mathematics

Article

Bayesian Influence Analysis of the Skew-Normal Spatial Autoregression Models

Yuanyuan Ju [1,2,3,*], Yan Yang [1], Mingxing Hu [1], Lin Dai [1] and Liucang Wu [4]

1 Faculty of Science, Kunming University of Science and Technology, Kunming 650500, China; yy1615106662@163.com (Y.Y.); 18208841368@139.com (M.H.); dailin1968@163.com (L.D.)
2 State Key Laboratory of Complex Nonferrous Metal Resources Clean Utilization, Kunming University of Science and Technology, Kunming 650093, China
3 Key Laboratory of Industrial Engineering Statistical Analysis, Faculty of Science, Kunming University of Science and Technology, Kunming 650500, China
4 Center for Applied Statistics, Kunming University of Science and Technology, Kunming 650500, China; wuliucang@163.com
* Correspondence: jundeyy@126.com

Abstract: In spatial data analysis, outliers or influential observations have a considerable influence on statistical inference. This paper develops Bayesian influence analysis, including the local influence approach and case influence measures in skew-normal spatial autoregression models (SSARMs). The Bayesian local influence method is proposed to evaluate the impact of small perturbations in data, the distribution of sampling and prior. To measure the extent of different perturbations in SSARMs, the Bayes factor, the ϕ-divergence and the posterior mean distance are established. A Bayesian case influence measure is presented to examine the influence points in SSARMs. The potential influence points in the models are identified by Cook's posterior mean distance and Cook's posterior mode distance ϕ-divergence. The Bayesian influence analysis formulation of spatial data is given. Simulation studies and examples verify the effectiveness of the presented methodologies.

Keywords: skew-normal distribution; spatial autoregression model; Bayesian local influence; Bayesian case influence; MCMC algorithm

MSC: 62F15

Citation: Ju, Y.; Yang, Y.; Hu, M.; Dai, L.; Wu, L. Bayesian Influence Analysis of the Skew-Normal Spatial Autoregression Models. *Mathematics* **2022**, *10*, 1306. https://doi.org/10.3390/math10081306

Academic Editor: Christophe Chesneau

Received: 12 March 2022
Accepted: 12 April 2022
Published: 14 April 2022

1. Introduction

In the fields of econometrics, air quality monitoring and epidemic monitoring, spatial data are often encountered. Spatial data are a kind of data with location attributes, which usually have spatial dependence. If the spatial correlation of data is not considered, the conclusion will often be inaccurate or even wrong. Spatial autoregressive models (SARMs) can effectively deal with the data with spatial correlation [1]. Therefore, SARMs have attracted the extensive attention of many scholars over the past few years. For example, Piribauer and Cuaresma [2] proposed two Bayesian variable selection approaches and compared their performance in SARMs. Xie et al. [3] investigated the variable selection problem in SARMs; Du et al. [4] studied the generalized method of moments estimator for partially linear additive SARMs; Li et al. [5] considered a variable selection method for SARMs based on the minimum prediction error criterion; Jay et al. [6] introduced SARMs for the statistical inference of ecological data, which discussed model selection, spatial regression, the estimation of autocorrelation, the estimation of other connectivity parameters, spatial prediction and the spatial smoothing of practical ecological inference; Anik et al. [7] investigated a Lagrange multiplier test of spatial dependence for SARMs with latent variables; Song et al. [8] proposed a class of penalized robust regression estimators based on exponential squared loss for SARMs with independent identically distributed errors. The above studies of SSARMs are mainly based on the hypothesis that the dependent variable follows a normal distribution. However, the normal hypothesis is often excessively restrictive of the variability of spatial data in the actual statistical inference, and unreasonable results may be generated.

In the actual spatial data, there are few cases of strict symmetry. At this time, if the spatial data is studied based on the assumption of normality, it is difficult to capture the changes in the data. In recent years, the skew-normal distribution has attracted widespread attention. A large number of studies on skew-normal distribution have appeared under various statistical models. For example, Marcos et al. [9] developed nonlinear mixed-effects models for a mixture of scales with a skew-normal distributions family and gained an estimate of the parameters by the maximum likelihood method; Yin et al. [10] introduced a variable selection procedure for a finite mixture of regression models using the skew-normal distribution; Tatsuya et al. [11] investigated the decision–theoretic properties of Stein-type shrinkage estimators in multivariate skew-normal distribution under the condition of quadratic loss; Liu et al. [12] considered an autoregressive model based on the skew-normal distribution; Mahdi [13] adopted the EM algorithm to research the maximum likelihood estimation of the mixed model with a skewed-normal assumption. However, as far as we know, there are few studies on SARMs with the response variables following skew-normal distribution.

In recent years, the Bayesian methods of local influence and case influence analysis have become widely used statistical diagnostic methods. For Bayesian local influence analysis, it is extensively studied within different models based on several objective functions. For example, Zhu et al. [14] established a universal framework of Bayesian influence analysis to evaluate the impact of simultaneous perturbations of prior, data and sample distribution for a set of statistical models; Zhang et al. [15] proposed Bayesian local influence analysis to evaluate the impact of different perturbations of individual observations, prior distribution and nonignorable missing data mechanisms in general EEs. Ouyang et al. [16] applied the Bayesian local influence method to semiparametric structural equation models and the effects of minor perturbations were evaluated using different perturbation options; Dai et al. [17] studied two Bayesian local diagnostic procedures for heteroscedastic SARMs; Ju et al. [18] developed a new SDPDMs by assuming that the random effects and error terms obey a skew-normal distribution and also developed a Bayesian local influence analysis method for it. For Bayesian case influence diagnostics, they are widely researched in various statistical models based on conditional predictive coordinates and K-L distance. Cancho et al. [19] presented Bayesian case deletion influence diagnostics for nonlinear regression models with scale mixtures of skew-normal distributions based on the K-L divergence; Zhu et al. [20] proposed three Bayesian case influence measures to identify influential points for a class of statistical models with missing data, which included the Cook's posterior mean and mode distance and ϕ-divergence; Tang et al. [21] detected the influence observations of generalized partial linear mixed models for longitudinal data based on ϕ-divergence and Cook's posterior mean distance; Hao et al. [22] investigated Bayesian case influence analysis based on the K-L divergence of a generalized autoregressive conditional heteroscedasticity model; Duan et al. [23] developed a Bayesian case deletion influence measure based on the ϕ-divergence of a semiparametric reproductive dispersion mixed model and presented computationally feasible formulas. So far, there are few research results regarding the Bayesian influence analysis of skew-normal spatial autoregression models (SSARMs). The importance of SSARMs is mainly reflected in the fact that SSARMs not only consider the spatial correlation of economic individuals in different regions but also consider the skewness of spatial data. It is more effective to make statistical inferences by the autoregressive model of skewed normal space than the traditional regression model. However, the observed data may deviate greatly from the established SSARMs. The motivation of this paper is that there are outliers or strong influential cases in the actual spatial data, which have an effect on the statistical analysis and inference. Therefore, based on references [14,20], a methodology of Bayesian statistical diagnosis for

SSARMs is developed in this paper. The contribution of this paper is that Bayesian influence analysis, including the local influence approach and case influence measures in SSARMs, is proposed. The Bayesian local influence method is proposed to evaluate the impact of small perturbations in the data, distribution of sampling and prior. To measure the extent of different perturbations in SSARMs, the Bayes factor, the ϕ-divergence and the posterior mean distance are established. A Bayesian case influence measure is presented to examine the influence points in SSARMs. The potential influence points in the models are identified by Cook's posterior mean distance and Cook's posterior mode distance ϕ-divergence. The Bayesian influence analysis formulation of spatial data is given.

The rest of this paper is structured as follows. The Markov chain Monte Carlo (MCMC) algorithm and skew-normal spatial autoregression models are introduced in Section 2. Three Bayesian local influence methods are used to evaluate the effects of minor perturbation on the data, priors and distribution of sampling in Section 3. Section 4 shows three Bayesian case influence methods, which are used to recognize the outliers or influential points. Section 5 illustrates the effectiveness of the proposed approach through simulation studies and examples. The concluding section is presented in Section 6.

2. Models Introduction and MCMC Algorithm

$\boldsymbol{Y} = (y_1, \ldots, y_n)^T$ is a $n \times 1$ dimensional dependent variables vector; $\boldsymbol{X}$ is a $n \times p$ dimensional explanatory variables matrix; $\boldsymbol{W}$ is a $n \times n$ dimensional spatial weights matrix; ρ represents the strength of the spatial dependence, which can be used to measure the effect of geographic correlation on the dependent variable; $\boldsymbol{\beta}$ is a $p \times 1$ dimensional regression coefficients vector, which can indicate the degree of influence of the corresponding independent variable on the dependent variable; and $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_n)^T$ is a $n \times 1$ dimensional vector of disturbances. The traditional assumption on $\boldsymbol{\varepsilon}$ is $\boldsymbol{\varepsilon} \sim N_n(\boldsymbol{0}, \sigma_\varepsilon^2 \boldsymbol{I}_n)$. We assume the distribution of $\boldsymbol{\varepsilon}$ follows skew-normal, that is, $\boldsymbol{\varepsilon} \overset{\text{ind.}}{\sim} SN_n\left(\boldsymbol{0}, \sigma_\varepsilon^2 \boldsymbol{I}_n, \delta_\varepsilon^2\, \boldsymbol{I}_n\right)$. The advantages we assume are (1) the skewness and heavy tails can be simultaneously explained; (2) the assumption of normal distribution can be relaxed; and (3) the accurate representation of the structure present in spatial data can be provided. Therefore, SSARMs have the following expression:

$$\boldsymbol{Y} = \rho \boldsymbol{W}\boldsymbol{Y} + \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}. \tag{1}$$

Similar to Arellano-Valle et al. [24], skewed-normal distribution is defined as follows.

Definition 1. *If a n-dimensional random vector $\boldsymbol{z}$ obeys a n-variate skew-normal distribution, its position vector $\boldsymbol{\tau} \in \mathbb{R}^n$, the scale matrix $\sum$ is a $n \times n$ positive definite matrix and $\boldsymbol{\Delta}$ is a $n \times m$ skewness matrix, the density function is expressed as*

$$f(z|\boldsymbol{\tau}, \textstyle\sum, \boldsymbol{\Delta}) = 2^m \phi_n(z|\boldsymbol{\tau}, \sum + \boldsymbol{\Delta}\boldsymbol{\Delta}^T) \times \Phi_m(\boldsymbol{\Delta}^T (\sum + \boldsymbol{\Delta}\boldsymbol{\Delta}^T)^{-1} (\boldsymbol{z} - \boldsymbol{\tau}) | \boldsymbol{0}, (\boldsymbol{I}_m + \boldsymbol{\Delta}^T {\sum}^{-1} \boldsymbol{\Delta})^{-1} \tag{2}$$

which is denoted as $\boldsymbol{z} \sim SN_{n,m}(\boldsymbol{\tau}, \sum, \boldsymbol{\Delta})$. $\boldsymbol{z} \sim SN_n(\boldsymbol{\tau}, \sum, \boldsymbol{\Delta})$ when $n = m$, which can be given a stochastic representation as $\boldsymbol{z} \overset{d}{=} \boldsymbol{\Delta}|\boldsymbol{z_0}| + \boldsymbol{z}_1$, where $\boldsymbol{z}_0 \sim N_n(\boldsymbol{0}, \boldsymbol{I}_n)$, $\boldsymbol{z}_1 \sim N_n(\boldsymbol{\tau}, \sum)$ and '$\overset{d}{=}$' represents 'distributed as', with $\boldsymbol{z}_0$ and $\boldsymbol{z}_1$ being independent.

According to the Definition 1, Equation (1) is represented as follows:

$$\boldsymbol{Y} - \rho \boldsymbol{W}\boldsymbol{Y} \Big| \boldsymbol{\beta}, \delta_\varepsilon, \sigma_\varepsilon^2, \boldsymbol{R}_\varepsilon \overset{\text{ind.}}{\sim} N\Big(\boldsymbol{X}\boldsymbol{\beta} + \delta_\varepsilon \boldsymbol{R}_\varepsilon, \sigma_\varepsilon^2 \boldsymbol{I}_n\Big), \tag{3}$$

where δ_ε represents the skewness parameters, $\boldsymbol{R}_\varepsilon \overset{\text{ind.}}{\sim} N_n(\boldsymbol{0}, \boldsymbol{I}_n)\mathbb{I}\{\boldsymbol{R}_\varepsilon > 0\}$, and $\mathbb{I}$ represents the indicator function.

The log-likelihood function for SSARMs is given by

$$\ell(\boldsymbol{Y}|\theta) = (2\pi)^{-\frac{n}{2}} |\sigma_\varepsilon^2 \boldsymbol{I}_n|_{-\frac{1}{2}} |\boldsymbol{I}_n - \rho \boldsymbol{W}| \exp\left\{ -\frac{1}{2} \boldsymbol{e}^T (\sigma_\varepsilon^2 \boldsymbol{I}_n)^{-1} \boldsymbol{e} \right\}, \tag{4}$$

where $\boldsymbol{e} = \boldsymbol{Y} - \rho\boldsymbol{W}\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta} - \delta_\varepsilon\boldsymbol{R}_\varepsilon$, $\boldsymbol{\theta} = \left(\rho, \boldsymbol{\beta}, \sigma_\varepsilon^2, \delta_\varepsilon\right)$.

