Computational Statistics
https://doi.org/10.1007/s00180-024-01504-2

ORIGINAL PAPER

Bayesian diagnostics in a partially linear model with first-order autoregressive skew-normal errors

Yonghui Liu[1] · Jiawei Lu[1] · Gilberto A. Paula[2] · Shuangzhe Liu[3]

Received: 1 October 2023 / Accepted: 28 April 2024

Abstract
This paper studies a Bayesian local influence method to detect influential observations in a partially linear model with first-order autoregressive skew-normal errors. This method appears suitable for small or moderate-sized data sets ($n = 200{\sim}400$) and overcomes some theoretical limitations, bridging the diagnostic gap for small or moderate-sized data in classical methods. The MCMC algorithm is employed for parameter estimation, and Bayesian local influence analysis is made using three perturbation schemes (priors, variances, and data) and three measurement scales (Bayes factor, ϕ-divergence, and posterior mean). Simulation studies are conducted to validate the reliability of the diagnostics. Finally, a practical application uses data on the 1976 Los Angeles ozone concentration to further demonstrate the effectiveness of the diagnostics.

Keywords Bayesian local influence method · Gibbs algorithm · Matrix differential calculus · Time series model

1 Introduction

In regression models, the assumption of independent and identically distributed errors is often made. However, e.g. Seber and Wild (1989) proposed that this assumption is unrealistic for time series data, where errors often exhibit serial correlation. Autoregressive (AR) models are commonly used to address this issue in time series analysis. Additionally, real-world data may not follow a symmetric

✉ Shuangzhe Liu
shuangzhe.liu@canberra.edu.au

[1] School of Statistics and Information, Shanghai University of International Business and Economics, Shanghai, China

[2] Institute of Mathematics and Statistics, University of São Paulo, São Paulo, Brazil

[3] Faculty of Science and Technology, University of Canberra, Canberra, Australia

Published online: 11 July 2024

distribution, and the skew-normal (SN) distribution proposed by Azzalini (1985) can be used to analyze asymmetric data.

Partially linear models (PLMs), widely used in various areas, combine both linear and non-linear components. Different methods, such as kernel smoothing and penalized splines, can be used to handle the non-linear components in these models. For example, a semiparametric model for longitudinal data in HIV seroconverters is proposed by Zeger and Diggle (1994), while a semiparametric regression model for air pollution time series is improved by Dominici et al. (2004) using generalized additive forms. Recent studies on PLMs are made by Ferreira et al. (2013); Ferreira and Paula (2017); Oliveira and Paula (2021); Cardozo et al. (2022), and Ferreira et al. (2022a, 2022b).

On the other hand, local influence analysis is a valuable technique for statistical diagnostics and identifying influential observations. It allows us to scrutinize the impact of perturbations on the model fitting process. This method has been extensively explored in previous research, particularly in the context of regression and time series models where normal or elliptical distributions are assumed by e.g. Cook (1986); Galea et al. (1997); Liu (2000, 2004) and Liu et al. (2022a). Furthermore, Paula et al. (2012); Ferreira and Paula (2017) and Ferreira et al. (2022b) and others have demonstrated the effectiveness of the method, by examining various alternatives. Ferreira and Paula (2017); Ferreira et al. (2022a, 2022b) extended the local influence analysis to PLMs with autoregressive SN errors. Liu et al. (2017, 2020, 2022b, 2023a, 2023b) delved into the application of local influence analysis in autoregressive models, when considering SN distributions as well.

While traditional local influence analysis has been successful in many models, it has limitations in terms of computational intensity, theorem proofs, and estimation challenges in PLMs with non-Gaussian distributions. To address these limitations, Bayesian methods have gained attention in statistical modeling tasks, offering advantages such as overcoming difficult theorems, reducing computational complexity, and being suitable for moderate-sized datasets. Bayesian local influence techniques, including a general theory proposed by Zhu et al. (2011), have been enriched. Tang and Duan (2014) applied Bayesian local influence analysis to generalized partial linear mixed models for longitudinal data, while Dai et al. (2019) discussed its application to spatial autoregressive models assuming a normal distribution. Ju et al. (2022) extended the study by Dai et al. (2019) to the context of a SN distribution.

