symmetry

MDPI

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, Serdang 43400, Selangor, Malaysia; mbali@atbu.edu.ng (A.M.B.); bakri@upm.edu.my (M.B.A.); nurhaizum@upm.edu.my (N.H.A.R.)
[2] Department of Mathematical Sciences, Abubakar Tafawa Balewa University, Bauchi 0248, Nigeria
[3] Department of Mathematics and Statistics, Faculty of Science, Universiti Putra Malaysia, Serdang 43400, Selangor, Malaysia
* Correspondence: habshah@upm.edu.my

Abstract: 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 $ISRs - P_{osi}$ and $ESRs - P_{osi}$. 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, H_i^2 (using robust and non-robust measures), and our proposed $ISRs - P_{osi}$ and $ESRs - P_{osi}$ plots. Results of our simulation study and applications to real data showed that the Cook's distance, non-robust H_{si1}^2 and robust H_{si2}^2 were not very successful in detecting IOs. The H_{si1}^2 suffered from the masking effect, and the robust H_{si2}^2 suffered from swamping in general spatial models. Interestingly, the results showed that the proposed $ESRs - P_{osi}$ plot, followed by the $ISRs - P_{osi}$ plot, was very successful in classifying observations into the correct groups, hence correctly detecting the real IOs.

Keywords: spatial regression model; influential observation; outlier; leverage; prediction residual; masking and swamping; diagnostic

Citation: Baba, A.M.; Midi, H.; Adam, M.B.; Rahman, N.H. A. Detection of Influential Observations in Spatial Regression Model Based on Outliers and Bad Leverage Classification. *Symmetry* **2021**, *13*, 2030. https://doi.org/10.3390/sym13112030

Academic Editor: Jinyu Li

Received: 6 August 2021
Accepted: 22 October 2021
Published: 27 October 2021

1. Introduction

Belsley et al. [1] defined an influential observation (IO) as one which, either individually or together with several other observations, has a 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 (GLPs) and bad leverage points (BLPs). 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 misleading conclusions about the fitted regression models and various other estimates. Once the IOs are identified, there is a need to study their impact on the model and subsequent analyses. There is a handful

of studies on the 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., [13–17]. However, only a few articles deal with the detection of IOs in spatial regression models; some examples include [18–22]. Some robust estimation methods in spatial regression are [23–25]. Christensen et al. [18] and Haining [19] adapted one of the diagnostic measures in [3] to detect influential observations in spatial error autoregression model. They achieved this by defining correlated errors through the spatial weight matrix and coefficient of spatial autocorrelation in the error term. They also presented the spatial Studentized prediction residuals and the spatial leverage terms that contain error terms in spatial information.

The presence of high or low attribute value in the neighbourhood of a spatial location may result in the inability to detect the true spatial outlier, or the false identification of a good observation as an outlier [26]. Hadi [27] has also noted that spatial outlier detection methods inherit the problem of masking and swamping. Masking occurs when outlying observations are incorrectly declared as inliers. Swamping on the other hand, occurs when clean observations are incorrectly classified as outliers [28]. Aggarwal [29] observed that spatial outlier breaks the spatial autocorrelation and continuity of spatial locations. Spatial autocorrelation is a systematic pattern in attribute values that are recorded in several locations on a map. Attribute values in one location that are associated with values at neighbouring locations indicate the presence of autocorrelation. Positive autocorrelation indicates similar values that are clustered together. Negative autocorrelation indicates low attribute values in the neighbourhood of high attribute values and vice-versa [30].

Robust estimation methods mostly focus on estimations that are not influenced much by the effects of outliers. Anselin [23] has extended the bootstrap estimation to mixed regressive spatial autoregressive models, where pseudo error terms are generated by sampling from the vector of error terms. The spatial structure of the data is maintained through the generation of error terms. Politis et al. [31] and Heagerty and Lumley [32] also adopted the bootstrap method on blocks of contiguous locations to generate replicates of the estimates of the asymptotic standard error of statistics. Cerioli and Riana [24] argued that a robust estimator of the spatial autocorrelation parameters did not exist based on all datasets. They proposed a forward search algorithm based on blocks of contagious spatial locations (BFS). The BFS algorithm are drawn in such a way that the blocks retain the spatial dependence structure of the original data. Yildirim [25] proposed a robust estimation method of the log-likelihood with influence function in the spatial error model. This is achieved iteratively using scoring algorithm to estimate the parameters. Though they succeeded in obtaining robust estimates, identifying spatial outliers, which is vital in spatial statistics [26], was not achieved. Popular graphical techniques to detect spatial outliers are the scatterplot [33], the Moran's scatterplot [30] and the pockets of nonstationarity [34]. Besides being prone to the problem of masking and swamping [26], they focused mainly on spatial outliers in the Y- space only.

Diagnostic works on models that have both spatial autocorrelations in dependent variable and residual terms are missing in the literature. The problem of masking and swamping is prevalent in spatial regression model diagnostics, which may be due to the presence of vertical outliers as well as leverage points, as in the case of linear regression ([27]). This motivates us to represent the spatial Studentized prediction residuals and spatial leverage values in the general spatial model, and to adapt and extend some robust diagnostic measures of detection of outliers and IOs in linear regression, such as Hadi's potential (p_{oii}), Cook's distance (CD_i) [3], the overall potential influence (H_i^2) [10], and the external (ESRs) and internal (ISRs) 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 propose a robust spatial diagnostic plot and adapt some diagnostic measures in the linear regression model. Representations of the diagnostic measures in the spatial regression model are obtained, with a special emphasis on the general spatial regression model (GSM) that performs autoregression on both the dependent variable and error terms.

The main objective of this study is to propose a robust spatial diagnostic plot. Other objectives are: (1) to represent the leverage values of the hat matrix of the linear regression in the GSM model; (2) to extend the ISR of the linear regression to the GSM model; (3) to extend the *ESR* of the linear regression to the GSM model; (4) to extend the Cook's distance and the overall potential influence of the linear regression to the GSM model (5) to develop a method of identification of the influential observations of the GSM model by proposing a procedure of classification of the observations into regular observations, vertical outliers, good and bad leverage points, and hence IOs; (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, COVID-19 data in Georgia, USA, and the life expectancy data from USA counties. The significance of this study is that it can contribute to the development of a method of identification of influential observations in spatial regression models.