To conduct Bayesian inference on SSARMs, we first take the random sample from the joint posterior distribution $p(\rho, \boldsymbol{\beta}, \sigma_\varepsilon^2, \delta_\varepsilon \big| \boldsymbol{Y}, \boldsymbol{X})$ using the MCMC algorithm, and then use the random sample to compute Bayesian estimates of the unknown parameters. During the sampling process, the observations $\left\{\rho, \boldsymbol{\beta}, \sigma_\varepsilon^2, \delta_\varepsilon\right\}$ of the sample are generated from the following conditional distributions: $p(\rho \big| \boldsymbol{Y}, \boldsymbol{\beta}, \sigma_\varepsilon^2, \delta_\varepsilon)$, $p(\boldsymbol{\beta} \big| \boldsymbol{Y}, \rho, \sigma_\varepsilon^2, \delta_\varepsilon)$, $p(\sigma_\varepsilon^{-2} \big| \boldsymbol{Y}, \rho, \boldsymbol{\beta}, \delta_\varepsilon)$ and $p(\delta_\varepsilon \big| \boldsymbol{Y}, \rho, \boldsymbol{\beta}, \sigma_\varepsilon^2)$. To obtain the conditional distributions of unknown parameters in $\left\{\rho, \boldsymbol{\beta}, \sigma_\varepsilon^2, \delta_\varepsilon\right\}$, assigning their corresponding prior distributions is essential. Similar to Arellano-Valle et al. [22], the prior distributions of $\rho, \boldsymbol{\beta}, \sigma_\varepsilon^{-2}, \delta_\varepsilon$ are considered follows.

$$\begin{array}{c} p(\rho) \sim U(-1, 1) \\ p(\boldsymbol{\beta}) \sim N(\boldsymbol{\beta}_0, \boldsymbol{M}_\beta) \\ p(\sigma_\varepsilon^{-2}) \sim \Gamma(v, s) \\ p(\delta_\varepsilon) \sim N(\delta_{\varepsilon 0}, M_{\delta_\varepsilon})\mathbb{I}\{\delta_\varepsilon > 0\}, \end{array} . \quad (5)$$

where $\boldsymbol{\beta}_0$, $\boldsymbol{M}_\beta$, v, s, $\delta_{\varepsilon 0}$ and M_{δ_ε} are hyperparameters whose values are given in advance, and $\sigma_\varepsilon^{-2} = 1/\sigma_\varepsilon^2$.

The required posterior distributions are provided in Appendix A.

3. Bayesian Local Influence Analysis

This section uses the Bayesian local influence analysis method to evaluate the impact of small perturbations in the data, prior distribution and sample distribution on the posterior distribution in the SSARMs.

3.1. Perturbation Model and Manifold

Following Zhu et al. [14], the following perturbation model with skew-normal spatial data is considered:

$$p(\boldsymbol{Y}, \boldsymbol{\theta} \big| \boldsymbol{X}, \boldsymbol{\omega}) = p(\boldsymbol{\theta} \big| \boldsymbol{\omega}_p) p(\boldsymbol{Y} \big| \boldsymbol{X}, \boldsymbol{\theta}, \boldsymbol{\omega}_d, \boldsymbol{\omega}_s) \quad (6)$$

and $\int p(\boldsymbol{Y}, \boldsymbol{\theta} \big| \boldsymbol{X}, \boldsymbol{\omega}) d\boldsymbol{Y} d\boldsymbol{\theta} = 1$, where $\boldsymbol{Y} = (y_1, \ldots, y_n)^T$ is a $n \times 1$ dimensional dependent variables vector of expression (1), $\boldsymbol{\omega}_p \in \Re^{k_p}$ represents the perturbations to the prior, $\boldsymbol{\omega}_d \in \Re^{k_d}$ represents the perturbations to the data and $\boldsymbol{\omega}_s \in \Re^{k_s}$ represents the perturbations to sampling distribution. Denote $k = k_p + k_d + k_s$. Suppose $\boldsymbol{\omega}^0 = (\boldsymbol{\omega}_p^{0^T}, \boldsymbol{\omega}_d^{0^T}, \boldsymbol{\omega}_s^{0^T})^T \in \Re^k$ represents no perturbation.

Similar to Zhu et al. [14], the perturbed model $\mathscr{M} = \left\{p(\boldsymbol{Y}, \boldsymbol{\theta} \Big| \boldsymbol{X}, \boldsymbol{\omega}) : \boldsymbol{\omega} \in \Re^k\right\}$ is regarded as the k-dimensional manifold under some regular conditions. The tangent space $\boldsymbol{T}_\omega$ of $\mathscr{M}$ at each $\omega \in \mathscr{M}$ is spanned by k functions $\partial_{\omega_q}\ell(\boldsymbol{\omega})$, where $\partial_{\omega_q} = \partial/\partial\omega_q$, $\ell(\boldsymbol{\omega}) = \log p(\boldsymbol{Y}, \boldsymbol{\theta}|\boldsymbol{X}, \boldsymbol{\omega})$ and ω_q is the qth element of $\boldsymbol{\omega}$. So,

$$g_{iq}(\boldsymbol{\omega}) = E_\omega\left\{\partial_{\omega_j}\ell(\boldsymbol{\omega})\partial_{\omega_q}\ell(\boldsymbol{\omega})\right\} = E_\omega\left\{-\partial^2_{\omega_j\omega_q}\ell(\boldsymbol{\omega})\right\}, \ j, q = 1, \cdots, k. \quad (7)$$

where $\partial^2_{\omega_j\omega_q}\ell(\boldsymbol{\omega}) = \partial^2\ell(\boldsymbol{\omega})/\partial\omega_j\partial\omega_q$, the expectation of the joint probability density function $p(\boldsymbol{Y}, \boldsymbol{\theta}|\boldsymbol{X}, \boldsymbol{\omega})$ is denoted as E_ω. The diagonal element $g_{jj}(\boldsymbol{\omega})$ is considered to measure the amount of perturbation ω_j and the quantity $r_{iq} = g_{jq}(\boldsymbol{\omega})/\sqrt{g_{jj}(\boldsymbol{\omega})g_{qq}(\boldsymbol{\omega})}$ is considered to measure the association between perturbations ω_j and ω_q. If $\boldsymbol{G}(\boldsymbol{\omega}^0)$ is a diagonal matrix, all elements of $\boldsymbol{\omega}$ are mutually orthogonal and the relevant perturbation is referred to as a suitable perturbation. If $\boldsymbol{G}(\boldsymbol{\omega}^0)$ is not a diagonal matrix, the relevant perturbation is referred to as an improper perturbation. However, we can always find an appropriate perturbation vector $\widetilde{\omega} = \boldsymbol{\omega}^0 + \left\{\boldsymbol{G}(\boldsymbol{\omega}^0)\right\}^{\frac{1}{2}}(\omega - \boldsymbol{\omega}^0)$ such that $\boldsymbol{G}(\widetilde{\omega})$ evaluated at $\boldsymbol{\omega}^0$ equals $c\boldsymbol{I}_k$, where c is a positive number.

The Bayesian local influence analysis method is designed to quantify the degree of dependence of the posterior distribution on these three key elements of Bayesian analysis, including the prior distribution, the sampling distribution and the data. To illustrate the effect of small perturbations of the prior distribution, the sampling distribution and the data on SSARMs, the following four forms of perturbation are considered. The perturbation of covariate is considered in Example 1. The perturbation of the dependent variable is considered in Example 2. The perturbation of the prior distribution is considered in Example 3. The perturbation of the sampling distribution is considered in Example 4.

Example 1. *The Bayesian perturbation model for the* spatial data includes many perturbation schemes. The perturbation scheme to data points is proposed for identifying outliers and influential observations. Consider the perturbation covariate $\boldsymbol{X}: \ \boldsymbol{X}(\boldsymbol{\omega}_x) = (\boldsymbol{X}_1 + \omega_1 \mathbf{1}_p, \ldots, \boldsymbol{X}_n + \omega_n \mathbf{1}_p)^T = \boldsymbol{X} + \boldsymbol{\omega}_x \mathbf{1}_p^T$, in which $\boldsymbol{\omega}_x = (\omega_1, \ldots, \omega_n)^T$, $\mathbf{1}_p = (1, \ldots, 1)^T$ is a $p \times 1$ vector and $\boldsymbol{\omega}_x^0 = (0, \ldots, 0)^T$ stands for no perturbation. A Riemann manifold is formed by a perturbation model $\mathscr{M} = \{p(\boldsymbol{Y}, \boldsymbol{\theta} | \boldsymbol{X}, \boldsymbol{\omega}_x) : \boldsymbol{\omega}_x \in \Re^n\}$. The tangent space $\boldsymbol{T}_{\omega_x}$ of $\mathscr{M}$ can be spanned by

$$\partial_{\omega_x} \ell(\boldsymbol{\omega}_x) = (\mathbf{1}_p^T \boldsymbol{\beta})^T (\sigma_\varepsilon^2 \boldsymbol{I}_n)^{-1} \Big[\boldsymbol{Y} - \rho \boldsymbol{W} \boldsymbol{Y} - \Big(\boldsymbol{X} + \boldsymbol{\omega}_x \mathbf{1}_p^T\Big)\boldsymbol{\beta} - \delta_\varepsilon \boldsymbol{R}_\varepsilon\Big], \tag{8}$$

when $\boldsymbol{\omega}_x = \boldsymbol{\omega}_x^0$. We have

$$\partial_{\omega_x} \ell\Big(\boldsymbol{\omega}_x^0\Big) = (\mathbf{1}_p^T \boldsymbol{\beta})^T (\sigma_\varepsilon^2 \boldsymbol{I}_n)^{-1} (\boldsymbol{Y} - \rho \boldsymbol{W} \boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta} - \delta_\varepsilon \boldsymbol{R}_\varepsilon). \tag{9}$$

It is easily shown that $\boldsymbol{G}(\boldsymbol{\omega}_x^0) = E\{-\partial^2_{\omega_x} \ell(\boldsymbol{\omega}_x^0)\} = (\mathbf{1}_p^T \boldsymbol{\beta})^T (\sigma_\varepsilon^2 \boldsymbol{I}_n)^{-1} \Big(\mathbf{1}_p^T \boldsymbol{\beta}\Big)$. When $\boldsymbol{G}(\boldsymbol{\omega}_x^0)$ is a diagonal matrix, the perturbation is referred to as a suitable perturbation. Otherwise, the perturbation is deemed to be inappropriate.

Example 2. *Bayesian local influence analysis methods have many more perturbation schemes for spatial data points. In Example 1, perturbations of $\boldsymbol{X}$ are considered. However, perturbations of the dependent variable $\boldsymbol{Y}$ also have an impact on the Bayesian analysis. Consider the perturbation to the response variables $\boldsymbol{Y}$:$\boldsymbol{Y}(\boldsymbol{\omega}_y) = (y_1 + \omega_1, \ldots, y_n + \omega_n)^T = \boldsymbol{Y} + \boldsymbol{\omega}_y$, in which $\boldsymbol{\omega}_y = (\omega_1, \ldots, \omega_n)^T$ and $\boldsymbol{\omega}_y^0 = (0, \ldots, 0)^T$ stands for no perturbation. A Riemann manifold is formed by a perturbation model $\mathscr{M} = \{p(\boldsymbol{Y}, \boldsymbol{\theta} | \boldsymbol{X}, \boldsymbol{\omega}_y) : \boldsymbol{\omega}_y \in \mathfrak{R}^n\}$. The tangent space $\boldsymbol{T}_{\omega_y}$ of $\mathscr{M}$ can be spanned by*

$$\partial_{\omega_y} \ell(\boldsymbol{\omega}_y) = (\boldsymbol{I}_n - \rho \boldsymbol{W})^T (\sigma_\varepsilon^2 \boldsymbol{I}_n)^{-1} [\widetilde{\boldsymbol{Y}}(\boldsymbol{\omega}_y) - \boldsymbol{X}\boldsymbol{\beta} - \delta_\varepsilon \boldsymbol{R}_\varepsilon, \tag{10}$$

in which $\widetilde{\boldsymbol{Y}}(\boldsymbol{\omega}_y) = \boldsymbol{Y} + \boldsymbol{\omega}_y - \rho \boldsymbol{W}(\boldsymbol{Y} + \boldsymbol{\omega}_y)$. When$\boldsymbol{\omega}_y = \boldsymbol{\omega}_y^0$, we have

$$\partial_{\omega_y} \ell\Big(\boldsymbol{\omega}_y^0\Big) = (\boldsymbol{I}_n - \rho \boldsymbol{W})^T (\sigma_\varepsilon^2 \boldsymbol{I}_n)^{-1} (\boldsymbol{Y} - \rho \boldsymbol{W} \boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta} - \delta_\varepsilon \boldsymbol{R}_\varepsilon). \tag{11}$$

It is easily shown that $\boldsymbol{G}\Big(\boldsymbol{\omega}_y^0\Big) = E\Big\{-\partial^2_{\omega_y} \ell\Big(\boldsymbol{\omega}_y^0\Big)\Big\} = (\boldsymbol{I}_n - \rho \boldsymbol{W})^T (\sigma_\varepsilon^2 \boldsymbol{I}_n)^{-1} (\boldsymbol{I}_n - \rho \boldsymbol{W})$. When $\boldsymbol{G}\Big(\boldsymbol{\omega}_y^0\Big)$ is a diagonal matrix, the perturbation is referred to as a suitable perturbation. Otherwise, the perturbation is deemed to be inappropriate.

