the outlyingness of the locations classified as spatial IOs, the threshold of each outlier neighbourhood given by,

$$\text{med}_i + 3MAD_i \,,$$

is computed for the studentized residuals of the classified location and its immediate neighbourhood, where med_i is the median of the studentized residual and MAD_i is the median absolute deviation. The absolute value of the studentized residuals is compared to the neighbourhood threshold for confirmation as outlier.

The CD_{si} detected location 201, which has large ISRs, ESRs and p_{si}. $ISRs - P_{osi}$ and $ESRs - P_{osi}$ classified locations 1, 4, 35, 51, 91, 201 and 265 as IOs. As noted from Figure 4, $ISRs - P_{osi}$ and $ESRs - P_{osi}$ classified locations 1, 4, 35, 91 and 265 as outliers in y direction, and locations 51 and 201 in both X and y directions. The cut-off limits of $ESRs - P_{osi}$ are narrower than 2 for the 5% cut-off point of the student t distribution which is around 1.96 for large sample sizes.

H_{si1}^2 classified location 1 only as IO. Location 1 has large ISRs and ESRs with small p_{osi}. It is an outlier in the y direction. H_{si}^2 identified 60 locations as IOs, including all the locations classified by the other methods. However, diagnostic examination of the 53 other locations classified by H_{si2}^2 alone reveals that all locations that have small ISRs and ESRs with large potential values are classified as IOs. Moreover, the locations with small studentized residuals, which show no difference with their neighbourhood, are classified as IOs. This is a clear case of swamping. Perhaps, this is due local nature of the spatial IOs.

In a 1000 run of the simulation described above at different error variances of 0.01, 0.1, 0.2 and 0.3 as shown in Table 2, the CD_{si} consistently maintained low classification of

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 9 August 2021 doi:10.20944/preprints202108.0178.v1

influential observations with consistent swamping rates of 0%. The $\boldsymbol{ISRs - P_{osi}}$ has demonstrated high detection to the tune of 98% while $\boldsymbol{ESRs - P_{osi}}$ has 100% accurate classification of the IOs, both with swamping rates of 0%. H_{si1}^2 has less than 40% accurate classification with zero swamping rate, while the H_{si2}^2 attend up to 99% accurate IO classification, but usually with very high swamping rates.

Table 2. Influential observations classification rate based on large prediction studentized Residuals and large potentials.

σ^2	Method	Accurate classification (%)	Swamping (%)
0.01	CD_{si}	22.25	0.0
	$ISRs - P_{osi}$	98.54	0.0
	$ESRs - P_{osi}$	100.00	0.0
	H_{si1}^2	39.45	0.0
	H_{si2}^2	99.71	81.41
0.1	CD_{si}	20.64	0.0
	$ISRs - P_{osi}$	98.36	0.0
	$ESRs - P_{osi}$	100.00	0.0
	H_{si1}^2	38.09	0.0
	H_{si2}^2	99.14	76.48
0.2	CD_{si}	17.86	0.00
	$ISRs - P_{osi}$	97.51	0.00
	$ESRs - P_{osi}$	100.00	0.00
	H_{si1}^2	37.23	0.00
	H_{si2}^2	97.34	69.25
0.3	CD_{si}	16.36	0.00
	$ISRs - P_{osi}$	96.57	0.00
	$ESRs - P_{osi}$	100.00	0.00
	H_{si1}^2	36.23	0.00
	H_{si2}^2	96.00	64.42

1. Illustrative examples

5.1 Example 1

The gasoline price data for 61 retail sites in the South-West area of Sheffield from [18] are used in Example 1. The analysis indicates the presence of spatial interaction in the error term with a Moran's I of 0.239.

The fitted SEM model is given by Equation 21.

$$\hat{y}_M = 35.78 + 0.71X_F + \hat{\lambda}W\xi \quad (21)$$

where, y_M and X_M are March and February sales from the South-West Sheffield gasoline sale data, respectively, $\hat{\lambda} = 0.15$ is the estimate of coefficient of correlation in the residual, W is the standardized weight matrix and ξ is the vector of correlated residuals.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 9 August 2021 doi:10.20944/preprints202108.0178.v1

Table 3 shows the results of the detected IOs in the SEM model for the gasoline data with all the sites detected by the methods. A 'yes' under a method column indicates that the site has been detected by the method as IO and a 'no' means otherwise. The values in bold in columns ISRs, ESRs and p_{si} indicate large studentized residuals and potentials greater than 0.0335, respectively. Figure 5 shows the classification of observations by $ISRs - P_{osi}$ and $ESRs - P_{osi}$.