In this paper, we aim to provide statistical diagnostic methods in Bayesian local influence analysis for PLMs with first-order autoregressive SN errors for small or moderate-sized data sets. We construct three scales of Bayesian perturbation manifold internal structures for perturbations of priors, variances, and data. Using the principles of local influence analysis, we conduct statistical diagnostics of the model. With the introduction of concepts such as prior and posterior distributions, Bayesian methods can successfully compensate for the limitations of classical methods for small- or moderate-sized data sets. In addition, Bayesian methods have some very good advantages: they skip cumbersome theorem proofs and reduce computational complexity. Finally, in the empirical sanalysis, we once again prove our point, obtaining results that are almost the same as those of classical methods.

Moreover, based on the previously obtained results of classical methods, we accurately detect an influential observation and explain its cause using relevant knowledge in geography.

The structure of this paper is as follows: Sect. 2 introduces two representations of the SN distribution and selects the simpler one. Section 3 introduces PLMs with first-order autoregressive SN errors, modeling the non-parametric part to facilitate Bayesian estimation. Section 4 presents Bayesian estimation, including likelihood function calculation, prior distribution specification, detailed posterior distribution calculation, and MCMC implementation. Section 5 introduces the theory of local influence analysis, focusing on three perturbations (priors, variances, and data) and three metrics (Bayes factor, ϕ divergence, and posterior mean) to quantify the influence of perturbations on PLM-SNAR(1). Section 6 presents numerical simulations with a moderate-sized dataset ($n = 200$), while Sect. 7 demonstrates an empirical study to validate the proposed method. Section 8 summarizes the paper and suggests future research directions. The appendix includes additional content related to the model reduction and diagnostic matrices for perturbations involving priors, variances, and data.

2 The SN distribution

If the probability density function (pdf) of the random variable Y has the following form (Azzalini 1985):

$$f(y|\mu,\sigma_A^2,\lambda_A)=\frac{2}{\sigma_A}\,\varphi\Big(\frac{y-\mu}{\sigma_A}\Big)\,\Phi\Big(\frac{\lambda_A(y-\mu)}{\sigma_A}\Big), \tag{1}$$

where $\varphi(*)$ and $\Phi(*)$ represent the pdf and cumulative distribution function (cdf) of the standard normal distribution, respectively, then we refer to Y to follow a univariate SN distribution with location parameter μ, scale parameter σ_A^2, and skewness parameter λ_A, namely $Y\sim \mathrm{SN}(\mu,\sigma_A^2,\lambda_A)$. When $\lambda_A=0$, the distribution becomes a normal distribution.

Lemma 1 (Azzalini 1985) *If* $Y\sim \mathrm{SN}(\mu,\sigma_A^2,\lambda_A)$, *then* $Y\overset{d}{=}\mu+\frac{\lambda_A\sigma_A}{\sqrt{1+\lambda_A^2}}|X_0|+\frac{\sigma_A}{\sqrt{1+\lambda_A^2}}X_1$, *where* X_0 *and* X_1 *are mutually independent standard normal random variables.*

However, there is another representation of the SN distribution. If the pdf of the random variable Y has the following form (Sahu et al. 2003):

$$f(y|\mu,\sigma_S^2,\lambda_S)=\frac{2}{\sqrt{\sigma_S^2+\lambda_S^2}}\,\varphi\Big(\frac{y-\mu}{\sqrt{\sigma_S^2+\lambda_S^2}}\Big)\,\Phi\Big(\frac{\lambda_S}{\sigma_S}\frac{y-\mu}{\sqrt{\sigma_S^2+\lambda_S^2}}\Big), \tag{2}$$

where $\varphi(*)$ and $\Phi(*)$ represent the pdf and cdf of the standard normal distribution, respectively, then we refer to Y to follow a univariate SN distribution with

location parameter μ, scale parameter σ_S^2, and skewness parameter λ_S, namely $Y \sim SN(\mu, \sigma_S^2, \lambda_S)$. When $\lambda_S = 0$, the distribution becomes a normal distribution.

Lemma 2 (Ferreira et al. 2022b) *If* $Y \sim \mathrm{SN}(\mu, \sigma_S^2, \lambda_S)$, *then* $Y \overset{d}{=} \mu + \lambda_S |X_0| + \sigma_S X_1$, $\boldsymbol{E}(Y) = \mu + \sqrt{2/\pi}\lambda_S$, *where* X_0 *and* X_1 *are mutually independent standard normal random variables*.

From Table 1, it is evident that employing the Azzalini (1985) representation of the SN distribution results in a more intricate form of the model's coefficients. For computational ease, we utilize the representation proposed by Sahu et al. (2003) in this paper.