Example 3. *The Bayesian perturbation model for the prior distribution includes many existing schemes. Consider the perturbation to the prior distribution of $\boldsymbol{\beta}$: $p(\boldsymbol{\beta} | \omega_\beta) \sim N(\omega_\beta \boldsymbol{\beta}, \boldsymbol{M}_\beta)$, in which ω_β is a positive number. Additionally, $\omega_\beta^0 = 1$ stands for no perturbation. A Riemann manifold is formed by a perturbation model $\mathscr{M} = \Big\{p\Big(\boldsymbol{Y}, \boldsymbol{\theta} \Big| \boldsymbol{X}, \omega_\beta\Big) : \omega_\beta \in \mathfrak{R}^1\Big\}$. The tangent space $\boldsymbol{T}_{\omega_\beta}$ of $\mathscr{M}$ can be spanned by*

$$\partial_{\omega_\beta} \ell(\omega_\beta) = \boldsymbol{\beta}_0^T (\boldsymbol{M}_\beta)^{-1} (\boldsymbol{\beta} - \omega_\beta \boldsymbol{\beta}_0). \tag{12}$$

It is easily shown that $\boldsymbol{G}\left(\omega_{\beta}^{0}\right) = E\left\{-\partial_{\omega_\beta}^{2}\ell\left(\omega_{\beta}^{0}\right)\right\} = \boldsymbol{\beta}_0^T(\boldsymbol{M}_\beta)^{-1}\boldsymbol{\beta}_0$. *When* $\boldsymbol{G}\left(\omega_{\beta}^{0}\right)$ *is a diagonal matrix, the perturbation is referred to as a suitable perturbation. Otherwise, the perturbation is deemed to be inappropriate.*

Example 4. *The Bayesian perturbation model for the sampling distribution includes many perturbation schemes. Consider the perturbation to the distribution of the disturbances* $\boldsymbol{\varepsilon}$*:* $p(\boldsymbol{\varepsilon}|\omega_\varepsilon) \sim SN_n\left(\mathbf{0}, \sigma_\varepsilon^2\boldsymbol{I}_n, \omega_\varepsilon\delta_\varepsilon\boldsymbol{I}_n\right)$*, in which* ω_ε *is a scalar and* $\omega_\varepsilon^0 = 1$ *means no perturbation. A Riemann manifold is formed by a perturbation model* $\mathscr{M} = \left\{p(\boldsymbol{Y}, \boldsymbol{\theta}\middle|\boldsymbol{X}, \omega_\varepsilon) : \omega_\varepsilon \in \Re^1\right\}$*. The tangent space* $\boldsymbol{T}_{\omega_\varepsilon}$ *of* $\mathscr{M}$ *is spanned by*

$$\partial_{\omega_\varepsilon}\ell(\omega_\varepsilon) = (\delta_\varepsilon\boldsymbol{R}_\varepsilon)^T\left(\sigma_\varepsilon^2\boldsymbol{I}_n\right)^{-1}(\boldsymbol{Y} - \rho\boldsymbol{W}\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta} - \omega_\varepsilon\delta_\varepsilon\boldsymbol{R}_\varepsilon). \tag{13}$$

It is easily shown that $\boldsymbol{G}\left(\omega_\varepsilon^0\right) = E\left\{-\partial_{\omega_\varepsilon}^2\ell\left(\omega_\varepsilon^0\right)\right\} = (\delta_\varepsilon\boldsymbol{R}_\varepsilon)^T\left(\sigma_\varepsilon^2\boldsymbol{I}_n\right)^{-1}(\delta_\varepsilon\boldsymbol{R}_\varepsilon)$. *When* $\boldsymbol{G}\left(\omega_\varepsilon^0\right)$ *is a diagonal matrix, the perturbation is referred to as a suitable perturbation. Otherwise, the perturbation is deemed to be inappropriate.*

3.2. Local Influence Measures

Following Zhu et al. [14], let objective function $f(\boldsymbol{\omega}) : \mathscr{M} \to \Re^r$ be the objective function. The influence of perturbation can be measured by considering the Bayes factor, ϕ-divergence and posterior mean distance. If $\boldsymbol{\omega}(t)$ is a geodesic on the finite dimensional manifold $\mathscr{M}$ with $\boldsymbol{\omega}(0) = \boldsymbol{\omega}^0$ and $\partial_t\boldsymbol{\omega}(t)|_{t=0} = \boldsymbol{h} \in \Re^m$, in practice, the geodesic can be taken as $\boldsymbol{\omega}^0 = t\boldsymbol{h}$. The Taylor expansion of objective function $f(\boldsymbol{\omega}(t))$ at $t = 0$ is $f(\boldsymbol{\omega}(t)) = f(\boldsymbol{\omega}(0)) + \dot{f}_h(0)t + O\left(t^2\right)$, where $\dot{f}_h(0) = \boldsymbol{\nabla}_f^T\boldsymbol{h}$, $\boldsymbol{\nabla}_f = \partial f(\boldsymbol{\omega})/\partial\boldsymbol{\omega}$ and $\boldsymbol{h} = d\boldsymbol{\omega}(t)/dt|_{t=0} \in \Re^m$.

Firstly, when $\boldsymbol{\nabla}_f \neq 0$, the first-order influence measure in the defined direction $\boldsymbol{h} \in \Re^m$ is

$$\boldsymbol{FI}_{f,h} = \boldsymbol{FI}_{f(\omega(0)),h} = \frac{\boldsymbol{h}^T\boldsymbol{\nabla}_f\boldsymbol{W}_f\boldsymbol{\nabla}_f^T\boldsymbol{h}}{\boldsymbol{h}^T\boldsymbol{G}\boldsymbol{h}}, \tag{14}$$

where $\boldsymbol{G} = \boldsymbol{G}\left(\boldsymbol{\omega}^0\right)$, $\boldsymbol{W}_f$ represents a positive semi-definite matrix. Notably, the first-order influence measure can be rewritten with appropriate perturbations $\widetilde{\boldsymbol{\omega}}(\boldsymbol{\omega})$, then further defined as

$$\boldsymbol{FI}_{f(\widetilde{\omega}),h|\widetilde{\omega}=\omega^0} = \boldsymbol{h}^T\boldsymbol{G}^{-\frac{1}{2}}\boldsymbol{\nabla}_f\boldsymbol{W}_f\boldsymbol{\nabla}_f^T\boldsymbol{G}^{-\frac{1}{2}}\boldsymbol{h}. \tag{15}$$

Following Poon and Poon [25], the first-order adjusted influence measure $\boldsymbol{FIC}_{f(\widetilde{\omega}^0),h}$ can be defined as

$$\boldsymbol{FIC}_{f(\widetilde{\omega}^0),h} = \boldsymbol{h}^T\mathscr{B}\boldsymbol{h}, \tag{16}$$

in which $\mathscr{B} = \boldsymbol{Q}/\mathrm{tr}(\boldsymbol{Q})$ in which $\boldsymbol{Q} = \boldsymbol{G}^{-\frac{1}{2}}\boldsymbol{\nabla}_f\boldsymbol{W}_f\boldsymbol{\nabla}_f^T\boldsymbol{G}^{-\frac{1}{2}}$, $\mathrm{tr}(A)$ denotes the trace of a matrix A.

Let $\eta_1 \geq \ldots\eta_t > 0$ be the non-zero eigenvalues of matrix $\mathscr{B}$ and $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_t$ be the corresponding orthogonal eigenvectors, where $\boldsymbol{u}_i = (u_{i1}, \ldots, u_{ik})^T$ for $i = 1, \ldots, \iota$. The local maximum influence degree $\widetilde{\omega}$ is reflected by the maximum eigenvalue η_1. The corresponding eigenvector $\boldsymbol{u}_1$ is the direction of the most significant perturbations showing the perturbations of the three components of the SSARMs to obtain the largest local variation in $\boldsymbol{f}\left(\boldsymbol{\omega}^0\right)$ and further identify influential observations, misspecified priors and insufficient sampling distributions. Regrettably, it is not sufficient to assess $\boldsymbol{u}_1$ local impacts by inspection alone (Poon and Poon [25]). Through their arguments, local influence is evaluated using the overall contribution vector of all eigenvectors associated with all non-zero eigenvalues: $\boldsymbol{M}(0) = \eta_1\boldsymbol{u}_1^2 + \ldots + \eta_t\boldsymbol{u}_t^2$, where $\boldsymbol{u}_i^2 = \left(u_{i1}^2, \ldots, u_{ik}^2\right)^T$. It is easy to show that the jth component of $\boldsymbol{M}(0)$ is $M(0)_j = \sum_{i=1}^{\iota}\eta_i u_{ij}^2 = FIC_{f(\widetilde{\omega}^0),v_j}$, for $j = 1, \ldots, k$, where v_j is a $k \times 1$ primary perturbation vector, the jth element is 1, the rest of the features are 0 and

b_{jj} is the jth diagonal element of matrix $\mathscr{B}$. Similar to Zhu et al. [14], $M(0) + 2SM(0)$ is considered as a benchmark, where $M(0)$ is the mean of $\left\{M(0)_j : j = 1, \ldots, k\right\}$ and $SM(0)$ is the standard error of $\left\{\boldsymbol{M}(0)_j : j = 1, \ldots, k\right\}$. That is to say, if $M(0) + 2SM(0) \leq \mathrm{b}_{jj}$, the jth observation is identified as an influential point.

Example 5 (Bayes factor). *In this example, the perturbation change is measured by the Bayes factor. It measures this change mainly by the difference between* $log\{p(\boldsymbol{Y}|\boldsymbol{\omega}^0)\}$ *without perturbation and* $log\{p(\boldsymbol{Y}|\boldsymbol{\omega})\}$ *with perturbation. Consider objective function as the Bayes factor:* $f(\boldsymbol{\omega}) = log\{p(\boldsymbol{Y}|\boldsymbol{\omega})\} - log\{p(\boldsymbol{Y}|\boldsymbol{\omega}^0)\}$*, where* $p(\boldsymbol{Y}|\boldsymbol{\omega}) = log\int \; p(\boldsymbol{Y}, \boldsymbol{\theta}|\boldsymbol{X}, \boldsymbol{\omega})d\boldsymbol{\theta}$*. It is easy to prove that*

$$\boldsymbol{\nabla}_B = \boldsymbol{E}_{\theta|Y,X,\omega^0}\left\{\partial_\omega logp(\boldsymbol{Y}, \boldsymbol{\theta}\middle|\boldsymbol{X}, \boldsymbol{\omega}^0)\middle|\boldsymbol{Y}\right\} \neq 0, \tag{17}$$

in which $\boldsymbol{\nabla}_B$ *can be approximated by* $\boldsymbol{\nabla}_B \approx \mathcal{S}^{-1}\sum_{s=1}^{\mathcal{S}} \partial_\omega logp(\boldsymbol{Y}, \boldsymbol{\theta}^{(s)}\middle|\boldsymbol{X}, \boldsymbol{\omega}^0)$ *and* $\left\{\boldsymbol{\theta}^{(s)} : s = 1, \ldots, \mathcal{S}\right\}$ *is generated via the MCMC algorithm, unifying theGibbs sampler with the MH algorithm.*

Secondly, when $\boldsymbol{\nabla}_f = 0$, the second-order Taylor expansion of the objective function $f(\boldsymbol{\omega})$ is $f(\boldsymbol{\omega}(t)) = f(\boldsymbol{\omega}(0)) + 0.5\ddot{f}_h(0)t^2 + O(t^3)$, where $\ddot{f}_h(0) = \boldsymbol{h}^T\boldsymbol{H}_f\boldsymbol{h}$ and $\boldsymbol{H}_f = \partial^2_\omega f(\boldsymbol{\omega})\Big|_{\omega=\omega^0}$. Similar to Zhu et al. [14], the second-order influence measure in the direction $\boldsymbol{h} \in \Re^m$ is given as

$$\boldsymbol{SI}_{f,h} = \boldsymbol{SI}_{f(\omega(0)),h} = \frac{\boldsymbol{h}^T\boldsymbol{H}_f\boldsymbol{h}}{\boldsymbol{h}^T\boldsymbol{G}\boldsymbol{h}}. \tag{18}$$

In particular, when the appropriate perturbation $\widetilde{\boldsymbol{\omega}}(\boldsymbol{\omega})$ is added, the second-order influence measure can be further simplified as

$$\boldsymbol{SI}_{f(\widetilde{\omega}),h|\widetilde{\omega}=\omega^0} = \boldsymbol{h}^T\boldsymbol{G}^{-\frac{1}{2}}\boldsymbol{H}_f\boldsymbol{G}^{-\frac{1}{2}}\boldsymbol{h}. \tag{19}$$

Therefore, the second-order adjusted influence measure can be defined as

$$\boldsymbol{SCI}_{f(\widetilde{\omega}^0),h} = \boldsymbol{h}^T\mathscr{B}_s\boldsymbol{h}, \tag{20}$$

in which $\mathscr{B}_s = \boldsymbol{Q}_s/\mathrm{tr}(\boldsymbol{Q}_s)$, $\boldsymbol{Q}_s = \boldsymbol{G}^{-\frac{1}{2}}\boldsymbol{H}_f\boldsymbol{G}^{-\frac{1}{2}}$. The largest eigenvalue of $\mathscr{B}_s$ is an important indicator to evaluate the local influence behavior of the model and the corresponding eigenvector is the direction of the most significant local influence, which reveals how the three elements of SSARMs are perturbed and can be identified as the potential influence points, erroneous priors and improper modelling assumptions.