2. Identification of Influential Observations in a Linear Regression Model

Consider a *k*-variable regression model:

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon} \tag{1}$$

where $\boldsymbol{y}$ is an $n \times 1$ vector of observations of dependent variables, $\boldsymbol{X}$ is an $n \times k$ matrix of independent variables, $\boldsymbol{\beta}$ is a $k \times 1$ vector of unknown regression parameters, $\boldsymbol{\varepsilon}$ is an $n \times 1$ vector of random errors with identical normal distributions, that is, $\boldsymbol{\varepsilon} \sim NID(0, \sigma^2)$.

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

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{y} \tag{2}$$

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

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

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

Hoaglin and Welsh [3] suggested that an observation which exceeds $\frac{2k}{n}$, where $\frac{2k}{n}$ is the average value of p_{ii}, is considered as a leverage point, while Vellman and Welsch suggested $\frac{3k}{n}$ as a cut-off point 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. As a result, 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 element of a potential denoted as p_{oii}, is given by:

$$p_{0ii} = \boldsymbol{x}_i^T\left(\boldsymbol{X}_{(i)}^T\boldsymbol{X}_{(i)}\right)\boldsymbol{x}_i, \quad i = 1,2,\cdots,n \tag{3}$$

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:

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

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

$$\boldsymbol{r} = \boldsymbol{y} - \hat{\boldsymbol{y}} = (\boldsymbol{I} - \boldsymbol{P})\boldsymbol{y} = \boldsymbol{Q}\boldsymbol{y},$$

The Studentized residuals (internally Studentized residuals) denoted as ISRs and R-Student residuals (externally Studentized residuals) denoted as ESRs are widely used measures for the identification of outliers (see [7]). The *ISR*, denoted as t_i, is defined as:

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

where $\hat{\sigma}$ is the standard deviation of the residuals, r_i and p_{ii} are the 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 t_i^* and given by:

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

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 the influence on the regression coefficient estimate or the predicted values. The Cook's distance is given by

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

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

$$\widehat{CD}_i(\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{X}, k\hat{\sigma}^2) = \frac{(\hat{\boldsymbol{y}} - \hat{\boldsymbol{y}}_{\boldsymbol{i}})^{\boldsymbol{T}}(\hat{\boldsymbol{y}} - \hat{\boldsymbol{y}}_{\boldsymbol{i}})}{k\hat{\sigma}^2} \tag{5}$$

Computing the $\widehat{CD}_i(\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{X}, k\hat{\sigma}^2)$ 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]):

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

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 the 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 drawback 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 methods 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 $\boldsymbol{H_i^2}$, and defined as follows:

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

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

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

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

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

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

$$l_1 = mean(p_{oi}) + c\sqrt{Var(p_{oi})}$$
$$= \frac{k}{n} + c\sqrt{\frac{ns - k^2}{n(n-1)}},$$

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

$$\boldsymbol{l_2 = Med(p_{oi}) + cMAD(p_{oi})},$$

3. Influential Observations in Spatial Regression Models

The general spatial autoregressive model (GSM) ([21,35,36]) includes the spatial lag term and spatially correlated error structure. The data generating process (DGP) of the general spatial model is given by:

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

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

The general spatial autoregressive model in Equation (9) can be rewritten as:

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

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

The log-likelihood function (L) is given by:

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

Let $\hat{\rho}, \hat{\lambda}, \hat{\sigma}^2, \widehat{\boldsymbol{\beta}}$ be the maximum likelihood estimates (MLEs) of $\rho, \lambda, \sigma^2, \boldsymbol{\beta}$, respectively. The MLEs are obtained iteratively using numerical methods in the maximum likelihood estimation. Anselin [35] and LeSage [36] 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 $\boldsymbol{\beta_{ay}}$, $\widehat{\boldsymbol{\beta}}_{\boldsymbol{ay}}$, is given by:

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

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

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

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

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

Let $\boldsymbol{Q_{ay}} = \boldsymbol{I_n} - \boldsymbol{P_{ay}}$. Though $\boldsymbol{P_{ay}}$ and $\boldsymbol{Q_{ay}}$ have satisfied the idempotence property and their sum of diagonal elements equals k and $n - k$, respectively, they are not symmetric. As a result, they are not positive semi-definite, and as such, the diagonal elements of $\boldsymbol{P_{ay}}$ will have negative values. The hat matrices $\boldsymbol{P_{ay}}$ and $\boldsymbol{Q_{ay}}$ 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 $\boldsymbol{P^*_{ay}}$, such that:

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

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

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

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

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 $\boldsymbol{z^* = BAy}$. Schall and Dunne [14] defined the matrix $\boldsymbol{V^{-1}}$ as a singular value decomposition such that $\boldsymbol{V^{-1} = B\Delta B^T}$; where $\boldsymbol{B}$ is of the same order as $\boldsymbol{V^{-1}}$ and $\boldsymbol{\Delta}$ is a diagonal matrix. The transformation $\boldsymbol{z^*}$ is the principal component score. Puterman [13] and Haining [19] defined it as canonical variates such that $\boldsymbol{BX(X^TV^{-1}X)X^TB^T}$ is positive semi-definite. By setting $\boldsymbol{z^* = BAy}$, Equation (9) is rewritten in a generalized least squares (GLS) form as:

$$\boldsymbol{z^* = X^*\beta_s + \varepsilon}, \boldsymbol{\varepsilon} \sim N(0, \sigma^2\boldsymbol{I_n}) \tag{12}$$

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

The estimate $\boldsymbol{\widehat{\beta}_s}$ of $\boldsymbol{\beta_s}$ is now given by:

$$\boldsymbol{\widehat{\beta}_s} = \left(\boldsymbol{X^{*T}X^*}\right)^{-1}\boldsymbol{X^*}^{\boldsymbol{T}}\boldsymbol{z^*}$$

Thus,

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

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

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

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

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

The properties of the leverage in the transformed spatial model in Equation (13) are:
Property I: idempotent and symmetric.
Property Ia: idempotence

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

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

$$= \boldsymbol{P_s}$$

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

Property Ib: symmetric

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

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

$$= \boldsymbol{P_s}$$

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

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

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

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

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

$$= trace(I_k)$$

$$= k,$$

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

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

Property III: bounds on the spatial leverage.

The bound on the leverage of the classical regression is $0 \leq p_{ii} \leq 1$ 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 [37–39]. The transformation in Equation (11) yields 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:

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

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

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

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

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 regressions 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 the model residual; model (9) (GSM) combines both features.