3 PLM-SNAR(1) model

The PLM-SNAR(1) model is defined as follows:

$$\begin{aligned} &Y_i = \mathbf{x}_i{}^\top \boldsymbol{\beta} + g(t_i) + \epsilon_i, \\ &\epsilon_i = \rho \epsilon_{i-1} + e_i, \qquad -1 < \rho < 1, \\ &e_i \overset{iid}{\sim} \mathrm{SN}(-\sqrt{2/\pi}\lambda, \sigma^2, \lambda), \quad i = 1, 2, 3 \ldots n, \end{aligned} \tag{3}$$

where Y_i represents the response variable, $\mathbf{x}_i = (1, x_{i1}, x_{i2}, \ldots, x_{ip})^\top$ represents the $p+1$ linear explanatory variable values, $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \ldots, \beta_p)^\top$ is the $(p+1) \times 1$ vector of coefficients of the linear explanatory variables, t_i represents the nonlinear explanatory variable, $g(\cdot)$ denotes a smoothing function, ϵ_i is the error term of an AR(1) structure with ρ as the AR(1) coefficient, and e_i is the error to follow an iid SN distribution with zero mean, $i = 1, \cdots, n$. Like Ferreira et al. (2022b), we assume $\epsilon_0 = 0$ in this paper.

To simplify the expression of the model, we can represent model (3) in matrix form as follows (the detailed proof can be found in the appendix):

$$\begin{aligned} &Y_i = \mu_i - \sqrt{2/\pi}\lambda + \lambda h_i + \eta_i, \\ &\eta_i \overset{iid}{\sim} \mathrm{N}(0, \sigma^2), \quad i = 1, 2, 3 \ldots n, \end{aligned} \tag{4}$$

where Y_i represents the response variable, $\mu_1 = m_1$, $\mu_i = m_i + \rho(y_{i-1} - m_{i-1})$, $m_i = \mathbf{w}_i{}^\top \boldsymbol{\Lambda} + \mathbf{z}_i^\top \mathbf{b}$, $\mathbf{w}_i = (\mathbf{x}_i^\top, \mathbf{T}_i^\top)^\top$ is a $(p+3) \times 1$ vector, $\mathbf{x}_i = (1, x_{i1}, x_{i2}, \ldots, x_{ip})^\top$, $\mathbf{T}_i = (1, t_i)^\top$, $\boldsymbol{\Lambda} = (\boldsymbol{\beta}^\top, \boldsymbol{\alpha}^\top)^\top$ is a $(p+3) \times 1$ vector, $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \ldots, \beta_p)^\top$,

Table 1 Comparison of two representations

	Stochastic representation
Azzalini	$Y \overset{d}{=} \mu + \frac{\lambda_A \sigma_A}{\sqrt{1+\lambda_A^2}} \lvert X_0 \rvert + \frac{\sigma_A}{\sqrt{1+\lambda_A^2}} X_1$
Sahu	$Y \overset{d}{=} \mu + \lambda_S \lvert X_0 \rvert + \sigma_S X_1$

$\boldsymbol{\alpha} = (\alpha_0, \alpha_1)^\top$, $\mathbf{z}_i$ is the i-th row of the $n \times K$ matrix $\boldsymbol{Z}$, $\boldsymbol{Z} = \boldsymbol{Z}_K \boldsymbol{\Omega}_K^{-\frac{1}{2}}$, $\mathbf{b} = \boldsymbol{\Omega}_K^{\frac{1}{2}} \boldsymbol{u}$ is a $K \times 1$ vector, $\boldsymbol{u} = (u_1, u_2, \ldots, u_K)^\top$, $\boldsymbol{Z}_K$ is an $n \times K$ matrix with i-th row $\boldsymbol{Z}_{Ki} = \{|t_i - \kappa_1|^3, \ldots, |t_i - \kappa_K|^3\}$, $\boldsymbol{\Omega}_K$ is a $K \times K$ penalty coefficient matrix with k-th row $\boldsymbol{\Omega}_{Kk} = \{|\kappa_k - \kappa_1|^3, \ldots, |\kappa_k - \kappa_K|^3\}$ (to avoid overfitting), $h_i = |h_{0i}|$, $\eta_i = \sigma h_{1i}$, h_{0i} and h_{1i} are mutually independent standard normal random variables, n is the sample size and K is the number of knots.