Example 6 ($\boldsymbol{\phi}$-divergence). *In this example, the perturbation change is measured by the* ϕ*-divergence. It measures this change mainly by the difference between two posterior probability density functions. Setting the objective function* $f(\boldsymbol{\omega})$ *between the two posterior probability density functions before and after the introduction of perturbation* $\boldsymbol{\omega}$ *as* ϕ*-divergence can be defined as*

$$D_\phi(\boldsymbol{\omega}) = \int \; \phi(R(\boldsymbol{\theta}\middle|\boldsymbol{Y}, \boldsymbol{X}, \boldsymbol{\omega}))p(\boldsymbol{\theta}\middle|\boldsymbol{Y}, \boldsymbol{X})d\boldsymbol{\theta}, \tag{21}$$

where $R(\boldsymbol{\theta}\middle|\boldsymbol{Y}, \boldsymbol{X}, \boldsymbol{\omega}) = \frac{p(\boldsymbol{\theta}|\boldsymbol{Y},\boldsymbol{X},\boldsymbol{\omega})}{p(\boldsymbol{\theta}|\boldsymbol{Y},\boldsymbol{X})}$ *and* $\phi(\cdot)$ *is a convex function with* $\phi(1) = 0$*. It is easy to prove that* $\boldsymbol{\nabla}_\phi = 0$ *and*

$$\boldsymbol{H}_\phi = \ddot{\phi}(1)[E_{\omega^0}\left\{\partial_\omega logp(\boldsymbol{Y}, \boldsymbol{\theta}\middle|\boldsymbol{X}, \boldsymbol{\omega}^0)\right\}^{\otimes 2} - \left\{E_{\omega^0}\partial_\omega logp(\boldsymbol{Y}, \boldsymbol{\theta}\middle|\boldsymbol{X}, \boldsymbol{\omega}^0)\right\}^{\otimes 2}], \tag{22}$$

where $a^{\otimes 2} = aa^T$. $\boldsymbol{H}_\phi$ *is difficult to calculate directly because it involves higher dimensional integrals. To solve this problem, the approximate expression of* $\boldsymbol{H}_\phi$ *can be written as*

$$\boldsymbol{H}_\phi \approx \ddot{\phi}(1)\left[\frac{1}{\mathcal{S}}\sum_{s=1}^{\mathcal{S}}\left\{\partial_\omega logp(\boldsymbol{Y},\boldsymbol{\theta}^{(s)}\middle|\boldsymbol{X},\boldsymbol{\omega}^0)\right\}^{\otimes 2} - \left(\frac{1}{\mathcal{S}}\sum_{s=1}^{\mathcal{S}}\partial_\omega logp\left(\boldsymbol{Y},\boldsymbol{\theta}^{(s)}\middle|\boldsymbol{X},\boldsymbol{\omega}^0\right)\right)^{\otimes 2}\right]. \quad (23)$$

Example 7 (Posterior Mean Distance). *In this example, the perturbation change is measured by Cook's posterior mean distance. It measures this change mainly by the difference between two posterior means. Consider the objective function* $f(\boldsymbol{\omega})$ *as Cook's posterior mean distance between two posterior means of known function* $g(\boldsymbol{\theta})$ *before and after the introduction of perturbation* $\boldsymbol{\omega}$*, which is defined as*

$$CM_g(\boldsymbol{\omega}) = \left\{\boldsymbol{Q}_g(\boldsymbol{\omega}) - \boldsymbol{Q}_g\left(\boldsymbol{\omega}^0\right)\right\}^T \boldsymbol{G}_g\left\{\boldsymbol{Q}_g(\boldsymbol{\omega}) - \boldsymbol{Q}_g\left(\boldsymbol{\omega}^0\right)\right\}, \quad (24)$$

where $\boldsymbol{Q}_g(\boldsymbol{\omega}) = \int g(\boldsymbol{\theta})p(\boldsymbol{\theta}\middle|\boldsymbol{Y},\boldsymbol{X},\boldsymbol{\omega})d\boldsymbol{\theta}$ *is the posterior mean of* $g(\boldsymbol{\theta})$ *and* $\boldsymbol{G}_g = [var\{g(\boldsymbol{\theta})|\boldsymbol{Y},\boldsymbol{X},\boldsymbol{\omega}^0\}]^{-1}$. *It is easily shown that* $\boldsymbol{\nabla}_{CM} = 0$ *and* $\boldsymbol{H}_{CM} = \boldsymbol{Q}_g^{*T}\boldsymbol{G}_g\boldsymbol{Q}_g^*$*, where* $\boldsymbol{Q}_g^* = cov_{\omega^0}\{g(\boldsymbol{\theta}), \partial_\omega logp(\boldsymbol{Y},\boldsymbol{\theta}|\boldsymbol{X},\boldsymbol{\omega}^0)\}$*, which can be approximated by*

$$\boldsymbol{Q}_g^* \approx \mathcal{S}^{-1}\sum_{s=1}^{\mathcal{S}}\left\{g\left(\boldsymbol{\theta}^{(s)}\right)\partial_\omega \log p(\boldsymbol{Y},\boldsymbol{\theta}^{(s)}\middle|\boldsymbol{X},\boldsymbol{\omega}^0)\right\} - \left(\mathcal{S}^{-1}\sum_{s=1}^{\mathcal{S}} g\left(\boldsymbol{\theta}^{(s)}\right)\right)\left(\mathcal{S}^{-1}\sum_{s=1}^{\mathcal{S}}\partial_\omega \log p(\boldsymbol{Y},\boldsymbol{\theta}^{(s)}\middle|\boldsymbol{X},\boldsymbol{\omega}^0)\right). \quad (25)$$

For the above considered objective functions and perturbation schemes, the following steps are used to achieve Bayesian local influence analysis.

Step 1. A Bayesian perturbation model $p(\boldsymbol{Y},\boldsymbol{\theta}|\boldsymbol{X},\boldsymbol{\omega})$ is constructed.

Step 2. $\boldsymbol{\nabla}_f = \partial_\omega f(\boldsymbol{\omega}(0))$, $\boldsymbol{H}_f = \partial_\omega^2 f(\boldsymbol{\omega})\Big|_{\omega=\omega^0}$ and $\boldsymbol{G} = \boldsymbol{G}(\boldsymbol{\omega}^0)$ are calculated.

Step 3. If $\boldsymbol{\nabla}_f \neq 0$, the local effects of small perturbations are evaluated by computing the influence measure $\boldsymbol{FCI}_{f(\tilde{\omega}^0),\boldsymbol{e}_j}$ of the first-order adjustment. However, if $\boldsymbol{\nabla}_f = 0$, the second-order adjusted influence measure $\boldsymbol{SCI}_{f(\tilde{\omega}^0),\boldsymbol{e}_j}$ is calculated.

Step 4. For some given objective functions $f(\boldsymbol{\omega})$, if $M(0)_j = \boldsymbol{FCI}_{f(\tilde{\omega}^0),\boldsymbol{e}_j} \geq M(0) + 2SM(0)$ or $M(0)_j = \boldsymbol{SCI}_{f(\tilde{\omega}^0),\boldsymbol{e}_j} \geq M(0) + 2SM(0)$, the jth observation is detected as an influential observation.

4. Bayesian Case Influence Analysis

According to the case deletion approach provided by Cook and Weisberg [26,27], three Bayesian case influence measures are considered in this section. Let $\boldsymbol{Y}_S$ and $\boldsymbol{X}_S$ denote subsamples of $\boldsymbol{Y}$ and $\boldsymbol{X}$, respectively, for which all observations are in S. Let $\boldsymbol{Y}_{[S]}$ and $\boldsymbol{X}_{[S]}$ represent subsamples of $\boldsymbol{Y}$ and $\boldsymbol{X}$, respectively, and delete all observations in $\boldsymbol{Y}_S$ and $\boldsymbol{X}_S$.

To order to measure the impact of deleting observations in $\{\boldsymbol{Y}_S, \boldsymbol{X}_S\}$ on the posterior distribution of $\boldsymbol{\theta}$, the first type of Bayesian case influence measure is the ϕ-influence of $\boldsymbol{Y}_{[S]}$, defined as

$$D_\phi(S) = \int \phi\left(R_{[S]}(\boldsymbol{\theta})\right)p(\boldsymbol{\theta}\middle|\boldsymbol{Y},\boldsymbol{X}_{[S]})d\boldsymbol{\theta}, \quad (26)$$

where $\phi(\cdot)$ is a convex function with $\phi(1) = 0$, and $R_{[S]}(\boldsymbol{\theta}) = p(\boldsymbol{\theta}\middle|\boldsymbol{Y}_{[S]})/p(\boldsymbol{\theta}\middle|\boldsymbol{Y})$. $D_\phi(S)$ measures the difference between $p(\boldsymbol{\theta}\middle|\boldsymbol{Y}_{[S]},\boldsymbol{X}_{[S]})$ and $p(\boldsymbol{\theta}|\boldsymbol{Y},\boldsymbol{X})$ posterior distributions. If $D_\phi(S)$ is large, the observations in S are detected to be influential.

In order to measure the impact of deleting observations in S on the posterior mean of $\boldsymbol{\theta}$, the second type of Bayesian case influence measure that Cook's posterior mean distance, denoted by $CM(S)$, is considered as:

$$CM(S) = (\widetilde{\boldsymbol{\theta}}_{[S]} - \widetilde{\boldsymbol{\theta}})^T \boldsymbol{W}_\theta \left(\widetilde{\boldsymbol{\theta}}_{[S]} - \widetilde{\boldsymbol{\theta}}\right), \tag{27}$$

where $\boldsymbol{W}_\theta$ is a positive definite matrix and can be seen as the inverse of $\boldsymbol{\theta}$'s posterior covariance matrix. $\widetilde{\boldsymbol{\theta}} = \int \boldsymbol{\theta} p(\boldsymbol{\theta} \big| \boldsymbol{Y}) d\boldsymbol{\theta}$ and $\widetilde{\boldsymbol{\theta}}_{[S]} = \int \boldsymbol{\theta} p(\boldsymbol{\theta} \big| \boldsymbol{Y}_{[S]}) d\boldsymbol{\theta}$ are the posterior means of $\boldsymbol{\theta}$ based on data $\boldsymbol{Y}$ and $\boldsymbol{Y}_{[S]}$, respectively. If the value of $CM(S)$ is large, it points out that deleting observations in S will have a significant impact on the posterior mean. Therefore, the observations in S are considered to be strong influence points.

In order to measure the impact of deleting observations in S on the posterior mode of $\boldsymbol{\theta}$, the third type of Bayesian case influence measure is Cook's posterior mode distance, denoted by $CP(S)$, which is considered as:

$$CP(S) = (\hat{\boldsymbol{\theta}}_{[S]} - \hat{\boldsymbol{\theta}})^T \boldsymbol{G}_\theta \left(\hat{\boldsymbol{\theta}}_{[S]} - \hat{\boldsymbol{\theta}}\right), \tag{28}$$

where, $\boldsymbol{G}_\theta$ is a positive definite matrix and $\hat{\boldsymbol{\theta}} = \mathrm{argmax}_\theta logp(\boldsymbol{\theta} \big| \boldsymbol{Y})$ and $\hat{\boldsymbol{\theta}}_{[S]} = \mathrm{argmax}_\theta \log p(\boldsymbol{\theta} \big| \boldsymbol{Y}_{[S]})$ represent the posterior mode of parameter $\boldsymbol{\theta}$ concerning $\boldsymbol{Y}$ and $\boldsymbol{Y}_{[S]}$, respectively. If the value of $CP(S)$ is large, it indicates that deleting observations in S will have a significant impact on the posterior mode. Therefore, the observations in S are considered strong influence points.