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., $\lambda = 0$. Thus, the SAR model is given by:

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

The $\boldsymbol{P_s}$ corresponding to the model in Equation (13) reduces to:

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

with the transformation in Equation (11) simplifying to $\boldsymbol{z^*} = \boldsymbol{Ay}$, since $\boldsymbol{V} = (\boldsymbol{B^T B})^{-1}$ and $\boldsymbol{B} = \boldsymbol{I_n}$, when $\lambda = 0$. 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., $\rho = 0$. This yields the model:

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

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

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

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

Works on spatial regression diagnostics in the literature mainly focus on the autocorrelation in the residuals, mostly using a time series analogy [13–15]. Some remarkable works on the spatial regression model can be found in [18,19,21].

3.2. Influential Observations in Spatial Regression Model

The leverages $\boldsymbol{P_s}$ and $\boldsymbol{Q_s}$ in Equation (11) satisfy all the properties of a projection matrix, including that the sum of the diagonal terms of $\boldsymbol{P_s}$ and $\boldsymbol{Q_s}$ equal $\boldsymbol{k}$ and $n - k$, respectively. It also incorporates the autocorrelation in the dependent variables, $\boldsymbol{Wy}$. Hence, it can be used as a diagnostic measure of leverage points in a 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 CD_{si}, can be formulated as follows:

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

$$= \frac{\left(\widehat{\boldsymbol{\beta}}_{\boldsymbol{s}}^{(-i)} - \widehat{\boldsymbol{\beta}}_{\boldsymbol{s}}\right)^{\boldsymbol{T}} \left((\boldsymbol{BX})^{\boldsymbol{T}} (\boldsymbol{BX})\right) \left(\widehat{\boldsymbol{\beta}}_{\boldsymbol{s}}^{(-i)} - \widehat{\boldsymbol{\beta}}_{\boldsymbol{s}}\right)}{\boldsymbol{k\hat{\sigma}^2}}$$

$$= \frac{\left(\widehat{\boldsymbol{\beta}}_{\boldsymbol{s}}^{(-i)} - \widehat{\boldsymbol{\beta}}_{\boldsymbol{s}}\right)^{\boldsymbol{T}} (\boldsymbol{X^T B^T BX}) \left(\widehat{\boldsymbol{\beta}}_{\boldsymbol{s}}^{(-i)} - \widehat{\boldsymbol{\beta}}_{\boldsymbol{s}}\right)}{\boldsymbol{k\hat{\sigma}^2}}$$

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

where:

$$\widehat{\boldsymbol{\beta}}_{\boldsymbol{s}}^{(-\boldsymbol{i})} = \boldsymbol{X}_{(\boldsymbol{i})}\left(\boldsymbol{X}_{(\boldsymbol{i})}^{\boldsymbol{T}}\widehat{\boldsymbol{V}}_{(\boldsymbol{i},\boldsymbol{i})}^{-\boldsymbol{1}}\boldsymbol{X}_{(\boldsymbol{i})}\right)^{-\boldsymbol{1}}\boldsymbol{X}_{(\boldsymbol{i})}^{\boldsymbol{T}}\widehat{\boldsymbol{V}}_{(\boldsymbol{i},\boldsymbol{i})}^{-\boldsymbol{1}}\widehat{\boldsymbol{A}}_{(\boldsymbol{i},\boldsymbol{i})}\boldsymbol{Y}_{(\boldsymbol{i})}.$$

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

The spatial Cook distance, CD_{si}, is declared large if $CD_{si} > 0.70$ [19]. In its simplified form, the Cook's distance in spatial regression is written as

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

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 $\boldsymbol{P_s}$, 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 $\boldsymbol{b_i}$ and $\boldsymbol{a_i}$ are the i^{th} columns of matrices $\boldsymbol{B}$ and $\boldsymbol{A}$, respectively. The spatial Studentized residual has a cut-off point of 2 to declare a point large [19,40].

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

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

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

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

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

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

We define the spatial measure of overall potential influence as

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

When measuring the influence of an observation in a 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 is deleted in the linear regression and admitted into the model if it is proven clean (BACON [41], [42], DGRP [11]). This is because IOs in linear regression are global in nature; however, in a spatial regression model, IOs are local. Haining [20] 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 spatiotemporal statistics usually carry vital information in applications. Kou et al. [26] further pointed out that detecting spatial outliers can help in locating extreme meteorological events such as tornadoes and hurricanes, identify aberrant genes or tumour 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 a contiguous set of points to reveal the unusual features related to that neighbourhood.

Although methods that detect multiple outliers in spatial regression work well (see [21]), we refer to methods that group observations as clean or suspect, irrespective of their positions (with reference to spatial data), and admit them into the model as clean observations according to some conditions.

According to Hadi [10], examining each value of influence measure alone, such as $\boldsymbol{P_{si}}$, ISRs, ESRs, CD_{si} and H_{si}^2, might not be successful to indicate the IOs or the source of influence. Imon [43] and Mohammed [44] 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 [43,45], we adopt their rules for the 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 the local nature of spatial IOs, we have to make some modifications to the classification scheme. In this paper, a new diagnostic plot is proposed by plotting the ISRs (or ESRs) on the Y-axis against the spatial potential, $\boldsymbol{P_{osi}}$, on the X-axis. We consider the ISRs and ESRs because both measures contain spatial information. On the other hand, the potentials that are obtained from the transformed model in Equation (13) are considered in order to reflect spatial dependence. Hence, the proposed diagnostic plots are denoted as $ISRs - P_{osi}$ and $ESRs - P_{osi}$ plot, and they are based on the following classification schemes:

(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 $p_{osi} > l_2$.
iii. i^{th} observation is BLP if $|ISRs| > 2.0$ and $p_{osi} > l_2$.
iv. i^{th} observation is IO if $|ISRs| \geq 2.0$ and $p_{osi} \leq l_2$.