4 Bayesian implementation

4.1 Likelihood function

From model (4), we can get

$$\begin{aligned} &Y_1 | \boldsymbol{\Lambda}, \mathbf{b}, \sigma^2, \lambda, h_1 \sim \mathrm{N}(m_1 - \sqrt{2/\pi}\lambda + \lambda h_1, \sigma^2), \\ &Y_i | y_{i-1}, \boldsymbol{\Lambda}, \mathbf{b}, \rho, \sigma^2, \lambda, h_i \sim \mathrm{N}(\mu_i - \sqrt{2/\pi}\lambda + \lambda h_i, \sigma^2), \quad i = 2, 3, \ldots, n. \end{aligned} \tag{5}$$

We can observe that the sequence $\{y_i\}$ $(i = 1, \cdots, n)$ has the Markov property from the above model. Therefore, we can express the likelihood function for an observed sample $\mathbf{y} = (y_1, y_2, \ldots, y_n)^\top$ by

$$\begin{aligned} L(\boldsymbol{\theta}|\mathbf{y}) &= f(y_1 | m_1, \sigma^2, \lambda, h_1) \prod_{i=2}^{n} f(y_i | y_{i-1}, y_{i-2}, \ldots, y_1, \mu_i, \sigma^2, \lambda, h_i) \\ &= f(y_1 | m_1, \sigma^2, \lambda, h_1) \prod_{i=2}^{n} f(y_i | y_{i-1}, \mu_i, \sigma^2, \lambda, h_i) \\ &= (\sqrt{2\pi})^{-n} \exp\left(-\frac{1}{2\sigma^2} [(y_1 - m_1 + \sqrt{2/\pi}\lambda - \lambda h_1)^2 + \sum_{i=2}^{n} (y_i - \mu_i + \sqrt{2/\pi}\lambda - \lambda h_i)^2] \right), \end{aligned}$$

where $\boldsymbol{\theta} = (\boldsymbol{\Lambda}^\top, b^\mathrm{T}, \sigma^2, \lambda, \rho)^\top$.

4.2 Prior distributions

In Bayesian statistics, specifying prior distributions for the parameters is crucial. These prior distributions reflect our prior beliefs about the parameters and can lead to more accurate parameter estimation. To refine the Bayesian model, we partition the parameter $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \boldsymbol{\alpha}^\top, b^\mathrm{T}, \sigma^2, \lambda, \rho)^\top$ into two components and specify separate prior distributions for each component - one for the parametric part and another for the non-parametric part.

4.2.1 Parametric component

For setting prior distributions for the linear part of the model, a popular approach is to use conjugate prior distributions. In our case, following Ferreira et al. (2013), we choose to reparameterize ρ with $2\rho_0 - 1$, where $\rho_0 \in (0, 1)$ follows a beta distribution. Specifying the prior distribution for the coefficient ρ_0 poses a challenge.

A uniform distribution is not suitable in this context, as it is more appropriate for random walk processes. Instead, we opt for a beta distribution, which is better suited for fitting autoregressive processes (Marriott and Newbold 1998). We choose the following prior distributions:

$$\begin{aligned}\boldsymbol{\beta} &\sim \mathrm{N}_{p+1}\left(\boldsymbol{\beta_0}, \Omega_0\right),\\ \sigma^2 &\sim \mathrm{IGamma}\left(\frac{q_0}{2}, \frac{q_1}{2}\right),\\ \lambda &\sim \mathrm{N}\left(\mu_\lambda, {\sigma_\lambda}^2\right),\\ \rho_0 &\sim \mathrm{Beta}\left(a_{\rho_0}, b_{\rho_0}\right),\end{aligned}$$

where $\mathbf{N}_p$ denotes a p variate normal distribution, IGamma denotes an inverse gamma distribution, and Beta denotes a beta distribution.

4.2.2 Non-parametric component

Like Crainiceanu et al. (2005), we assume that the coefficients in the additive mixture model are normally distributed

$$\begin{aligned}\boldsymbol{\alpha} &\sim \mathrm{N}_2\left(\mathbf{0}, \Sigma_0\right),\\ \mathbf{b} &\sim \mathrm{N}_{20}\left(\mathbf{0}, {\sigma_b}^2\mathbf{I}_{20}\right),\end{aligned}$$

where $\boldsymbol{\alpha}$ is a 2×1 vector, $\mathbf{b}$ is a 20×1 vector ($K = 20$), Σ_0 is a 2×2 positive definite matrix, $\boldsymbol{I}_{20}$ is a 20×20 identity matrix, and ${\sigma_b}^2$ is a positive constant.