To obtain $D_\phi(S)$, $CM(S)$ and $CP(S)$, it is necessary to calculate $p(\boldsymbol{\theta}|\boldsymbol{Y})$ and $p(\boldsymbol{\theta} \big| \boldsymbol{Y}_{[S]})$. When $p_S(\boldsymbol{\theta}) = p(\boldsymbol{Y} \big| \boldsymbol{\theta}) / p(\boldsymbol{Y}_{[S]} \big| \boldsymbol{\theta}) = p(\boldsymbol{Y}_S \big| \boldsymbol{\theta})$, it is easily shown that

$$p\left(\boldsymbol{\theta} \middle| \boldsymbol{Y}_{[S]}\right) = \frac{\{p_S(\boldsymbol{\theta})\}^{-1} p(\boldsymbol{\theta} \big| \boldsymbol{Y})}{\int \{p_S(\boldsymbol{\theta})\}^{-1} p(\boldsymbol{\theta}|\boldsymbol{Y}) d\boldsymbol{\theta}}, \tag{29}$$

where $R_{[S]}(\boldsymbol{\theta}) = \{p_S(\boldsymbol{\theta})\}^{-1} / E_{\boldsymbol{\theta}|\boldsymbol{Y}}\{p_S(\boldsymbol{\theta})\}^{-1}$. Thus, $D_\phi(S)$ can be written as

$$D_\phi(S) = \mathrm{E}_{\boldsymbol{\theta}|\boldsymbol{Y}}\left[\phi\left(\frac{\{p_S(\boldsymbol{\theta})\}^{-1}}{E_{\boldsymbol{\theta}|\boldsymbol{Y}}\{p_S(\boldsymbol{\theta})\}^{-1}}\right)\right], \tag{30}$$

where $\mathrm{E}_{\boldsymbol{\theta}|\boldsymbol{Y}}$ represents the mathematical expectation of the posterior distribution $p(\boldsymbol{\theta}|\boldsymbol{Y})$. The K-L divergence $(\phi(k) = -\log(k))$, $D_\phi(S) = \log \mathrm{E}_{\boldsymbol{\theta}|\boldsymbol{Y}}\{p_S(\boldsymbol{\theta})\}^{-1} + \mathrm{E}_{\boldsymbol{\theta}|\boldsymbol{Y}}\{\log p_S(\boldsymbol{\theta})\}$ can be considered a computational formula of $D_\phi(S)$. To compute $CM(S)$, we need to evaluate $\widetilde{\boldsymbol{\theta}}$ and $\widetilde{\boldsymbol{\theta}}_{[S]}$. Where $\widetilde{\boldsymbol{\theta}} = \mathrm{E}_{\boldsymbol{\theta}|\boldsymbol{Y}}(\boldsymbol{\theta})$ and $\widetilde{\boldsymbol{\theta}}_{[S]} = \mathrm{E}_{\boldsymbol{\theta}|\boldsymbol{Y}}(\boldsymbol{\theta}\{p_S(\boldsymbol{\theta})\}^{-1}) / \mathrm{E}_{\boldsymbol{\theta}|\boldsymbol{Y}}\{p_S(\boldsymbol{\theta})\}^{-1}$. In order to obtain $CP(S)$, it is necessary to calculate $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\theta}}_{[S]}$. Typically, the posterior mode of $\boldsymbol{\theta}$ is not a closed-form, so an iterative algorithm is needed to obtain $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\theta}}_{[S]}$. $\boldsymbol{G}_\theta$ in $CP(S)$ can be analytically obtained by evaluating $\boldsymbol{J}(\boldsymbol{\theta}) = -\partial_\theta^2 \log p(\boldsymbol{\theta} \big| \boldsymbol{Y}) = -\partial_\theta^2 \log p(\boldsymbol{Y} \big| \boldsymbol{\theta}) - \partial_\theta^2 \log p(\boldsymbol{\theta})$ at $\hat{\boldsymbol{\theta}}$. $\boldsymbol{W}_\theta$ can be obtained by computing the value of $\boldsymbol{J}_\theta$ at $\boldsymbol{\theta}$, which can also be equal to the inverse of the posterior covariance matrix of θ. Thus, $D_\phi(S)$, $CM(S)$ and $CP(S)$ are derived directly by averaging MCMC samples over the posterior distribution $p(\boldsymbol{\theta}|\boldsymbol{Y})$.

5. Simulation Studies and Real Examples

The proposed approaches in Sections 3 and 4 will be illustrated by three simulation studies and practical cases of air quality from China (2020) in this section.

5.1. Simulation Studies

The data set $\{y_i : i = 1, \cdots, n\}$ was generated from Equation (1) with $n = 30$ and $p = 3$. $\boldsymbol{X}$ was generated from the normal distribution $N(\boldsymbol{0}, \boldsymbol{I}_{30})$; the measurement errors $\boldsymbol{\varepsilon}$ were generated from the skew-normal distribution $\boldsymbol{\varepsilon} \sim SN_{30}(\boldsymbol{0}, \sigma_\varepsilon^2 \boldsymbol{I}_{30}, \delta_\varepsilon \boldsymbol{I}_{30})$. The spatial weight matrix $\boldsymbol{W}$ was generated from the Delaunay program in MATLAB. The true values of ρ, $\boldsymbol{\beta}$, δ_ε and σ_ε^2 were taken to be $\rho = 0.2$, $\boldsymbol{\beta} = (1.25, 2, 1.2)^T$, $\delta_\varepsilon = 0.1$ and $\sigma_\varepsilon^2 = 0.1$, respectively. The hyperparameters of $\boldsymbol{\beta}_0$, $\boldsymbol{M}_{\boldsymbol{\beta}}$, $\delta_{\varepsilon 0}$, M_{δ_ε}, v_0 and s_0 were taken to be $\boldsymbol{\beta}_0 = (1.25, 2, 1.2)^T$, $\boldsymbol{M}_{\boldsymbol{\beta}} = 0.001\boldsymbol{I}_3$, $\delta_{\varepsilon 0} = 0.1$, $M_{\delta_\varepsilon} = 0.001$, $v_0 = 6$ and $s_0 = 2$, respectively. Such true values of the parameters were selected randomly. The purpose of selecting parameter values is only to test whether the proposed method is effective. The hyperparameters of the prior distribution were selected to be similar to the true value to make the algorithm converge quickly.

To illustrate the previously proposed Bayesian theory of local influence analysis, the following two simulation studies were considered.

Simulation 1. *In this simulation study, the outliers were generated in the following two ways (i) x_{16} was changed to $x_{16} + 2$ and x_{29} was changed to $x_{29} + 2$, and (ii) y_7 was changed to $y_7 + 8$ and y_{22} was changed to $y_{22} + 8$. For case (i), for the covariate perturbation in Example 1, the Bayesian local influence measures$\boldsymbol{FIC}_{B,e_j}$,$\boldsymbol{SIC}_{D_\phi,e_j}$ and $\boldsymbol{SIC}_{M_d,e_j}$ were calculated by using an MCMC algorithm iteration 10,000 times, discarding the first 5000 times and using the last 5000 iteration values to calculate them. The results are shown in Figure 1a–c, respectively. Cases 16 and 29 turned out to be influential observations, as we expected. For case (ii), for the response variables perturbation in Example 2, the Bayesian local influence measures $\boldsymbol{FIC}_{B,e_j}$, $\boldsymbol{SIC}_{D_\phi,e_j}$ and $\boldsymbol{SIC}_{M_d,e_j}$ were calculated by using an MCMC algorithm iteration 10,000 times, discarding the first 5000 times and using the last 5000 iteration values to calculate them. The results are shown in Figure 2a–c, respectively. Cases 7 and 22 turned out to be influential observations, as expected.*

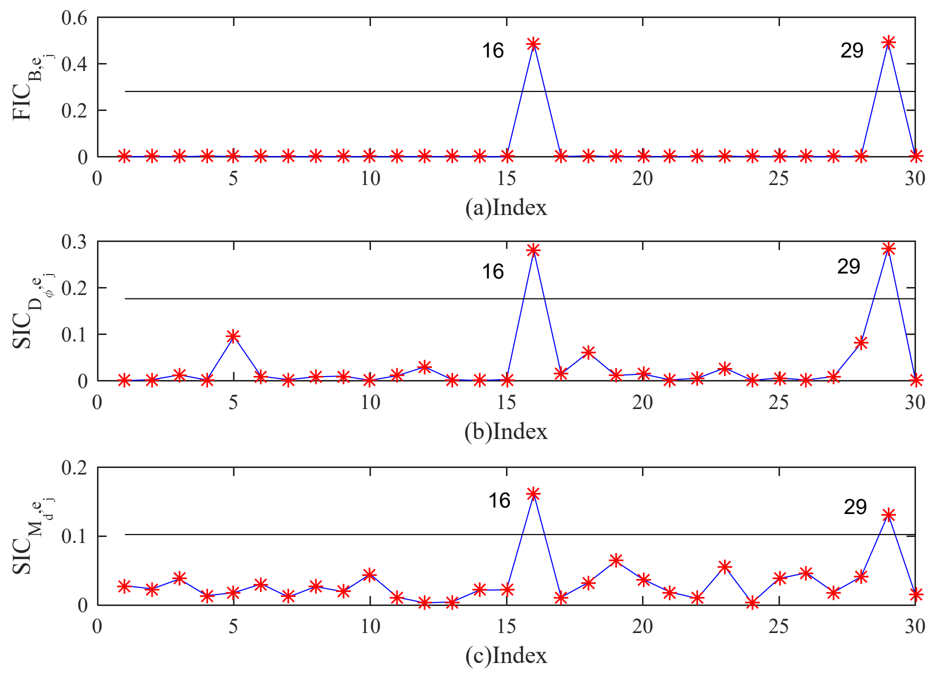

Figure 1. Plots of covariate perturbation in Simulation 1.

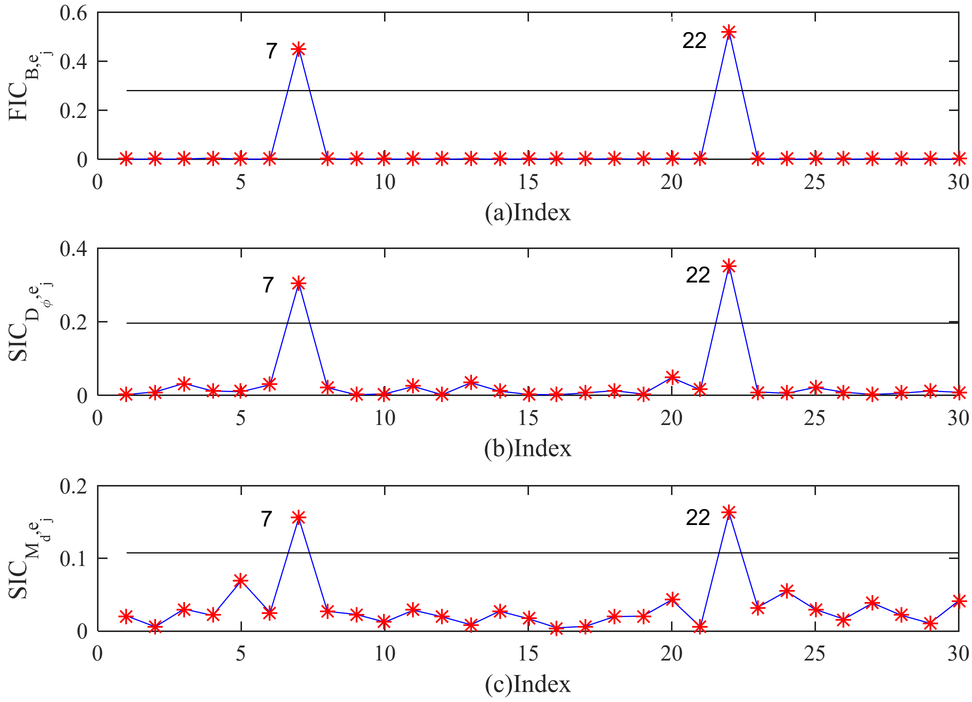

Figure 2. Plots of response variables perturbation in Simulation 1.