Table 3: Sites with IOs in the analysis of the South-West Sheffield gasoline data.

S/N	Site	ISRs (2.00)	ESRs (2.00)	p_{osi} (0.0335)	CD_{si}	$ISRs - P_{osi}$	$ESRs - P_{osi}$	H^2_{si1}	H^2_{si2}
1.	3	-1.8879	-1.9301	**0.3538**	no	No	No	no	yes
2.	9	1.4810	1.4962	0.0223	no	No	No	no	yes
3.	22	1.0127	1.0129	**0.0779**	no	No	No	no	yes
4.	25	**5.4292**	**7.5481**	**0.2773**	yes	Yes	Yes	yes	yes
5.	26	1.4438	1.4573	**0.1352**	no	No	No	no	yes
6.	30	**2.2054**	**2.2813**	**0.2489**	no	No	No	no	yes
7.	40	1.5692	1.5890	0.0194	no	No	No	no	yes
8.	41	1.1974	1.2058	0.0218	no	No	No	no	yes
9.	42	-1.9150	-1.9598	**0.0378**	no	No	No	no	yes
10.	46	0.1003	0.0995	**0.1319**	no	No	No	no	yes
11.	55	-1.2042	-1.2089	0.0219	no	No	No	no	yes
12.	61	-1.8011	-1.8363	0.0319	no	No	No	no	Yes

The CD_{si}, $ISRs - P_{osi}$, $ESRs - P_{osi}$, and H^2_{si1} coincidentally identified site 25 only as IO. H^2_{si2} detects 11 more sites as IOs in addition to site 25. Haining [18] has made elaborate diagnostic analysis of the data where he emphasized the effect of site 25 as IO in the data. Our methods have classified site 30 in addition to location 25 as IO. Figure 5 shows the graph of the lagged residuals against the residuals. It is noticeable from the graph that site 30 has also been marked, and hence IO.

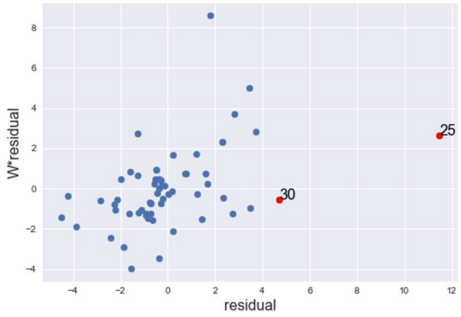
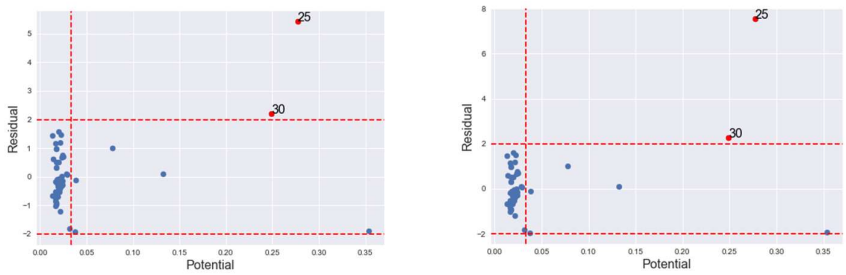


Figure 5. Graph of the lagged residuals against the residuals, of the 61 sites of the south-west Sheffield fitted for SEM, showing the IO points in red dots.

Though the H^2_{si2} has detected all the suspected IO, it's prone to swamping. The remaining high potentials are classified as GLP by $ISRs - P_{osi}$ and $ESRs - P_{osi}$ since their studentized values are small.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 9 August 2021 doi:10.20944/preprints202108.0178.v1



(a) $\boldsymbol{ISRs - P_{osi}}$. (b) $\boldsymbol{ESRs - P_{osi}}$.