4.3 The implementation of MCMC algorithm

4.3.1 Posterior distribution

Now that we have computed the likelihood function and specified the prior distributions, the next step is to calculate the posterior distributions, so we obtain:

$$\begin{aligned}\boldsymbol{\Lambda}|\boldsymbol{\theta}_{\boldsymbol{\Lambda}}, \mathbf{y}, \mathbf{h} &\sim \mathrm{N}_{p+3}\left(\boldsymbol{A}_{\boldsymbol{\Lambda}}\boldsymbol{\mu}_{\boldsymbol{\Lambda}}, \boldsymbol{A}_{\boldsymbol{\Lambda}}\right),\\ \sigma^2|\boldsymbol{\theta}_{\sigma^2}, \mathbf{y}, \mathbf{h} &\sim \mathrm{IGamma}\left(\frac{q_0+n}{2}, \frac{q_1+l}{2}\right),\\ \lambda|\boldsymbol{\theta}_{\boldsymbol{\lambda}}, \mathbf{y}, \mathbf{h} &\sim \mathrm{N}(\mu_p, \sigma_p^2),\\ \rho_0|\boldsymbol{\theta}_{\boldsymbol{\rho}}, \mathbf{y}, \mathbf{h} &\propto {\rho_0}^{a_{\rho_0}-1}(1-\rho_0)^{b_{\rho_0}-1}\\ \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=2}^{n}[(2\rho_0-1)(y_{i-1}-m_{i-1}) - y_i + m_i\right.&\\ \left. - \sqrt{2/\pi}\lambda + \lambda h_i]\right\},&\\ \mathbf{b}|\boldsymbol{\theta}_{\boldsymbol{b}}, \mathbf{y}, \mathbf{h} &\sim \mathrm{N}_{20}(\boldsymbol{A}_{\boldsymbol{b}}\boldsymbol{\mu}_{\boldsymbol{b}}, \boldsymbol{A}_{\boldsymbol{b}}),\end{aligned}$$

where

$$
\begin{aligned}
\boldsymbol{A}_{\boldsymbol{\Lambda}} &= \left(\Delta_{\boldsymbol{\Lambda}}^{-1} + \sigma^{-2} \left[\mathbf{w}_1 \mathbf{w}_1^{\top} + \sum_{i=2}^{n} (\mathbf{w}_i - \rho \mathbf{w}_{i-1})(\mathbf{w}_i - \rho \mathbf{w}_{i-1})^{\top} \right] \right)^{-1}, \\
\boldsymbol{\mu}_{\boldsymbol{\Lambda}} &= \Delta_{\boldsymbol{\Lambda}}^{-1} \boldsymbol{\Lambda}_0 + \sigma^{-2} \left[\mathbf{w}_1 (\tau_1 + \sqrt{2/\pi}\lambda - \lambda h_1) \right. \\
&\quad \left. + \sum_{i=2}^{n} (\tau_i - \rho \tau_{i-1} + \sqrt{2/\pi}\lambda - \lambda h_i)(\mathbf{w}_i - \rho \mathbf{w}_{i-1}) \right], \\
l &= (\mathbf{w}_1^{\top} \boldsymbol{\Lambda} - \tau_1 - \sqrt{2/\pi}\lambda + \lambda h_1)^2 \\
&\quad + \sum_{i=2}^{n} [(\mathbf{w}_i^{\top} - \rho \mathbf{w}_{i-1}^{\top})\Lambda - \tau_i + \rho \tau_{i-1} - \sqrt{2/\pi}\lambda + \lambda h_i]^2, \\
\mu_p &= \frac{\sigma^2 \mu_\lambda - {\sigma_\lambda}^2 [(y_1 - m_1)(\sqrt{2/\pi} - h_1) + \sum_{i=2}^{n} (y_i - \mu_i)(\sqrt{2/\pi} - h_i)]}{\sigma^2 + {\sigma_\lambda}^2 \sum_{i=1}^{n} (\sqrt{2/\pi} - h_i)}, \\
\sigma_p^2 &= \frac{\sigma^2 {\sigma_\lambda}^2}{\sigma^2 + {\sigma_\lambda}^2 \sum_{i=1}^{n} (\sqrt{2/\pi} - h_i)}, \\
\boldsymbol{A}_{\boldsymbol{b}} &= \left(\sigma_b^{-2} \mathbf{I} + \sigma^{-2} \left[\mathbf{z}_1 \mathbf{z}_1^{\top} + \sum_{i=2}^{n} (\mathbf{z}_i - \rho \mathbf{z}_{i-1})(\mathbf{z}_i - \rho \mathbf{z}_{i-1})^{\top} \right] \right)^{-1}, \\
\boldsymbol{\mu}_{\boldsymbol{b}} &= \sigma^{-2} \left[\mathbf{z}_1 (\xi_1 + \sqrt{2/\pi}\lambda - \lambda h_1) + \sum_{i=2}^{n} (\xi_i - \rho \xi_{i-1} + \sqrt{2/\pi}\lambda - \lambda h_i)(\mathbf{z}_i - \rho \mathbf{z}_{i-1}) \right], \\
\boldsymbol{\Lambda}_{\mathbf{0}} &= (\boldsymbol{\beta}_0^{\top}, 0, 0)^{\top}, \\
\boldsymbol{\Delta}_{\boldsymbol{\Lambda}} &= \operatorname{diag}(\Omega_0, \Sigma_0), \\
\tau_i &= y_i - \mathbf{z}_i^{\top} \mathbf{b}, \\
\xi_i &= y_i - \mathbf{w}_i^{\top} \boldsymbol{\Lambda}, \\
\tau_1 &= y_1 - \mathbf{z}_1^{\top} \mathbf{b}, \\
\xi_1 &= y_1 - \mathbf{w}_1^{\top} \boldsymbol{\Lambda}, \\
\rho &= 2\rho_0 - 1.
\end{aligned}
$$