Simulation 2. *In this simulation study, perturbations to the data, priors and sampling distribution were considered simultaneously. The outliers generation method was similar to the first method in Simulation 1. (i) The covariate perturbations were considered (see Example 1), (ii) the prior distribution of $\boldsymbol{\beta}$ perturbations was considered (see Example 3) and (iii) the distributions of $\boldsymbol{\varepsilon}$ perturbations were considered (see Example 4). The perturbed log-likelihood function is shown as follows*

$$\log p(\boldsymbol{Y},\boldsymbol{\theta}|\boldsymbol{\omega}) \propto -\frac{1}{2}(\boldsymbol{Y}-\rho\boldsymbol{WY}-\boldsymbol{X}_\omega\boldsymbol{\beta}-\delta_\varepsilon\boldsymbol{R}_\varepsilon)^T(\sigma_\varepsilon^2\boldsymbol{I}_n)^{-1}(\boldsymbol{Y}-\rho\boldsymbol{WY}-\boldsymbol{X}_\omega\boldsymbol{\beta}-\delta_\varepsilon\boldsymbol{R}_\varepsilon)-\frac{1}{2}(\boldsymbol{\beta}-\omega_\beta\boldsymbol{\beta}_0)^T\boldsymbol{M}_\beta^{-1}(\boldsymbol{\beta}-\omega_\beta\boldsymbol{\beta}_0)- \frac{1}{2}(\boldsymbol{Y}-\rho\boldsymbol{WY}-\boldsymbol{X\beta}-\omega_\varepsilon\delta_\varepsilon\boldsymbol{R}_\varepsilon)^T(\sigma_\varepsilon^2\boldsymbol{I}_n)^{-1}(\boldsymbol{Y}-\rho\boldsymbol{WY}-\boldsymbol{X\beta}-\omega_\varepsilon\delta_\varepsilon\boldsymbol{R}_\varepsilon)$$

where $\boldsymbol{\omega}=(\omega_1,\ldots,\omega_{30},\omega_\beta,\omega_\varepsilon)^T$ and $\boldsymbol{X}_\omega=\boldsymbol{X}+\boldsymbol{\omega}_x\boldsymbol{1}_p^T$. In this case, $\boldsymbol{\omega}^0=(0,\ldots,0,1,1)^T$ represents no perturbation. After some calculations, we obtained $\boldsymbol{G}(\boldsymbol{\omega}^0)=diag(a_{11},a_{22},a_{33})$, where $a_{11}=\boldsymbol{\beta}^T\boldsymbol{1}_p(\sigma_\varepsilon^2\boldsymbol{I}_n)^{-1}\boldsymbol{1}_p^T\boldsymbol{\beta}$, $a_{22}=\boldsymbol{\beta}_0^T\boldsymbol{M}_\beta^{-1}\boldsymbol{\beta}_0$ and $a_{33}=(\delta_\varepsilon\boldsymbol{R}_\varepsilon)^T(\sigma_\varepsilon^2\boldsymbol{I}_n)^{-1}\delta_\varepsilon\boldsymbol{R}_\varepsilon$. The Bayesian local influence measures $\boldsymbol{FIC}_{B,e_j}$, $\boldsymbol{SIC}_{D_\phi,e_j}$ and $\boldsymbol{SIC}_{M_d,e_j}$ were concluded using an MCMC algorithm iteration 10,000 times, discarding the first 5000 times and using the last 5000 iteration values to calculate them based on prior distribution $N(10\boldsymbol{\beta}_0,\boldsymbol{M}_\beta)$ of $\boldsymbol{\beta}$ and the distributions $SN_{30}(\boldsymbol{0},\sigma_\varepsilon^2\boldsymbol{I}_n,30\delta_\varepsilon\boldsymbol{R}_\varepsilon)$ of $\boldsymbol{\varepsilon}$. The results are shown in Figure 3a–c. The results showed that cases 16 and 29 were identified to be influential observations, and the prior distribution of $\boldsymbol{\beta}$ and distributions of $\boldsymbol{\varepsilon}$ were detected to be misspecified distributions with having a large effect on $\boldsymbol{FIC}_{B,e_j}$, $\boldsymbol{SIC}_{D_\phi,e_j}$ and $\boldsymbol{SIC}_{M_d,e_j}$.

To demonstrate the Bayesian case influence diagnostics methods, we consider the following simulation research:

Simulation 3. *In this simulation research, the outliers were generated in the following two ways (i) x_{16} was changed to $x_{16}+2$ and x_{29} was changed to $x_{29}+2$, and (ii) y_7 was changed to y_7+8 and y_{22} was changed to $y_{22}+8$. The Bayesian case influence measures $D_\phi(S)$,$CM(S)$ and $CP(S)$ were calculated using an MCMC algorithm iteration 10,000 times, discarding the first 5000 times and using the last 5000 iteration values to calculate them. The results are shown in Figures 4a–c and 5a–c. In this simulation, samples 16 and 29 are created as outliers for x and the other points as normal. Therefore, results are correct when samples 16 and 29 are detected to be outliers and the* *other points are less influenced or have no influence. Similarly, samples 7 and 22 are creased as outliers for y and the other points as normal. Therefore, the results are correct when samples 7 and 22 are detected to be outliers and the other points are less influenced or have no influence. It can be seen from Figure 4 that cases 16 and 29 were detected to be influential; it can be seen from Figure 5 that cases 7 and 22 were diagnosed as influence points. As we expected, the results are the same as the previous diagnosis of Bayesian local influence measures.*

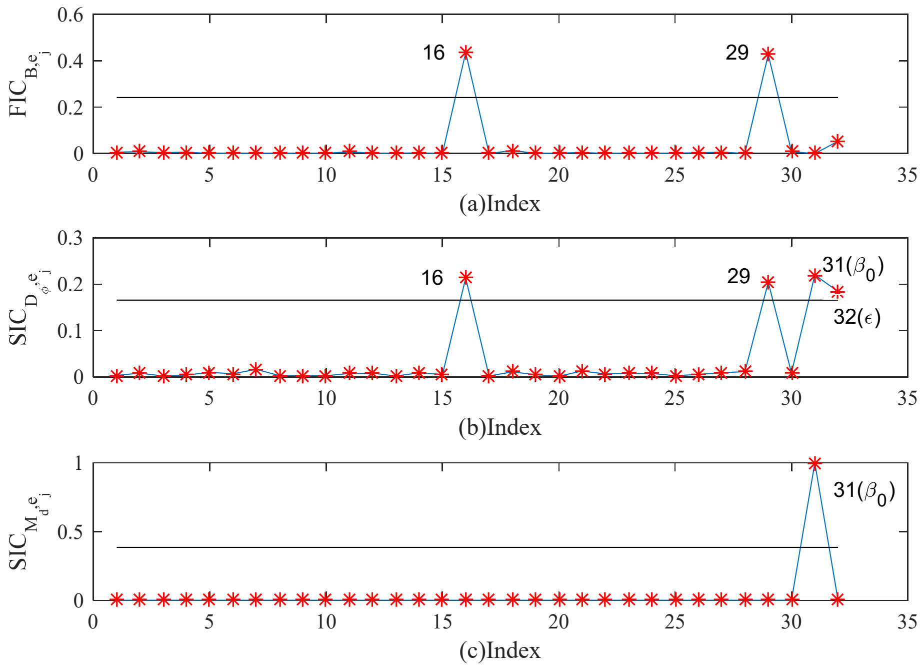

Figure 3. Plots of simultaneous perturbation in Simulation 2.

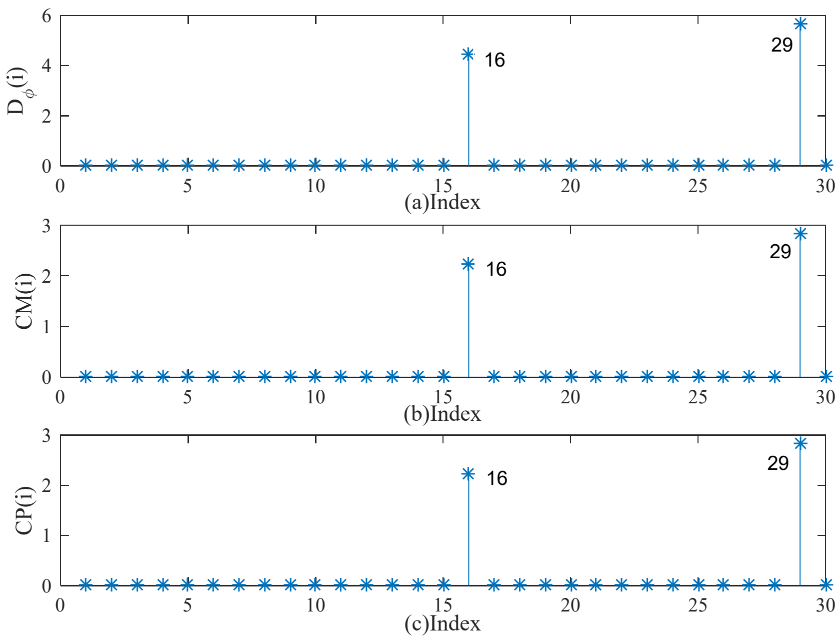

Figure 4. Diagnostic plots of covariate outliers in Simulation 3.

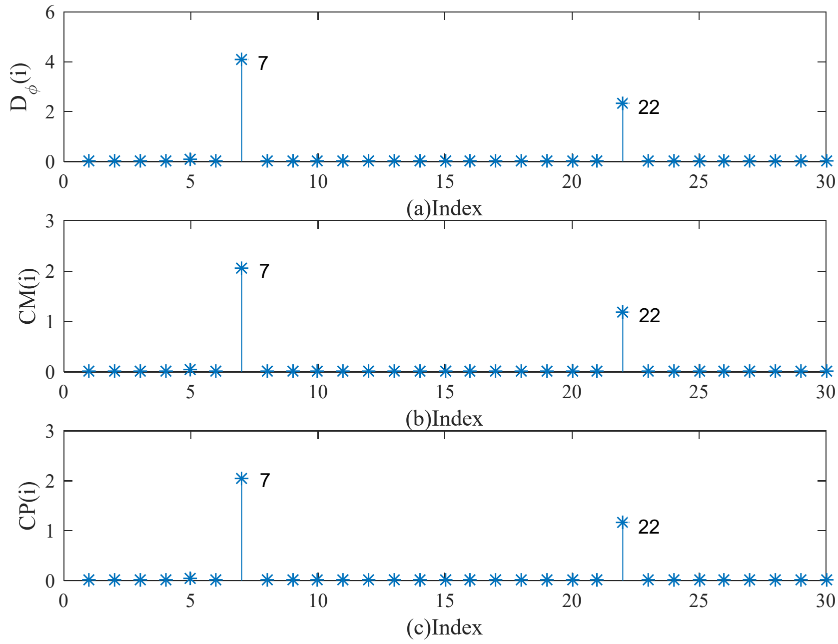

Figure 5. Diagnostic plots of response variables outliers in Simulation 3.

5.2. Real Example

A dataset concerning China's air quality was used to demonstrate our proposed approach in the subsection. Let $\boldsymbol{X}_p$ be a 31×1 vector of the fine particulate matter ($PM_{2.5}$) in 31 provinces, let $\boldsymbol{X}_s$ be a 31×1 vector of the sulfur dioxide (SO_2) in 31 provinces, let $\boldsymbol{X}_c$ be a 31×1 vector of the carbon monoxide (CO) in 31 provinces and let $\boldsymbol{Y} = (y_1, \ldots, y_{31})^T$ be a 31×1 vector of the air quality index (AQI) in 31 provinces.

Firstly, the above data were exploited to illustrate the application of the Bayesian local influence analysis method. The following three perturbation options were considered: (i) perturbations to covariates; (ii) perturbations to response variables; (iii) covariates, prior distributions of $\boldsymbol{\beta}$ and sampling distributions of $\boldsymbol{\varepsilon}$ are perturbed simultaneously. Bayesian local influence measures $\boldsymbol{FIC}_{B,e_j}$, $\boldsymbol{SIC}_{D_\phi,e_j}$ and $\boldsymbol{SIC}_{M_d,e_j}$ were calculated for three different perturbation scenarios. The results are shown in Figures 6–8. The results of Figure 6a–c show that cases 6 and 19 were diagnosed as influence points by $\boldsymbol{FIC}_{B,e_j}$, $\boldsymbol{SIC}_{D_\phi,e_j}$ and $\boldsymbol{SIC}_{M_d,e_j}$ in perturbation scheme (i). The results of Figure 7a–c show that cases 1 and 12 were diagnosed as influence points by $\boldsymbol{FIC}_{B,e_j}$, $\boldsymbol{SIC}_{D_\phi,e_j}$ and $\boldsymbol{SIC}_{M_d,e_j}$ in perturbation scheme (ii). The results of Figure 8a,b show that cases 6 and 19 were diagnosed as influence points by $\boldsymbol{FIC}_{B,e_j}$, $\boldsymbol{SIC}_{D_\phi,e_j}$ and $\boldsymbol{SIC}_{M_d,e_j}$ in perturbation scheme (iii). The results of Figure 8b show that the prior distribution of $\boldsymbol{\beta}$, the distributions of ε were diagnosed to have a significant effect by $\boldsymbol{SIC}_{D_\phi,e_j}$. The results of Figure 8c showed that the prior distribution of $\boldsymbol{\beta}$ was diagnosed to have a significant effect by $\boldsymbol{SIC}_{M_d,e_j}$.

Secondly, the above data are used to illustrate the application of the Bayesian case influence analysis method. We calculated Bayesian case influence measures $D_\phi(S)$, $CM(S)$ and $CP(S)$. The results are presented in Figures 9 and 10. Figure 9a–c demonstrate that cases 6 and 19 were diagnosed as influence points. Figure 10a–c demonstrate that cases 1 and 12 were diagnosed as influence points. As we expected, the results are the same as the previous diagnosis of Bayesian local influence measures.