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

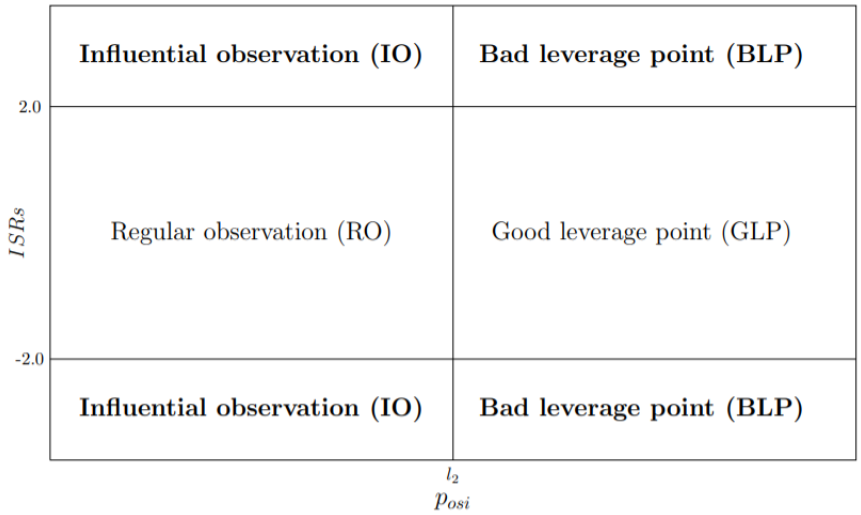


Figure 1. Classification of RO, GLP, and IO according $ISRs - P_{osi}$.

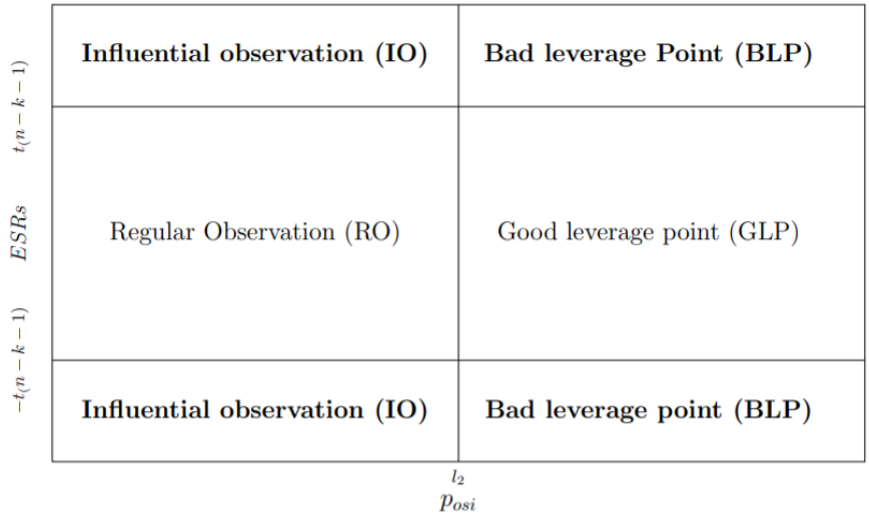


Figure 2. Classification of RO, GLP, and IO according $ESRs - P_{osi}$.

(b) $ESRs - P_{osi}$

i. i^{th} observation is declared RO if $|ESRs| < t_{n-k-1}$ and $p_{osi} < l_2$.
ii. i^{th} observation is GL if $|ESRs| < t_{n-k-1}$ and $p_{osi} > l_2$.
iii. i^{th} observation is IO if $|ESRs| > t_{n-k-1}$ and $p_{si} > l_2$.
iv. i^{th} observation is IO if $|ESRs| \geq t_{n-k-1}$ and $p_{si} \leq l_2$.

4. Results and Discussions

In this section, the performance of all the proposed methods, i.e., the Cook's Distance ($\widehat{CD}_{si}$), H_{si}^2(H_{si1}^2(non-robust) and H_{si2}^2(robust)), $ISRs - P_{osi}$ and $ESRs - P_{osi}$, is evaluated using a simulation study, artificial data and real datasets of gasoline price data in the southwest area of Sheffield, UK, COVID-19 data in the counties of the State of Georgia, USA and the life expectancy data in counties of the USA.

Simulated Data

We simulated the spatial regression model in Equation (9) for a square spatial grid with sample size, $n = 400$, $\rho = 0.4$, $\lambda = 0.5$ and $\boldsymbol{W_1} = \boldsymbol{W_2}$, using row-standardized Queen's contiguity spatial weights. $\boldsymbol{x_0} = \boldsymbol{1}$, $\boldsymbol{x_1} \sim N(0,1)$, $\boldsymbol{\beta_0} = \boldsymbol{0}$, $\boldsymbol{\beta_1} = \boldsymbol{1}$ (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 $\boldsymbol{X}$ and $\boldsymbol{y}$ directions. The contamination in the $\boldsymbol{y}$ direction is taken from the Cauchy distribution because of its fat tails. Contamination in the $\boldsymbol{X}$ direction is taken from the following multivariate distribution,

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

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 [46], and spatial simulation is conditioned to a real dataset.

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.

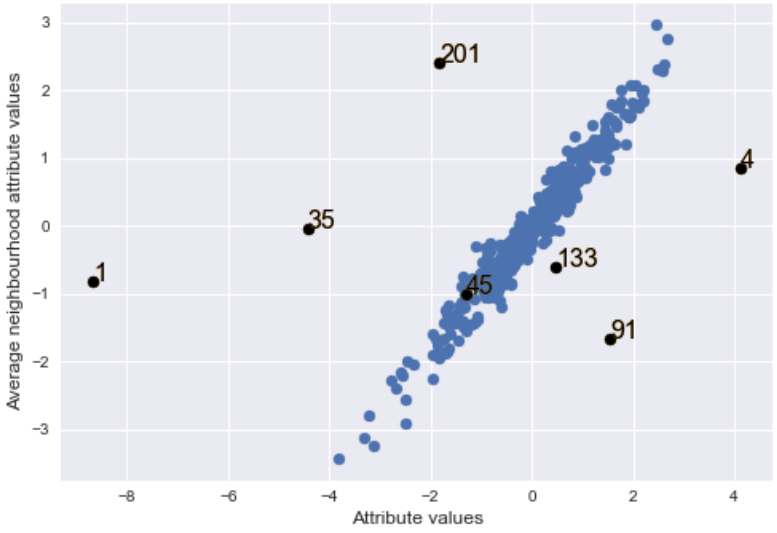


Figure 3. Graph of average attribute values in neighbourhood of locations against the attribute values in the locations with contamination (black points).

Table 1 presents the values of ISRs, ESRs and $\boldsymbol{p_{osi}}$, where values in parentheses 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 fifty-four potentials correspond to Studentized residuals greater than the thresholds of ISRs and ESRs (locations 51 and 201).

Table 1. ISRs, ESRs and $\boldsymbol{p_{osi}}$ of locations with large Studentized residuals in the simulated GSM model, with their cut-off points in parentheses.