Figure 6. Graph of IO classification according to GLP, BLP and vertical outlier in South-West gasoline data.

Figure 6 shows the graph of classification of the $\boldsymbol{ISRs - P_{osi}}$ (a) and $\boldsymbol{ESRs - P_{osi}}$ (b) indicating the outliers in red dots, where both are classified as outliers in both X and y directions.

5.2. Example 2

The data for example 2 is the Covid-19 data for the 159 counties of Georgia state, USA, as at 30/06/2020 (http://dph.georgia.gov/covid-19-daily-status-report). The health ranking (http://www.countyheathrankings.org). The case-rate per 100000 of Covid19 is the dependent variable. The independent variables are population of black race in the county (X_1), population of Asians (X_2), population of Hispanic (X_3), population of people that are 65 years and above (X_4), population of female in the county (X_5) and life expectancy (X_6).

The model is fitted to the SAR model (Model with lowest Akaike Information Criteria (AIC) value of 2192). The SAR model is presented in Equation 22.

$$\hat{y} = \hat{\rho}W\hat{y} + \hat{\beta}_0 + \sum_{i=1}^{6} \hat{\beta}_i X_i, \tag{22}$$

where, $\hat{\rho} = 0.6967$, $\hat{\beta}_0 = 1087.7388$, $\hat{\beta}_1 = 9.7831$, $\hat{\beta}_2 = -6.2210$, $\hat{\beta}_3 = -54.1402$, $\hat{\beta}_4 = -28.5874$, $\hat{\beta}_5 = 4.8288$ and $\hat{\beta}_6 = 40.3323$. X_1 and X_3 and $\hat{\rho}$ are significant at 5%, while X_2 and X_5 are significant at 10%. X_4 and X_6 are not significant.

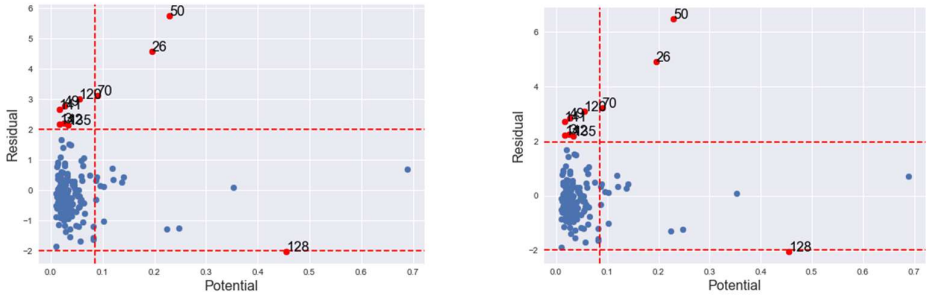
The Cook's distance classified only county 50 as IO. The $\boldsymbol{ISRs - P_{osi}}$ and $\boldsymbol{ESRs - P_{osi}}$ coincided in detecting counties 3, 26, 49, 50, 70, 120, 135, 141 and 142 as IOs. The H^2_{si1} (non-robust) detected 26 and 50 as IOs. The H^2_{si2} (robust) detected 3, 26, 50, 58, 67, 70, 98, 118, 120, 128, 131, 134, 135, 139, 141, 142, 153 and 155 counties. Table 4 shows the detected locations by the various methods with large ISRs, ESRs and high potentials in bold fond.

The IOs identified by $\boldsymbol{ISRs - P_{osi}}$ and $\boldsymbol{ESRs - P_{osi}}$ have both large studentized residuals and large potentials as could be observed in Table 4. Figure 7 shows the outliers in X, y and both X and y directions. The CD_{si} detected largest studentized residual with high potential as IO. The H^2_{si} identified two observations with large studentized value and high potential values. The H^2_{si2} detected all suspected IOs, but with many having both small values of studentized residuals and potential values.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 9 August 2021 doi:10.20944/preprints202108.0178.v1

Table 4. Detected IOs Counties by different methods in the Georgia Covid-19 data.