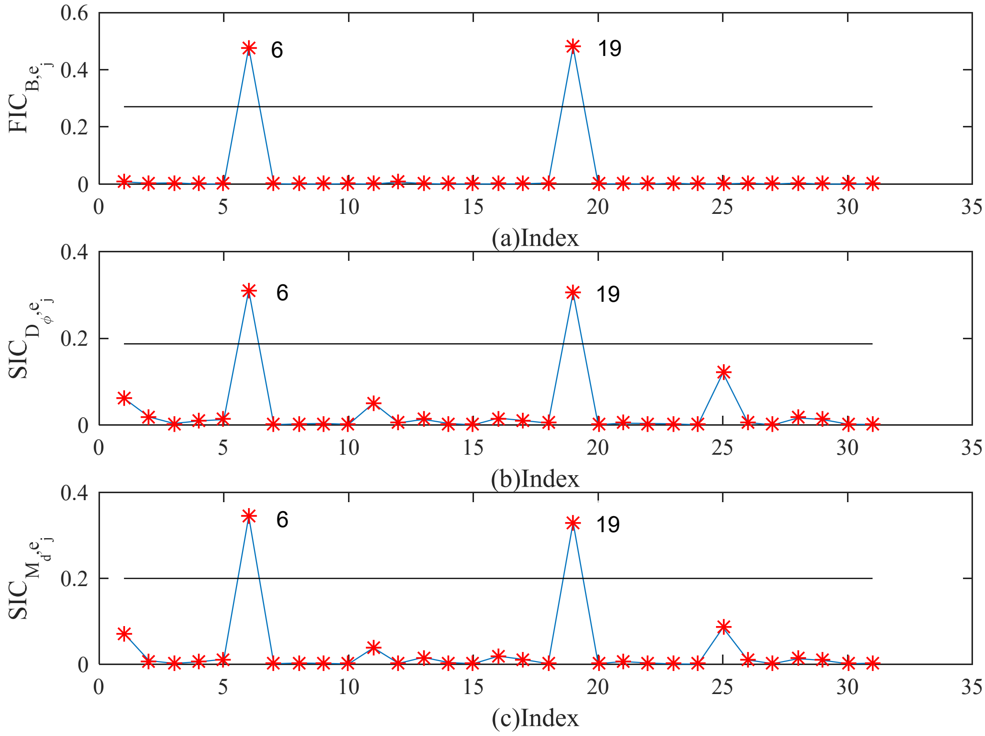

Figure 6. Plots of covariate perturbation in the real example.

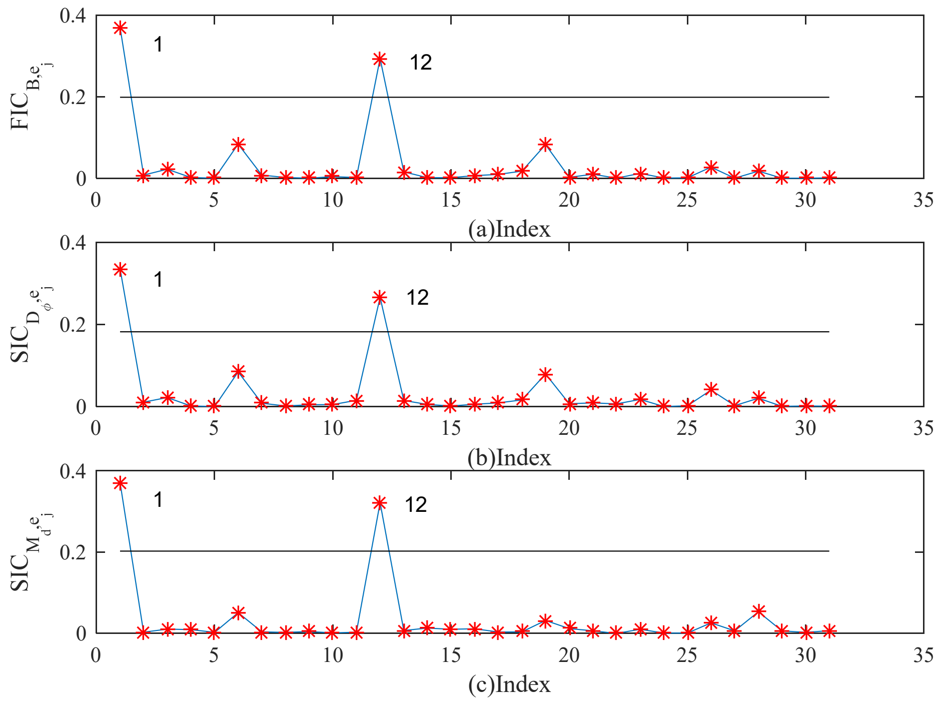

Figure 7. Plots of response variables perturbation in the real example.

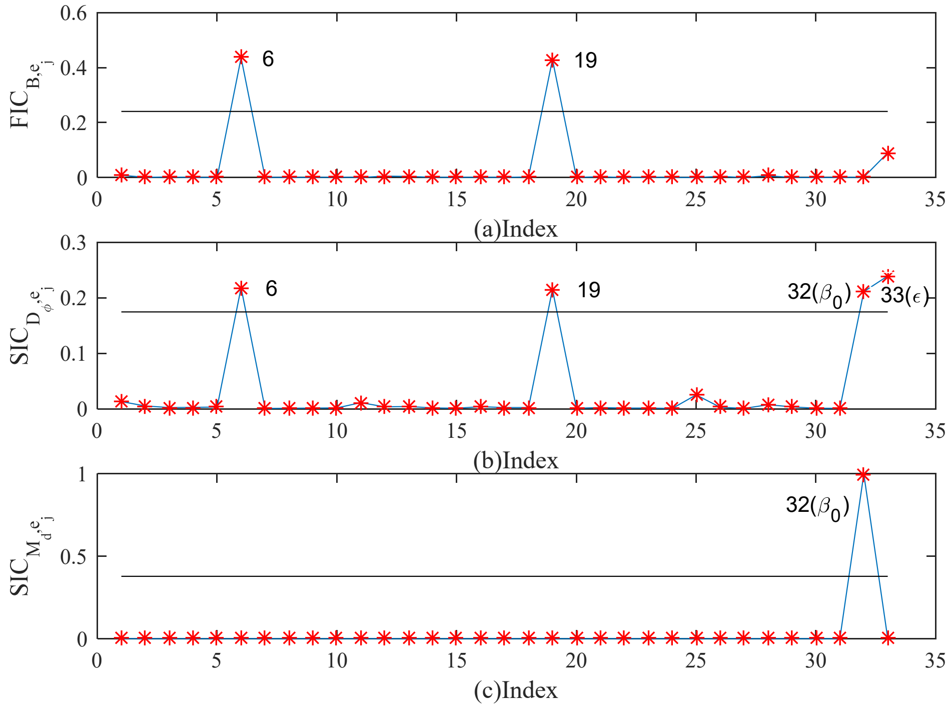

Figure 8. Plots of simultaneous perturbation in the real example.

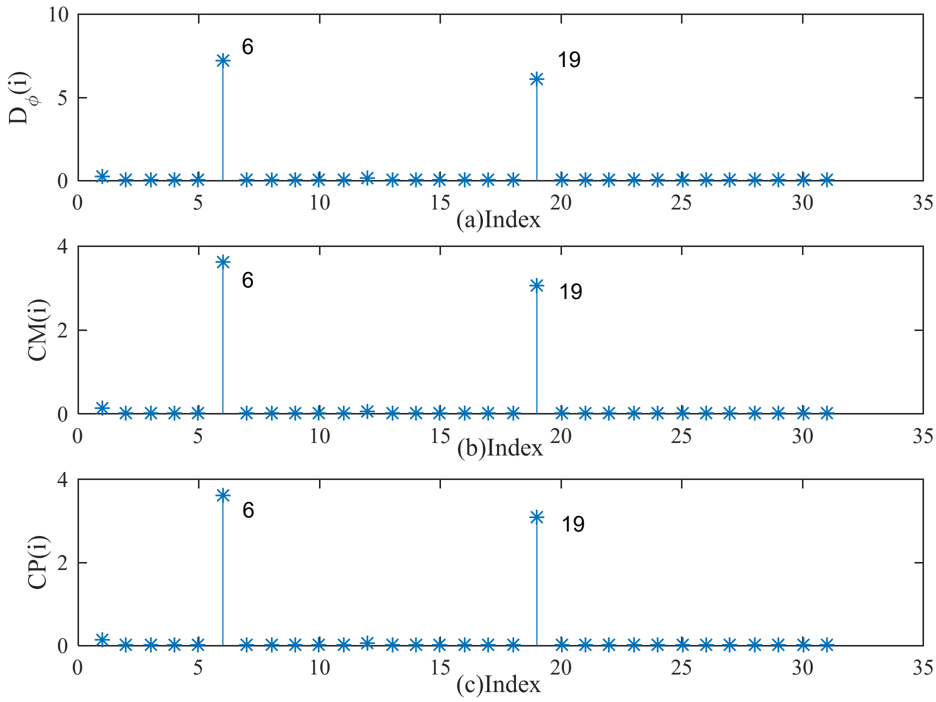

Figure 9. Diagnostic plots of covariate outliers in the real example.

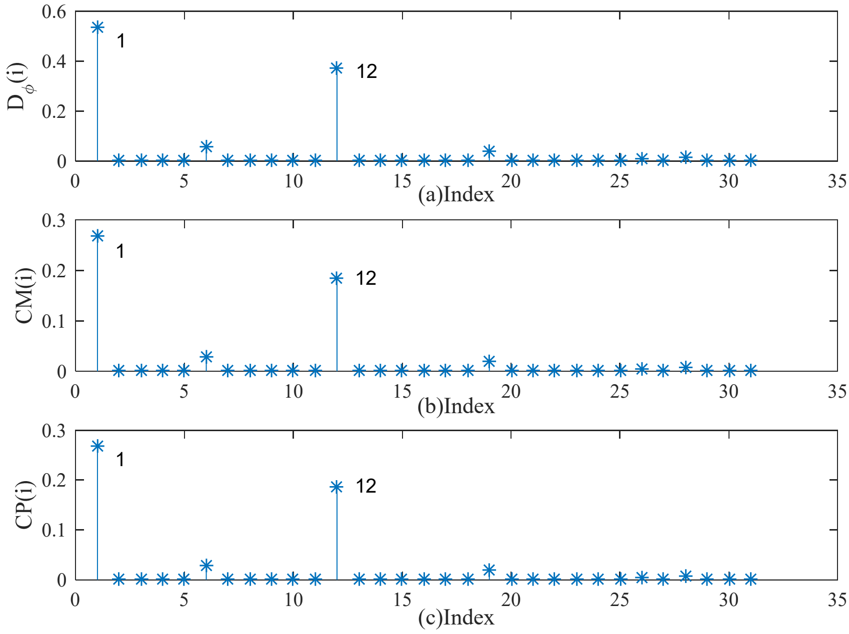

Figure 10. Diagnostic plots of response variables outliers in the real example.

Finally, by reviewing China's air quality data for April 2020, it was found that cases 6 and 19's influence points represent Harbin and Shenyang, respectively. In particular, Harbin's air quality is the second-lowest among the 31 provincial capitals in China, while Shenyang's air quality is the third-lowest among the 31 provincial capitals in China. This indicates that the air quality situation in these two cities is more prominent among the 31 provinces. Similarly, it was found that cases 1 and 12 of the diagnosed impact points represented Beijing and Lhasa, respectively. Among them, Beijing's air quality ranks 17th among 31 provincial capital cities. The air quality of Lhasa ranks fourth among the 31 provinces, indicating that the air quality of Lhasa is indeed outstanding among the 31 provincial capital cities. Although Beijing's air quality ranks in the middle of the 31 provincial capitals, there are many sandstorms in Beijing in April, which will lead to abnormal air quality in Beijing. It is further illustrated that the impact point diagnosed is consistent with the relevant air quality reality.

To further illustrate that the observed values of samples 6 and 19 may be the outliers, the parameter estimates before and after deleting the 6th and 19th observations are calculated, respectively, as shown in Table 1. Table 1 show that after deleting the 6th and 19th observations, the estimators of $\beta_1, \beta_2, \beta_3, \rho, \sigma_\varepsilon^2, \delta_\varepsilon$ are changed from 1.1988, 1.9275, 1.1440, 0.1795, 1.0437, 0.0979 to 1.3440, 2.1451, 1.2767, 0.4755, 1.3488, 0.0000, respectively. Similarly, to further illustrate that the observed values of samples 1 and 12 are the outliers, the parameter estimates before and after deleting the 1st and 12th observations are calculated, respectively, as shown in Table 2. Table 2 show that after deleting the 1st and 12th observations, the estimators of $\beta_1, \beta_2, \beta_3, \rho, \sigma_\varepsilon^2, \delta_\varepsilon$ are changed from 1.1988, 1.9275, 1.1440, 0.1795, 1.0437, 0.0979 to 0.9898, 1.5838, 0.9491, 0.2086, 3.2987, 0.1989, respectively.

Table 1. Bayesian estimation with or without influence point 6 and 19 in the real example.

Par.	**With**		**Without**	
	Est.	**SD.**	**Est.**	**SD.**
β_1	1.1988	0.0315	1.3440	0.0311
β_2	1.9275	0.0320	2.1451	0.0315
β_3	1.1440	0.0315	1.2767	0.0313
ρ	0.1795	0.1348	0.4755	0.2005
σ_ε^2	1.0437	0.1023	1.3488	0.1430
δ_ε	0.0979	0.0076	0.0000	0.0000

Table 2. Bayesian estimation with or without influence point 1 and 12 in the real example.