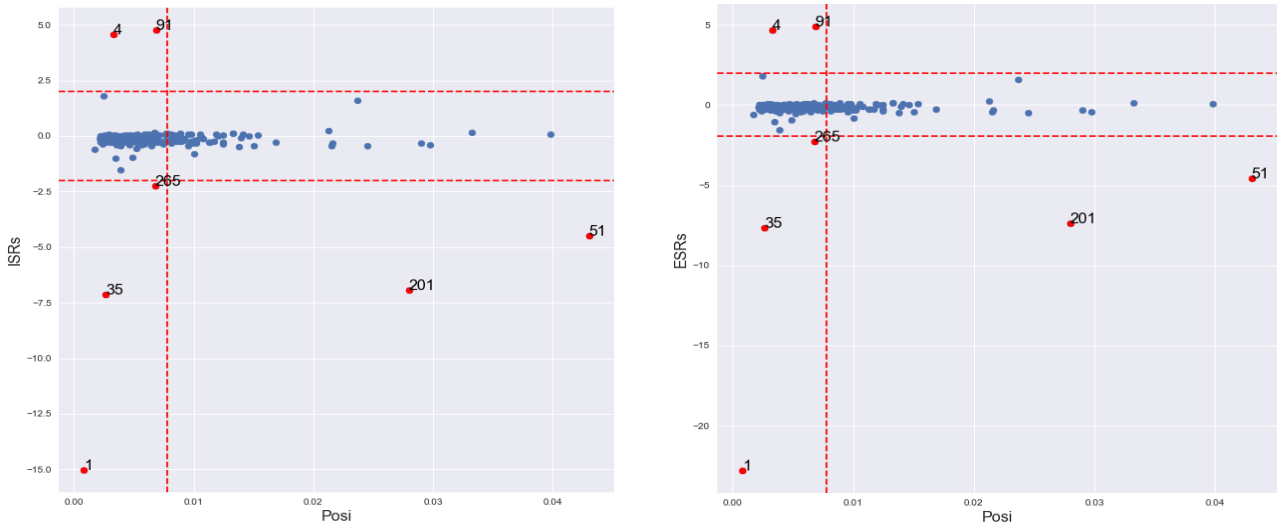
Location	$\boldsymbol{ISRs}$ (2.00)	$\boldsymbol{ESRs}$ (1.97)	$\boldsymbol{p_{osi}}$ (0.0078)
1	15.0378	22.8179	0.0008
4	4.5847	4.7046	0.0033
35	−7.1434	−7.6397	0.0026
51	−4.4695	−4.5801	0.0430
91	4.7613	4.8965	0.0068
201	−6.9336	−7.3840	0.0280
265	−2.2644	−2.2762	0.0068

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 residuals and MAD_i is the median absolute deviation. The absolute value of the Studentized residuals is compared to the neighbourhood threshold for confirmation as an 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 on Figure 4, $ISRs - P_{osi}$ and $ESRs - P_{osi}$ classified locations 1, 4, 35, 91 and 265 as outliers in the 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's t-distribution, which is around 1.96 for large sample sizes.



(**a**) (**b**)

Figure 4. Graph of IO classification according to GLP, BLP and vertical outlier in $ISRs - P_{osi}$ and $ESRs - P_{osi}$ for simulated data. (**a**) $ISRs - P_{osi}$; (**b**) $ESRs - P_{osi}$.

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_{si2}^2 identified 60 locations as IOs, including all the locations classified by the other methods. However, a 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 due the 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 influential observations with consistent swamping rates of 0%. The $ISRs - P_{osi}$ demonstrated a high detection to the tune of 98% while $ESRs - P_{osi}$ had 100% accurate classification of the IOs, both with swamping rates of 0%. H_{si1}^2 had less than 40% accurate classification with zero swamping rate, while the H_{si2}^2 had 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

5. Illustrative Examples

5.1. Example 1

The gasoline price data for 61 retail sites in the southwest area of Sheffield from [19] were used in Example 1. The analysis indicated 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):

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

where, $\boldsymbol{y}_M$ and $\boldsymbol{X}_M$ are the March and February sales from the southwest Sheffield gasoline sale data, respectively, $\hat{\lambda} = 0.15$ is the estimate of coefficient of correlation in the residuals, $\boldsymbol{W}$ is the standardized weight matrix and $\boldsymbol{\xi}$ is the vector of correlated residuals.

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 southwest Sheffield gasoline data.

S/N	Site	ISRs (2.00)	ESRs (2.00)	$\boldsymbol{p_{osi}}$ $(\boldsymbol{0.0335})$	$\boldsymbol{CD_{si}}$	$\boldsymbol{ISRs - P_{osi}}$	$\boldsymbol{ESRs - P_{osi}}$	$\boldsymbol{H_{si1}^2}$	$\boldsymbol{H_{si2}^2}$
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

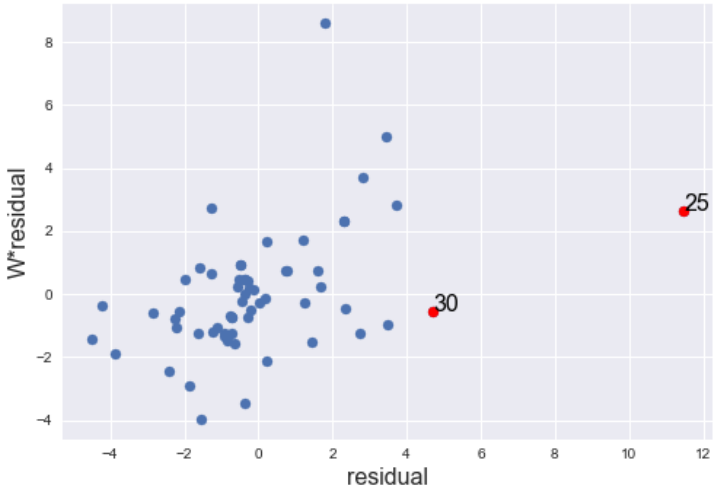
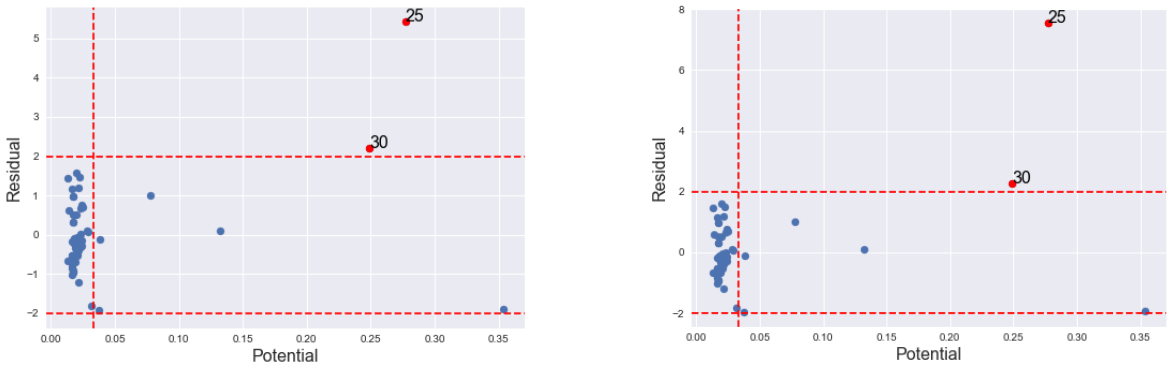