Note that in this context, we use $\boldsymbol{\theta}_{\text{par}}$ to refer to all parameters in $\boldsymbol{\theta}$ except for *par*. Once the posterior distribution is computed, we can employ it as the target distribution for sampling, thus obtaining estimations for various parameters.

4.3.2 MCMC algorithm

For Bayesian practitioners, the MCMC algorithm stands as a potent tool, allowing the extraction of samples from the posterior distribution for parameter estimation. Moreover, it adeptly tackles computational hurdles, particularly in evaluating

complex integrals, by employing numerical approximations through Monte Carlo numerical integration. In this study, we leverage specific algorithms within the MCMC framework, such as the Metropolis-Hastings (MH) algorithm and Gibbs sampling (Albert 2007; Gelman et al. 2013).

As delineated in Sect. 4.3.1, explicit posterior distributions can be computed for all parameters except ρ_0. Consequently, all parameters except ρ_0 lend themselves to parameter estimation via Gibbs sampling, given that Gibbs sampling mandates explicit posterior distributions for parameters. However, owing to the unattainability of an explicit posterior distribution for ρ_0, the Gibbs sampling method proves impracticable. In this scenario, the MH algorithm emerges as the method of choice for sampling from the undetermined posterior distribution.

To streamline the integration of data and parameters from R into the WinBUGS program for MCMC sampling, we rely on the R package R2WinBUGS, as detailed in Sturtz et al. (2005). This integration facilitates seamless and efficient sampling in our Bayesian analysis (Fig. 1).

Building on the preceding context, we choose to employ the MH algorithm and the Gibbs sampling, a fundamental component of WinBUGS, with our corresponding pseudo-code outlined in Appendix. To streamline the parameter estimation process, we use WinBUGS for a single-step solution. This strategy proves efficient in exploring the target distribution, thereby enhancing the effectiveness of Bayesian analysis and parameter estimation.