County	ISRs (2.00)	ESRs (1.98)	p_{si} (0.0851)	CD_{si}	$ISRs - P_{osi}$	$ESRs - P_{osi}$	H^2_{si1}	H^2_{si2}
3	**2.2245**	**2.2539**	0.0257	no	no	no	no	yes
26	**4.5733**	**4.9060**	**0.1956**	no	yes	yes	yes	yes
49	**2.7685**	**2.8313**	0.0265	no	yes	Yes	no	yes
50	**5.7504**	**6.4737**	**0.2298**	yes	yes	Yes	yes	yes
58	0.7090	0.7079	**0.6893**	no	no	No	yes	yes
67	0.1018	0.1015	**0.3524**	no	no	No	no	yes
70	**3.1334**	**3.2285**	**0.0895**	no	yes	Yes	no	yes
98	-1.8549	-1.8699	0.0105	no	no	No	no	yes
118	-1.5657	-1.5731	**0.0827**	no	no	No	no	yes
120	**3.0168**	**3.1006**	0.0544	no	yes	Yes	no	yes
128	**-2.0152**	**-2.0359**	**0.4557**	no	yes	yes	no	yes
131	-1.6718	-1.6862	0.0565	no	no	No	no	yes
134	-1.6168	-1.6253	**0.0818**	no	no	No	no	yes
135	**2.1674**	**2.1942**	0.0338	no	yes	Yes	no	yes
141	**2.6726**	**2.7283**	0.0163	no	Yes	yes	no	yes
142	**2.1805**	**2.2079**	0.0174	Yes	yes	no	no	yes
153	-1.2693	-1.2718	**0.2234**	No	No	no	no	yes
155	-1.2334	-1.2359	**0.2472**	No	No	no	no	yes



kljbh

(a) $ISRs - P_{osi}$ (b) $ESRs - P_{osi}$

Figure 7. Graph of IO classification according to GLP, BLP and vertical outlier in Georgia state, USA covid-19 data.

While examining the outlyingness of the classified counties, it is found that county 50 is clearly an IO since it has both large studentized residual and large potential value. It's outside the threshold value of its neighboughood.

Four of the counties classified by $ISRs - P_{osi}$ and $ESRs - P_{osi}$ (i.e., 26, 50, 70 and 128) are classified as vertical outliers while the counties 3, 49, 120, 135,141 and 142 have large

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 9 August 2021 doi:10.20944/preprints202108.0178.v1

potential values and studentized values greater than their threshold values, classified as BLP and hence IO.

Beside the counties classified by $\boldsymbol{ISRs - P_{osi}}$ and $\boldsymbol{ESRs - P_{osi}}$, all the other counties detected by H_{si2}^2 have their studentized difference residuals below their neighbourhood threshold. Though their potential values are mostly large, their prediction studentized residuals are small in both ISRs and ESRs.

5.3. Example 3

In Example 3, the life expectancy of the counties of the US is measured on population density (X_1), fair/poor health status (X_2), obesity (X_3), population in rural area (X_4), inactivity rate (X_5), population of smokers (X_6), population of black (X_7), population of Asians (X_8)and population of Hawaiians (X_9). The data are obtained from the Kaggle website(https://www.kaggle.com/johnjdavisiv/us-counties-covid19-weather-sociohealth-data).

The spatial error model (SEM) has the lowest AIC value, and is fitted for the data. The model is significant at 5% level with a significant Moran's I of 0.2160. X_1 and X_4 are not significant at 5%. All the other estimates are significant at 5% level.
The fitted model is given by

$$\hat{y} = \hat{\beta}_0 + \sum_{i=1}^{9} \hat{\beta}_i X_i + \hat{\lambda} W\xi \tag{23}$$

where $\hat{\lambda} = 0.4343$, $\hat{\beta}_0 = 88.4885$, $\hat{\beta}_1 = 0.0000$, $\hat{\beta}_2 = -0.0954\$$, $\hat{\beta}_3 = -0.0377$, $\hat{\beta}_4 = 0.0040$, $\hat{\beta}_5 = -0.0630$, $\hat{\beta}_6 = -0.3892$, $\hat{\beta}_7 = -0.0113$, $\hat{\beta}_8 = 0.1437$, $\hat{\beta}_9 = -0.2016$. Counties with fair/poor health facility have lower life expectancy at rate of 0.1 for increase in the population. Counties with larger number of obese people have a decrease in life expectancy at the rate of 0.03. Similarly, those counties with large number of people with in-activity have decreased life expectancy at the rate 0.06, counties with larger number of smokers has decrease at 0.04 per increase in population. Counties with higher number of Blacks and Hawaiians have decreased life expectancy at the rate 0.01 and 0.2 respectively, while those with higher number of Asians have an increased rate of 0.14 for population increased.