Par.	With		Without	
	Est.	SD.	Est.	SD.
β_1	1.1988	0.0315	0.9898	0.0316
β_2	1.9275	0.0320	1.5838	0.0314
β_3	1.1440	0.0315	0.9491	0.0315
ρ	0.1795	0.1348	0.2086	0.1701
σ_ε^2	1.0437	0.1023	3.2987	0.7886
δ_ε	0.0979	0.0076	0.1989	0.0019

6. Discussion

Based on SSARMs with response variables obeying skew-normal distribution, this paper proposed a Bayesian local influence method to evaluate the impact of small perturbations in data, prior distribution and sample distribution. Perturbation models with separate or simultaneous perturbations of covariates, response variables, parameter prior and sample distributions were developed. A Bayesian perturbation manifold was constructed to describe the internal structure of perturbation models and quantify the perturbation degree of different perturbation models. The Bayesian local influence measures, including the Bayes factor, the ϕ-distance and the posterior mean distance, were used to evaluate the impact of various perturbations. In addition, three Bayesian case deletion influence measures, including ϕ-distance, Cook's posterior mean distance and Cook's posterior mode distance, were proposed to identify potential outliers in skew-normal spatial autoregression models. The effectiveness of our proposed approach was illustrated by three simulation studies and a real example. The results showed that: (1) our proposed Bayesian local influence approach can effectively identify the potential influence points, misspecified prior distribution and misspecified sampling distribution; (2) our proposed Bayesian case influence approach can be used to effectively detect the potential influence observations; (3) the outliers detected by the Bayesian local influence approach and Bayesian case influence approach are consistent, which further explains the rationality of the two methods.

Author Contributions: Conceptualization, Y.J. and Y.Y.; methodology, Y.J. and M.H.; software, Y.Y. and M.H.; validation, Y.J., L.D. and L.W.; project management, Y.Y. and L.D.; investigation, Y.Y. and M.H.; Preparation of the original work draft, Y.J. and M.H.; revision, Y.J. and L.W.; visualization, M.H.; supervision, funding acquisition, Y.J., L.W., Y.J. and L.W. All authors have read and agreed to the published version of the manuscript.

Funding: This study was funded by the National Natural Science Foundation of China, grant number 11861041; by the Natural Science Foundation of Kunming University of Science and Technology, grant number KKSY201907003; and by the Postdoctoral Directional Training Funding of Yunnan Province, grant number 109820210027.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: This research data can be found at https://www.mee.gov.cn/ (accessed on 8 May 2021).

Acknowledgments: Sincere thanks to everyone who suggested revisions and improved this article.

Conflicts of Interest: No conflict of interest among all authors.

Appendix A

To obtain Bayesian parameter estimates in the SSARMs, a random sequence of observations was produced from the joint posterior distribution $p(\rho, \boldsymbol{\beta}, \sigma_\varepsilon^2, \delta_\varepsilon | \boldsymbol{Y}, \boldsymbol{X})$ using the MCMC algorithm combining Gibbs sampling and the MH algorithm. The sampler was implemented by iteratively drawing observations from the conditional distributions $p(\rho|\boldsymbol{Y},\boldsymbol{\beta},\sigma_\varepsilon^2,\delta_\varepsilon,\mathbf{R}_\varepsilon)$, $p(\beta|\boldsymbol{Y},\rho,\sigma_\varepsilon^2,\delta_\varepsilon,\mathbf{R}_\varepsilon)$, $p(\sigma_\varepsilon^{-2}|\boldsymbol{Y},\rho,\boldsymbol{\beta},\delta_\varepsilon)$, $p(\delta_\varepsilon|\boldsymbol{Y},\rho,\boldsymbol{\beta},\sigma_\varepsilon^2,\mathbf{R}_\varepsilon)$ and $p(\boldsymbol{R}_\varepsilon|\boldsymbol{Y}_n,\boldsymbol{\beta},\rho,\delta_\varepsilon,\sigma_\varepsilon^2)$.

(1) The posterior distribution density of $\boldsymbol{\beta}$ is

$$p\left(\boldsymbol{\beta}\middle|\boldsymbol{Y},\rho,\delta_\varepsilon,\sigma_\varepsilon^2,\mathbf{R}_\varepsilon\right)\sim N(\boldsymbol{\mu}_\beta,\boldsymbol{\Omega}_\beta),$$

where, $\boldsymbol{\Omega}_\beta=(\sigma_\varepsilon^{-2}\mathbf{X}^T\mathbf{X}+\boldsymbol{M}_\beta^{-1})^{-1}$, $\boldsymbol{\mu}_\beta=\boldsymbol{\Omega}_\beta\left(\boldsymbol{M}_\beta^{-1}\boldsymbol{\beta}_0+\sigma_\varepsilon^{-2}\mathbf{X}^T\boldsymbol{\Lambda}_1\right)$ and $\boldsymbol{\Lambda}_1=\boldsymbol{Y}-\rho\boldsymbol{W}\boldsymbol{Y}-\delta_\varepsilon\boldsymbol{R}_\varepsilon$.

(2) The posterior distribution density of δ_ε is

$$p\left(\boldsymbol{\beta}\middle|\boldsymbol{Y},\rho,\delta_\varepsilon,\sigma_\varepsilon^2,\mathbf{R}_\varepsilon\right)\sim N(\boldsymbol{\mu}_\beta,\boldsymbol{\Omega}_\beta),$$

where, $\boldsymbol{\Omega}_{\delta_\varepsilon}=(\sigma_\varepsilon^{-2}\boldsymbol{R}_\varepsilon^T\boldsymbol{R}_\varepsilon+\boldsymbol{M}_{\delta_\varepsilon}^{-1})^{-1}$, $\boldsymbol{\mu}_{\delta_\varepsilon}=\boldsymbol{\Omega}_{\delta_\varepsilon}\left(\boldsymbol{M}_{\delta_\varepsilon}^{-1}\delta_{\varepsilon 0}+\sigma_\varepsilon^{-2}\boldsymbol{R}_\varepsilon^T\boldsymbol{\Lambda}_2\right)$ and $\boldsymbol{\Lambda}_2=\boldsymbol{Y}-\rho\boldsymbol{W}\boldsymbol{Y}-\boldsymbol{X}^T\boldsymbol{\beta}$.

(3) The posterior distribution density of $\boldsymbol{R}_\varepsilon$ is

$$p\left(\boldsymbol{R}_\varepsilon\middle|\boldsymbol{Y},\boldsymbol{\beta},\rho,\delta_\varepsilon,\sigma_\varepsilon^2\right)\sim N(\boldsymbol{\mu}_{\boldsymbol{R}_\varepsilon},\boldsymbol{\Omega}_{\boldsymbol{R}_\varepsilon})$$

where, $\boldsymbol{\Omega}_{\boldsymbol{R}_\varepsilon}=(\sigma_\varepsilon^{-2}\delta_\varepsilon^T\delta_\varepsilon+\boldsymbol{I}_n)^{-1}$ and $\boldsymbol{\mu}_{\boldsymbol{R}_\varepsilon}=\boldsymbol{\Omega}_{\boldsymbol{R}_\varepsilon}(\sigma_\varepsilon^{-2}\delta_\varepsilon^T\boldsymbol{\Lambda}_2)$.

(4) The posterior distribution density of σ_ε^{-2} is

$$p\left(\sigma_\varepsilon^{-2}|\boldsymbol{Y},\boldsymbol{\beta},\rho,\delta_\varepsilon,\mathbf{R}_\varepsilon\right)\sim\Gamma\left(\frac{n}{2}+v,\frac{1}{2}\boldsymbol{e}^T\boldsymbol{e}+s\right)$$

(5) The posterior distribution density of ρ is

$$p\left(\rho\middle|\boldsymbol{Y},\boldsymbol{\beta},\delta_\varepsilon,\sigma_\varepsilon^2,\mathbf{R}_\varepsilon\right)\propto|\boldsymbol{I}_n-\rho\boldsymbol{W}|\exp\left\{-\frac{1}{2}\boldsymbol{e}^T(\sigma_\varepsilon^2\boldsymbol{I}_n)^{-1}\boldsymbol{e}\right\}$$

References

1. LeSage, J.; Pace, R.K. *Introduction to Spatial Econometrics*; Chapman and Hall: London, UK, 2009.
2. Piribauer, P.; Cuaresma, J.C. Bayesian variable selection in spatial autoregressive models. *Spat. Econ. Anal.* **2016**, *11*, 457–479. [CrossRef]
3. Xie, T.F.; Cao, R.Y.; Du, J. Variable selection for spatial autoregressive models with a diverging number of parameters. *Stat. Pap.* **2018**, *61*, 1125–1145. [CrossRef]
4. Du, J.; Sun, X.Q.; Cao, R.Y.; Zhang, Z.Z. Statistical inference for partially linear additive spatial autoregressive models. *Spat. Stat.-Neth.* **2018**, *25*, 52–67. [CrossRef]
5. Xie, L.; Wang, X.R.; Cheng, W.H.; Tang, T. Variable selection for spatial autoregressive models. *Commun. Stat.-Theor. Methods* **2019**, *50*, 1325–1340. [CrossRef]
6. Jay, M.; Ver, H.; Erin, E.; Peterson, M.B.; Hooten, E.M.; Hanks, M.J.F. Spatial autoregressive models for statistical inference from ecological data. *Ecol. Monogr.* **2018**, *88*, 36–59.
7. Anik, A.; Bambang, W.O.; Purhadi, P.; Sutikno, S. Lagrange multiplier test for spatial autoregressive model with latent variables. *Symmetry* **2020**, *12*, 1375.
8. Song, Y.Q.; Liang, X.J.; Zhu, Y.J.; Lin, L. Robust variable selection with exponential squared loss for the spatial autoregressive model. *Comput. Stat. Data Anal.* **2021**, *155*. [CrossRef]
9. Pereira, M.A.A.; Russo, C.M. Nonlinear mixed-effects models with scale mixture of skew-normal distributions. *J. Appl. Stat.* **2019**, *46*, 1602–1620. [CrossRef]
10. Yin, J.H.; Wu, L.C.; Dai, L. Variable selection in finite mixture of regression models using the skew-normal distribution. *J. Appl. Stat.* **2020**, *47*, 2941–2960. [CrossRef]
11. Tatsuya, K.; Strawderman, W.E.; Ryota, Y. Shrinkage estimation of location parameters in a multivariate skew-normal distribution. *Commun. Stat.-Theor. Methods* **2020**, *49*, 2008–2024.
12. Liu, Y.H.; Mao, G.; Leiva, V.; Liu, S.Z.; Alejandra, T. Diagnostic analytics for an autoregressive model under the skew-normal distribution. *Mathematics* **2020**, *8*, 693. [CrossRef]
13. Teimouri, M. EM algorithm for mixture of skew-normal distributions fitted to grouped data. *J. Appl. Stat.* **2021**, *48*, 1154–1179. [CrossRef]

14. Zhu, H.T.; Ibrahim, J.G.; Tang, N.S. Bayesian influence analysis: A geometric approach. *Biometrika* **2011**, *98*, 307–323. [CrossRef] [PubMed]
15. Zhang, Y.Q.; Tang, N.S. Bayesian local influence analysis of general estimating equations with nonignorable missing data. *Comput. Stat. Data Anal.* **2017**, *105*, 184–200. [CrossRef]
16. Ouyang, M.; Yan, X.D.; Chen, J.; Tang, N.S.; Song, X.Y. Bayesian local influence of semiparametric structural equation models. *Comput. Stat. Data Anal.* **2017**, *111*, 102–115. [CrossRef]
17. Dai, X.W.; Jin, L.B.; Tian, M.Z.; Shi, L. Bayesian local influence for spatial autoregressive models with heteroscedasticity. *Stat. Pap.* **2019**, *60*, 1423–1446. [CrossRef]
18. Ju, Y.Y.; Tang, N.S.; Li, X.X. Bayesian local influence analysis of skew-normal spatial dynamic panel data models. *J. Stat. Comput. Sim.* **2018**, *88*, 2342–2364. [CrossRef]
19. Vicente, G.; Cancho, D.K.D.; Victor, H.; Lachos, M.G.A. Bayesian nonlinear regression models with scale mixtures of skew-normal distributions: Estimation and case influence diagnostics. *Comput. Stat. Data Anal.* **2010**, *55*, 588–602.
20. Zhu, H.T.; Joseph, G.I.; Cho, Y.; Tang, N.S. Bayesian case influence measures for statistical models with missing data. *J. Comput. Graph. Stat.* **2012**, *21*, 253–271. [CrossRef]
21. Tang, N.S.; Duan, X.D. Bayesian influence analysis of generalized partial linear mixed models for longitudinal data. *J. Multivar. Anal.* **2014**, *126*, 86–99. [CrossRef]
22. Hao, H.X.; Lin, J.G.; Wang, H.X.; Huang, X.F. Bayesian case influence analysis for GARCH models based on Kullback–Leibler divergence. *J. Korean Stat. Soc.* **2016**, *45*, 595–609. [CrossRef]
23. Duan, X.D.; Fung, W.F.; Tang, N.S. Bayesian semiparametric reproductive dispersion mixed models for non-normal longitudinal data: Estimation and case influence analysis. *J. Stat. Comput. Sim.* **2017**, *87*, 1925–1939. [CrossRef]
24. Arellano-Valle, R.B.; Bolfarine, H.; Lachos, V.H. Bayesian inference for skew-normal linear mixed models. *J. Appl. Stat.* **2007**, *34*, 663–682. [CrossRef]
25. Poon, W.Y.; Poon, Y.S. Conformal normal curvature and assessment of Local Influence. *J. R. Stat. Soc. B* **1999**, *61*, 51–61. [CrossRef]
26. Cook, R.D.; Weisberg, S. Residuals and influence regression. *Biometrics* **1982**, *39*, 413–415.
27. Weiss, R.E.; Cook, R.D. A graphical case statistic for assessing posterior influence. *Biometrika* **1992**, *79*, 51–55. [CrossRef]