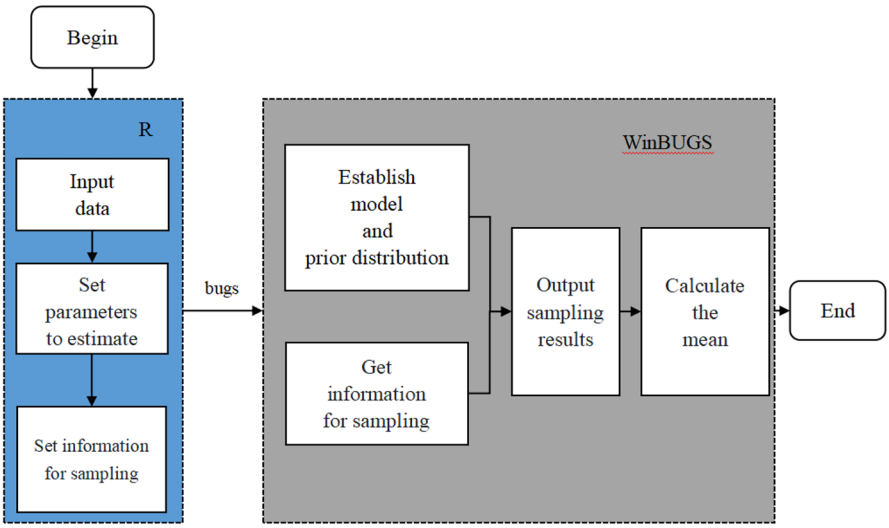


Fig. 1 The operating process of Gibbs sampling using R2winBUGS

5 Bayesian local influence analysis

5.1 Bayesian perturbation model and manifold

Under the assumption of regularity conditions, we can consider the perturbation model $\mathcal{M} = \{p(y, \theta|\boldsymbol{\omega}) : \boldsymbol{\omega} \in R^m\}$ as forming an m-dimensional Riemannian Hilbert manifold (Zhu et al. 2011). Let $l(\boldsymbol{\omega}) = \ln[p(y, \boldsymbol{\theta}|\boldsymbol{\omega})]$, the tangent space $T_{\boldsymbol{\omega}}$ of the manifold $\mathcal{M}$ is formed by all the elements in the tangent vector $\dot{l}(y, \boldsymbol{\theta}|\boldsymbol{\omega}) = \frac{\partial l(\boldsymbol{\omega})}{\partial \boldsymbol{\omega}}$.

Furthermore, under the fulfillment of regularity conditions, $g_{ij}(\boldsymbol{\omega})$, the element at the i-th row and j-th column of the matrix $\mathbf{G}(\boldsymbol{\omega})$, constitutes the measure tensor of $\mathcal{M}$, where

$$g_{ij}(\boldsymbol{\omega}) = E_{\boldsymbol{\omega}}[\partial_{\omega_i} l(\boldsymbol{\omega}) \partial_{\omega_j} l(\omega)] = -E_{\boldsymbol{\omega}}[\partial^2_{\omega_i \omega_j} l(\omega)], \quad i, j = 1, 2, \ldots, m,$$

$$\partial_{\omega_i} l(\boldsymbol{\omega}) = \frac{\partial l(\boldsymbol{\omega})}{\partial \omega_i},$$

and $E_{\boldsymbol{\omega}}[f(*)]$ represents the expectation of $f(*)$ with respect to $p(y, \boldsymbol{\theta}|\boldsymbol{\omega})$.

In $\mathbf{G}(\boldsymbol{\omega})$, the diagonal elements represent the perturbation values of the corresponding elements in $\boldsymbol{\omega}$. The value of $g_{ij}(\boldsymbol{\omega}) / \sqrt{g_{ii}(\boldsymbol{\omega}) g_{jj}(\boldsymbol{\omega})}$ indicates the degree of correlation between ω_i and ω_j. If $\mathbf{G}(\boldsymbol{\omega}^0)$ is a diagonal matrix, it implies that different perturbations are mutually orthogonal, ensuring no interdependence among them. On the other hand, if $\mathbf{G}(\boldsymbol{\omega}^0)$ is not diagonal, it would invalidate the diagnostic results.

To address this issue, we propose another perturbation vector $\tilde{\boldsymbol{\omega}} = \boldsymbol{\omega}^0 + \mathbf{G}(\boldsymbol{\omega}^0)^{\frac{1}{2}}(\boldsymbol{\omega} - \boldsymbol{\omega}^0)$. This ensures that the estimated value of $\mathbf{G}(\tilde{\boldsymbol{\omega}})$ at point $\boldsymbol{\omega}^0$ is represented as $C * \mathbf{I}_m$, where C is a positive constant and $\mathbf{I}_m$ denotes an $m \times m$ identity matrix.

Based on this theoretical foundation and inspired by the work of Hao et al. (2019), we conducted research using three different approaches: prior perturbation, model and prior joint perturbation, and data and prior joint perturbation. These approaches allow us to effectively perform statistical diagnostic analysis in our study.

5.1.1 Perturbation of priors

In Sect. 4.2, we have established the prior distributions for each coefficient. Now, we consider perturbations of the parameters, see Hao et al. (2019) and Appendix.

5.1.2 Perturbation of variances

In the context of Bayesian local influence analysis, it is common to apply a small perturbation to the PLM-SNAR(1) model by modifying its variance σ^2. Like Hao et al. (2019), this is achieved by rescaling the variance using perturbation coefficients, denoted as ω_i, resulting in the new variance $\omega_i^{-1}\sigma^2$, see Appendix. The goal of this

perturbation is to assess the impact of variance changes on the posterior distribution of the PLM-SNAR(1) model.

5.1.3 Perturbation of data

Like Hao et al. (2019), we not only explore variance perturbations but also investigate the impact of data perturbations. The unique aspect of our data perturbation approach is that it preserves the original data structure while introducing perturbations. By employing this method, we can derive an updated posterior distribution and assess the model's sensitivity and robustness to different variations in the input data. Data perturbations in this paper are categorized into the following two types: response variable perturbation and explanatory variable perturbation. This comprehensive analysis allows us to gain valuable insights into how the model responds to different types of data variations and enhances our understanding of its performance under various scenarios (we refer to Appendix for details).