The $\boldsymbol{ISRs - P_{osi}}$ classified 139 counties as IOs, while $\boldsymbol{ESRs - P_{osi}}$ classified additional 8 more counties, making a total of 147. H_{si1}^2 and H_{si2}^2 have respectively, classified 24 and 324 counties as IOs. CD_{si} classified no county.

2. Conclusions

In this article, we have demonstrated the employment of influential observations (IOs) detection techniques in the classical regression to the spatial regression model. Measures that contain spatial information in the spatial autoregression in the dependent variable and residuals are obtained. We have also evaluated the performance of some methods that are employed in the classical regression to their spatial counterparts. Though the methods work well in the classical regression models, they are mostly prone to either masking or swamping in spatial applications. This is attributable to the local nature of spatial outliers. Interestingly, measures that adopt studentized residuals and leverage values (through potentials), $ISRs - P_{osi}$ and $ESRs - P_{osi}$, that contain neighbourhood information of the data points do well in classifying the influential observations as demonstrated by both simulation result and examples.

Author Contributions: Conceptualization, A.M.B, H.M.; methodology, A.M.B; software, A.M.B; validation, A.M.B, H.M. ; formal analysis, A.M.B, N.H.A; investigation, A.M.B, N.H.A; resources, A.M.B, H.M., M.B.A; data curation, A.M.B; writing--original draft preparation, A.M.B; writing--review and editing, A.M.B, H.M., M.B.A, N.H.A; visualization, A.M.B, M.B.A; supervision, H.M.; project administration, H.M., funding acquisition, H.M..