Figure 5. Graph of the lagged residuals against the residuals, of the 61 sites of the southwest Sheffield fitted with SEM, showing the IO points in red dots.

The CD_{si}, $ISRs - P_{osi}$, $ESRs - P_{osi}$, and H_{si1}^2 coincidentally identified site 25 only as IO. H_{si2}^2 detects 11 more sites as IOs in addition to site 25. Haining [19] 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 as an IO.

Though the H_{si2}^2 has detected all the suspected IOs, it is 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.

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



(**a**) (**b**)

Figure 6. Graph of IO classification according to GLP, BLP and vertical outlier in South West gasoline data. (**a**) $ISRs - P_{osi}$; (**b**) $ESRs - P_{osi}$.

5.2. Example 2

The data for example 2 were the COVID-19 data for the 159 counties of the State of Georgia, USA, as of 30 June 2020 (http://dph.georgia.gov/covid-19-daily-status-report; accessed on 30 June 2020) and the health ranking (http://www.countyheathrankings.org; accessed on 30 June 2020). The case-rate per 100,000 of COVID-19 was the dependent variable. The independent variables were the population of black race in the county ($\boldsymbol{X}_1$), population of Asians ($\boldsymbol{X}_2$), population of Hispanic ($\boldsymbol{X}_3$), population of people that are 65 years and above ($\boldsymbol{X}_4$), population of female in the county ($\boldsymbol{X}_5$) and life expectancy ($\boldsymbol{X}_6$).

The model was fitted with the SAR model (model with the lowest Akaike information criteria (AIC) value of 2192). The SAR model is presented in Equation (22):

$$\hat{\boldsymbol{y}} = \hat{\rho}\boldsymbol{W}\hat{\boldsymbol{y}} + \hat{\boldsymbol{\beta}}_0 + \sum_{i=1}^{6} \hat{\boldsymbol{\beta}}_i \boldsymbol{X}_i \quad (22)$$

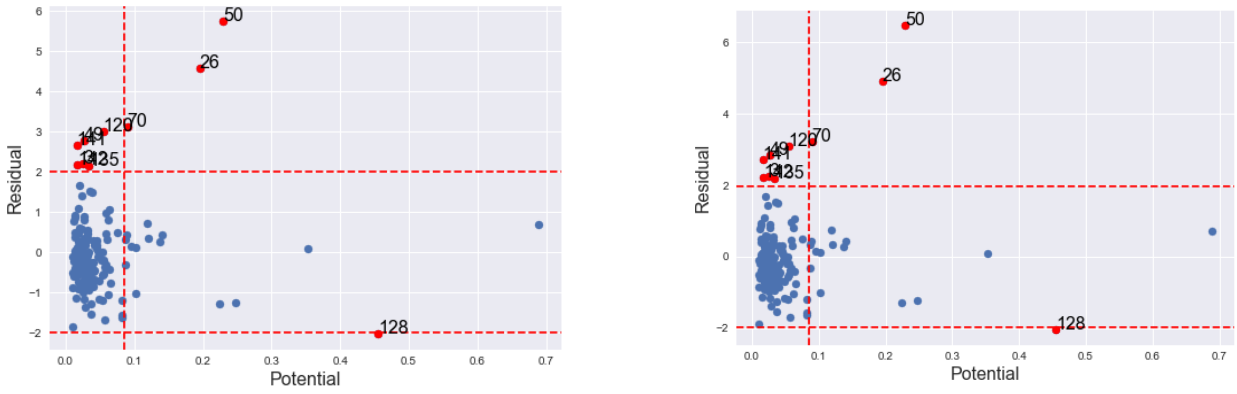
where $\hat{\rho} = 0.6967$, $\hat{\boldsymbol{\beta}}_0 = 1087.7388$, $\hat{\boldsymbol{\beta}}_1 = 9.7831$, $\hat{\boldsymbol{\beta}}_2 = -6.2210$, $\hat{\boldsymbol{\beta}}_3 = -54.1402$, $\hat{\boldsymbol{\beta}}_4 = -28.5874$, $\hat{\boldsymbol{\beta}}_5 = 4.8288$ and $\hat{\boldsymbol{\beta}}_6 = 40.3323$. $\boldsymbol{X}_1$, $\boldsymbol{X}_3$ 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 only classified county 50 as an IO. The $ISRs - P_{osi}$ and $ESRs - P_{osi}$ coincided in detecting counties 3, 26, 49, 50, 70, 120, 135, 141 and 142 as IOs. The H_{si1}^2 (non-robust) detected 26 and 50 as IOs. The H_{si2}^2 (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 font.

Table 4. Detected IOs counties by different methods in the Georgia COVID-19 data.

County	ISRs (2.00)	ESRs (1.98)	$\boldsymbol{p_{si}}$ $(\boldsymbol{0.0851})$	$\boldsymbol{CD_{si}}$	$\boldsymbol{ISRs - P_{osi}}$	$\boldsymbol{ESRs - P_{osi}}$	$\boldsymbol{H_{si1}^2}$	$\boldsymbol{H_{si2}^2}$
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

The IOs identified by $ISRs - P_{osi}$ and $ESRs - P_{osi}$ both have large Studentized residuals and large potentials as can be observed in Table 4. Figure 7 shows the outliers in X, y and both X and y directions. The CD_{si} detected the largest Studentized residual with a high potential as IO. The H_{si1}^2 identified two observations with large Studentized values and high potential values. The H_{si2}^2 detected all suspected IOs, but with many having both small values of Studentized residuals and potential values.



(**a**) (**b**)

Figure 7. Graph of IO classification according to GLP, BLP and vertical outlier in the State of Georgia, USA COVID-19 data. (**a**) $ISRs - P_{osi}$; (**b**) $ESRs - P_{osi}$.

While examining the outlyingness of the classified counties, we find that county 50 is clearly an IO since it has both large Studentized residual and a large potential value. It is outside the threshold value of its neighbourhood.

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 potential values and Studentized values greater than their threshold values and are classified as BLPs and hence IOs.

Besides the counties classified by $ISRs - P_{osi}$ and $ESRs - P_{osi}$, 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 was measured by population density ($\boldsymbol{X}_1$), fair/poor health status ($\boldsymbol{X}_2$), obesity ($\boldsymbol{X}_3$), population in rural area ($\boldsymbol{X}_4$), inactivity rate ($\boldsymbol{X}_5$), population of smokers ($\boldsymbol{X}_6$), population of black people ($\boldsymbol{X}_7$), population of Asians ($\boldsymbol{X}_8$) and population of Hawaiians (X_9). The data were obtained from the Kaggle website (https://www.kaggle.com/johnjdavisiv/us-counties-covid19-weather-sociohealth-data; accessed on 13 December 2020).

The spatial error model (SEM) had the lowest AIC value and was fitted to the data. The model was significant at the 5% level with a significant Moran's I of 0.2160. $\boldsymbol{X}_1$ and $\boldsymbol{X}_4$ were not significant at the 5%. All the other estimates were significant at the 5% level. The fitted model is given by:

$$\hat{\boldsymbol{y}} = \hat{\boldsymbol{\beta}}_0 + \sum_{i=1}^{9} \hat{\boldsymbol{\beta}}_i \boldsymbol{X}_i + \hat{\lambda} \boldsymbol{W} \boldsymbol{\xi}$$

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

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

6. Conclusions

In this article, we demonstrated the application of influential observations (IOs) detection techniques from the classical regression to the spatial regression model. Measures that contained spatial information in the spatial autoregression in the dependent variables and residuals were obtained. We also evaluated the performance of some methods employed in classical regression to their spatial counterparts. Though the methods work well in 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. Hence, we proposed new $ISRs - P_{osi}$ and $ESRs - P_{osi}$ plots to classify observations into four categories: regular observations, vertical outliers, good leverage points and bad leverage points, whereby IOs are those observations which fall in the vertical and bad leverage point categories. Interestingly, the proposed $ESRs - P_{osi}$ diagnostic plot was very successful in classifying observations into the correct categories followed by the $ISRs - P_{osi}$, as demonstrated by the results obtained from a simulation study and real data examples. Thus, the newly established $ESRs - P_{osi}$ plot can be a suitable alternative to identify IOs in the spatial regression model.

Author Contributions: Conceptualization, A.M.B. and H.M.; methodology, A.M.B.; software, A.M.B.; validation, A.M.B. and H.M.; formal analysis, A.M.B. and N.H.A.R.; investigation, A.M.B. and N.H.A.R.; resources, A.M.B., H.M. and 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. and N.H.A.R.; visualization, A.M.B. and M.B.A.; supervision, H.M.; project administration, H.M.; funding acquisition, H.M. All authors have read and agreed to the published version of the manuscript.

Funding: This article was partially supported by the Fundamental Research Grant Scheme (FRGS) under the Ministry of Higher Education, Malaysia with project number FRGS/1/2019/STG06/UPM/01/1

Data Availability Statement: Data are available online. Data for Example 1 are available in page 332 of [19]. Data for Example 2 are available online, website link (http://dph.georgia.gov/covid-19-daily-status-report; accessed on 30 June 2020 and http://www.countyheathrankings.org; accessed on 30 June 2020). Data for Example 3 are available online, website link (https://www.kaggle.com/johnjdavisiv/us-counties-covid19-weather-sociohealth-data; accessed on 13 June 2020)