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Belsley, D.A.; Kuh, E.; Welsch, R.E. *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*; John Wiley & Sons, 1980; Vol. 571;.
2. Rashid, A.M.; Midi, H.; Slwabi, W.D.; Arasan, J. An Efficient Estimation and Classification Methods for High Dimensional Data Using Robust Iteratively Reweighted SIMPLS Algorithm Based on *Nu* -Support Vector Regression. *IEEE Access* **2021**, *9*, 45955–45967, doi:10.1109/ACCESS.2021.3066172.
3. Cook, R.D. Influential Observations in Linear Regression. *Journal of the American Statistical Association* **1977**, *74*, 169–174.
4. Hoaglin, D.C.; Welsch, R.E. The Hat Matrix in Regression and ANOVA. *The American Statistician* **1978**, *32*, 17, doi:10.2307/2683469.
5. Andrews, D.F.; Pregibon, D. Finding the Outliers That Matter. *Journal of the Royal Statistical Society: Series B (Methodological)* **1978**, *40*, 85–93.
6. Hawkins, D.M. *Identification of Outliers*; Springer, 1980; Vol. 11;.
7. Huber, P. Robust Statistics. New York: John Wiley and Sons. *HuberRobust statistics1981* **1981**.
8. Cook, R.D.; Weisberg, S. *Residuals and Influence in Regression*; Monographs on statistics and applied probability; Chapman and Hall: New York, 1982; ISBN 978-0-412-24280-9.
9. Chatterjee, S.; Hadi, A.S. *Sensitivity Analysis in Linear Regression*; John Wiley & Sons, 1988; Vol. 327;.
10. Hadi, A.S. A New Measure of Overall Potential Influence in Linear Regression. *Computational Statistics & Data Analysis* **1992**, *14*, 1–27.
11. Habshah, M.; Norazan, M.R.; Rahmatullah Imon, A.H.M. The Performance of Diagnostic-Robust Generalized Potentials for the Identification of Multiple High Leverage Points in Linear Regression. *Journal of Applied Statistics* **2009**, *36*, 507–520, doi:10.1080/02664760802553463.
12. Meloun, M.; Militkỳ, J. *Statistical Data Analysis: A Practical Guide*; Woodhead Publishing Limited, 2011;
13. Puterman, M.L. Leverage and Influence in Autocorrelated Regression Models. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **1988**, *37*, 76–86.
14. Schall, R.; Dunne, T.T. A Unified Approach to Outliers in the General Linear Model. *Sankhyā: The Indian Journal of Statistics, Series B* **1988**, 157–167.
15. Martin, R.J. Leverage, Influence and Residuals in Regression Models When Observations Are Correlated. *Communications in statistics-theory and methods* **1992**, *21*, 1183–1212.
16. Shi, L.; Chen, G. Influence Measures for General Linear Models with Correlated Errors. *The American Statistician* **2009**, *63*, 40–42.
17. Christensen, R.; Johnson, W.; Pearson, L.M. Prediction Diagnostics for Spatial Linear Models. *Biometrika* **1992**, *79*, 583–591, doi:10.1093/biomet/79.3.583.
18. Haining, R. DIAGNOSTICS FOR REGRESSION MODELING IN SPATIAL ECONOMETRICS*. *J Regional Sci* **1994**, *34*, 325–341, doi:10.1111/j.1467-9787.1994.tb00870.x.
19. Haining, R.P.; Haining, R. *Spatial Data Analysis: Theory and Practice*; Cambridge University Press, 2003;
20. Dai, X.; Jin, L.; Shi, A.; Shi, L. Outlier Detection and Accommodation in General Spatial Models. *Stat Methods Appl* **2016**, *25*, 453–475, doi:10.1007/s10260-015-0348-1.
21. Midi, H.; Mohammed, A. The Identification of Good and Bad High Leverage Points in Multiple Linear Regression Model. *Mathematical Methods and System in Science and Engineering* **2015**, 147–158.
22. Anselin, L. *Spatial Econometrics: Methods and Models*; Studies in Operational Regional Science; Springer Netherlands: Dordrecht, 1988; Vol. 4; ISBN 978-90-481-8311-1.
23. LeSage, J.P. The Theory and Practice of Spatial Econometrics. *University of Toledo. Toledo, Ohio* **1999**, *28*.
24. Olver, P.J.; Shakiban, C.; Shakiban, C. *Applied Linear Algebra*; Springer, 2006; Vol. 1;.
25. Horn, R.A.; Johnson, C.R. *Matrix Analysis*; 2nd ed.; Cambridge University Press: Cambridge ; New York, 2012; ISBN 978-0-521-83940-2.
26. Liesen, J.; Mehrmann, V. *Linear Algebra*; Springer Undergraduate Mathematics Series; Springer International Publishing: Cham, 2015; ISBN 978-3-319-24344-3.
27. Shekhar, S.; Lu, C.-T.; Zhang, P. A Unified Approach to Detecting Spatial Outliers. *GeoInformatica* **2003**, *7*, 139–166.

28. Billor, N.; Hadi, A.S.; Velleman, P.F. BACON: Blocked Adaptive Computationally Efficient Outlier Nominators. *Computational statistics & data analysis* **2000**, *34*, 279–298.
29. Imon, A. Identifying Multiple High Leverage Points in Linear Regression. *Journal of Statistical Studies* **2002**, *3*, 207–218.
30. Kou, Y.; Lu, C.-T. Outlier Detection, Spatial. *Encyclopedia of GIS* **2008**, 1539–1546.
31. Rahmatullah Imon, A.H.M. Identifying Multiple Influential Observations in Linear Regression. *Journal of Applied Statistics* **2005**, *32*, 929–946, doi:10.1080/02664760500163599.
32. Alguraibawi, M.; Midi, H.; Imon, A.H.M.R. A New Robust Diagnostic Plot for Classifying Good and Bad High Leverage Points in a Multiple Linear Regression Model. *Mathematical Problems in Engineering* **2015**, *2015*, 1–12, doi:10.1155/2015/279472.
33. Bagheri, A.; Midi, H. Diagnostic Plot for the Identification of High Leverage Collinearity-Influential Observations. 20.
34. Dowd, P. The variogram and kriging: robust and resistant estimators. In *Geostatistics for natural resources characterization*; Springer, 1984; pp. 91–106.