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: New York, NY, USA, 1980; Volume 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. *J. Am. Stat. Assoc.* 1977, 74, 169–174.
4. Hoaglin, D.C.; Welsch, R.E. The Hat Matrix in Regression and ANOVA. *Am. Stat.* **1978**, *32*, 17, https://doi.org/10.2307/2683469.
5. Andrews, D.F.; Pregibon, D. Finding the Outliers That Matter. *J. R. Stat. Soc. Ser. B (Methodol.)* **1978**, *40*, 85–93.
6. Hawkins, D.M. *Identification of Outliers*; Springer: Berlin/Heidelberg, Germany, 1980; Volume 11.
7. Huber, P. *Robust Statistics*; John Wiley and Sons: New York, NY, USA, 1981.
8. Cook, R.D.; Weisberg, S. Monographs on statistics and applied probability. In *Residuals and Influence in Regression*; Chapman and Hall: New York, NY, USA, 1982; ISBN 978-0-412-24280-9.
9. Chatterjee, S.; Hadi, A.S. *Sensitivity Analysis in Linear Regression*; John Wiley & Sons: New York, NY, USA, 1988; Volume 327.
10. Hadi, A.S. A New Measure of Overall Potential Influence in Linear Regression. *Comput. Stat. Data Anal.* **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. *J. Appl. Stat.* **2009**, *36*, 507–520, https://doi.org/10.1080/02664760802553463.
12. Meloun, M.; Militkỳ, J. *Statistical Data Analysis: A Practical Guide*; Woodhead Publishing Limited: Sawston, Cambridge, 2011;.
13. Puterman, M.L. Leverage and Influence in Autocorrelated Regression Models. *J. R. Stat. Soc. Ser. C (Appl. Stat.)* **1988**, *37*, 76–86.
14. Schall, R.; Dunne, T.T. A Unified Approach to Outliers in the General Linear Model. *Sankhyā Indian J. Stat. Ser. B* **1988**, *50*, 157–167.
15. Martin, R.J. Leverage, Influence and Residuals in Regression Models When Observations Are Correlated. *Commun. Stat.-Theory Methods* **1992**, *21*, 1183–1212.
16. Shi, L.; Chen, G. Influence Measures for General Linear Models with Correlated Errors. *Am. Stat.* **2009**, *63*, 40–42.
17. Cerioli, A.; Riani, M. Robust Transformations and Outlier Detection with Autocorrelated Data. In *From Data and Information Analysis to Knowledge Engineering*; Spiliopoulou, M., Kruse, R., Borgelt, C., Nürnberger, A., Gaul, W., Eds.; Studies in Classification, Data Analysis, and Knowledge Organization; Springer: Berlin/Heidelberg, Germany, 2006; pp. 262–269, ISBN 978-3-540-31313-7.
18. Christensen, R.; Johnson, W.; Pearson, L.M. Prediction Diagnostics for Spatial Linear Models. *Biometrika* **1992**, *79*, 583–591, https://doi.org/10.1093/biomet/79.3.583.
19. Haining, R. Diagnostics for Regression Modeling in Spatial Econometrics*. *J. Reg. Sci.* **1994**, *34*, 325–341, https://doi.org/10.1111/j.1467-9787.1994.tb00870.x.
20. Haining, R.P.; Haining, R. *Spatial Data Analysis: Theory and Practice*; Cambridge University Press: Cambridge, England, 2003.
21. Dai, X.; Jin, L.; Shi, A.; Shi, L. Outlier Detection and Accommodation in General Spatial Models. *Stat. Methods Appl.* **2016**, *25*, 453–475, https://doi.org/10.1007/s10260-015-0348-1.
22. Singh, A.K.; Lalitha, S. A Novel Spatial Outlier Detection Technique. *Commun. Stat. Theory Methods* **2018**, *47*, 247–257.
23. Anselin, L. Some Robust Approaches to Testing and Estimation in Spatial Econometrics. *Reg. Sci. Urban Econ.* **1990**, *20*, 141–163, https://doi.org/10.1016/0166-0462(90)90001-J.
24. Cerioli, A.; Riani, M. Robust Methods for the Analysis of Spatially Autocorrelated Data. *Stat. Methods Appl.* **2002**, *11*, 335–358, https://doi.org/10.1007/BF02509831.
25. Yildirim, V.; Mert Kantar, Y. Robust Estimation Approach for Spatial Error Model. *J. Stat. Comput. Simul.* **2020**, *90*, 1618–1638, https://doi.org/10.1080/00949655.2020.1740223.
26. Kou, Y.; Lu, C.-T. Outlier Detection, Spatial. In *Encyclopedia of GIS*; Springer: Boston, MA, USA, 2008; 1539–1546.
27. Hadi, A.S.; Imon, A.H.M.R. Identification of Multiple Outliers in Spatial Data. *Int. J. Stat. Sci.* **2018**, *16*, 87–96.
28. Hadi, A.S.; Simonoff, J.S. Procedures for the Identification of Multiple Outliers in Linear Models. *J. Am. Stat. Assoc.* **1993**, *88*, 1264–1272, https://doi.org/10.1080/01621459.1993.10476407.
29. Aggarwal, C.C. Spatial Outlier Detection. In *Outlier Analysis*; Springer: New York, NY, USA, 2013; pp. 313–341, ISBN 978-1-4614-6395-5.
30. Anselin, L. Local Indicators of Spatial Association—LISA. *Geogr. Anal.* **1995**, *27*, 93–115.
31. Politis, D.; Romano, J.; Wolf, M. Bootstrap Sampling Distributions. In *Subsampling*; Springer: New York, NY, USA, 1999; Available online: https://link.springer.com/chapter/10.1007/978-1-4612-1554-7_1 (accessed on 16 September 2021).
32. Heagerty, P.J.; Lumley, T. Window Subsampling of Estimating Functions with Application to Regression Models. *J. Am. Stat. Assoc.* **2000**, *95*, 197–211, https://doi.org/10.1080/01621459.2000.10473914.
33. Anselin, L. Exploratory Spatial Data Analysis and Geographic Information Systems. *New Tools Spat. Anal.* **1994**, *17*, 45–54.
34. Cressie, N.A.C. *Statistics for Spatial* Data, Rev. ed.; Wiley series in probability and mathematical statistics; Wiley: New York, NY, USA, 1993; ISBN 978-0-471-00255-0.
35. Anselin, L. *Spatial Econometrics: Methods and Models*; Studies in Operational Regional Science; Springer: Dordrecht, The Netherlands, 1988; Volume 4, ISBN 978-90-481-8311-1.
36. LeSage, J.P. *The Theory and Practice of Spatial Econometrics*; University of Toledo: Toledo, OH, USA, 1999; Volume 28.
37. Olver, P.J.; Shakiban, C.; Shakiban, C. *Applied Linear Algebra*; Springer: Berlin/Heidelberg, Germany, 2006; Volume 1.
38. Horn, R.A.; Johnson, C.R. *Matrix Analysis*, 2nd ed.; Cambridge University Press: Cambridge, NY, USA, 2012; ISBN 978-0-521-83940-2.
39. Liesen, J.; Mehrmann, V. *Linear Algebra*; Springer Undergraduate Mathematics Series; Springer International Publishing: Cham, Germany, 2015; ISBN 978-3-319-24344-3.
40. Shekhar, S.; Lu, C.-T.; Zhang, P. A Unified Approach to Detecting Spatial Outliers. *GeoInformatica* **2003**, *7*, 139–166.
41. Billor, N.; Hadi, A.S.; Velleman, P.F. BACON: Blocked Adaptive Computationally Efficient Outlier Nominators. *Comput. Stat. Data Anal.* **2000**, *34*, 279–298.
42. Imon, A. Identifying Multiple High Leverage Points in Linear Regression. *J. Stat. Stud.* **2002**, *3*, 207–218.
43. Rahmatullah Imon, A.H.M. Identifying Multiple Influential Observations in Linear Regression. *J. Appl. Stat.* **2005**, *32*, 929–946, https://doi.org/10.1080/02664760500163599.
44. Midi, H.; Mohammed, A. The Identification of Good and Bad High Leverage Points in Multiple Linear Regression Model. *Math. Methods Syst. Sci. Eng.* **2015**, 147–158.
45. Bagheri, A.; Midi, H. Diagnostic Plot for the Identification of High Leverage Collinearity-Influential Observations. *Sort: Stat. Oper. Res. Trans.* **2015**, *39*, 51–70.

46. Dowd, P. The variogram and kriging: Robust and resistant estimators. In *Geostatistics for Natural Resources Characterization*; Springer: Berlin/Heidelberg, Germany, 1984; pp. 91–106.