5.2 Local influence measure

Since Cook (1986) introduced the concept of local influence, it has become a powerful tool for detecting influential observations in statistical models. Liu (2004) has previously validated its effectiveness, particularly in time series models. However, with the advancement of Bayesian statistics, traditional methods of local influence analysis may not be directly applicable to Bayesian models. Therefore, adopting the framework of Bayesian local influence analysis is more appropriate in this context.

According to Zhu et al. (2011), $f(\boldsymbol{\omega}) : \mathcal{M} \to \mathcal{R}^l$ represents the objective function, where commonly used objective functions include Bayes factor, ϕ-divergence and posterior mean. For a finite-dimensional manifold $\mathcal{M}$, if $\boldsymbol{\omega}(t)$ is a geodesic on $\mathbf{M}$ with $\boldsymbol{\omega}(0) = \boldsymbol{\omega}^0$ and $\partial_t \boldsymbol{\omega}(t)|_{t=0} = \mathbf{h} \in \mathcal{R}^m$, we can apply Taylor's expansion to obtain $f[\boldsymbol{\omega}(t)] = f[\boldsymbol{\omega}(0)] + f'_{\mathbf{h}}(0)t + o(t^2)$, where $f'_{\mathbf{h}}(0) = \nabla_f^{\top}\mathbf{h}$, and $\nabla_f = \partial_{\boldsymbol{\omega}} f(\boldsymbol{\omega}(0))$. This expansion enables us to assess the impact of small perturbations on the objective function, allowing us to conduct local influence analysis in the Bayesian setting.

5.2.1 First-order influence measure

When $\nabla_f \neq 0$, we consider using the first-order influence measure. The first-order influence measure (*FI*) is defined in the $\mathbf{h} \in \mathcal{R}^m$ direction by Zhu et al. (2011):

$$FI_{f,\mathbf{h}} = FI_{f(\omega(0)),\mathbf{h}} = \frac{\mathbf{h}^{\top}\nabla_f \mathbf{W}_f \nabla_f^{\top}\mathbf{h}}{\mathbf{h}^{\top}\mathbf{G}\mathbf{h}},$$

where $\mathbf{G} = \mathbf{G}(\boldsymbol{\omega}^0)$ and $\mathbf{W}_f$ is a positive semi-definite matrix.

Specifically, for an appropriate perturbation $\tilde{\boldsymbol{\omega}}$, *FI* can also be written as:

$$FI_{f(\tilde{\omega}),\mathbf{h}|\tilde{\omega}=\omega^0} = \frac{\mathbf{h}^\top \mathbf{G}^{-\frac{1}{2}} \nabla_f \mathbf{W}_f \nabla_f^\top \mathbf{G}^{-\frac{1}{2}} \mathbf{h}}{\mathbf{h}^\top \mathbf{h}}.$$

If the value of $FI_{f,\mathbf{h}}$ is larger, it indicates that the perturbation $\tilde{\boldsymbol{\omega}}$ has a greater local influence on the model. This conclusion is based on the fact that the maximum value of $FI_{f,\mathbf{h}}$ is equal to the largest eigenvalue of $\mathbf{G}^{-\frac{1}{2}} \nabla_f \mathbf{W}_f \nabla_f^\top \mathbf{G}^{-\frac{1}{2}}$, and the eigenvector corresponding to the largest eigenvalue of $\mathbf{G}^{-\frac{1}{2}} \nabla_f \mathbf{W}_f \nabla_f^\top \mathbf{G}^{-\frac{1}{2}}$ can be used to assess the robustness of the prior, influential observations, or improper sampling distributions, while also obtaining the worst perturbation direction corresponding to $f(\tilde{\boldsymbol{\omega}})$, the direction in which the objective function is maximally perturbed.

In order to evaluate $FI_{f(\tilde{\omega}),\mathbf{h}}$ objectively, we use the following first-order adjusted influence measure $FIC_{f(\boldsymbol{\omega}_0),\mathbf{h}}$ at $\boldsymbol{\omega}^0$ in an unit direction h :

$$FIC_{f(\boldsymbol{\omega}_0),\mathbf{h}} = \mathbf{h}^\top \mathbf{B} \mathbf{h}.$$

where $\mathbf{B} = \mathbf{Q}/\text{trace}(\mathbf{Q})$, and $\mathbf{Q} = \mathbf{G}^{-\frac{1}{2}} \nabla_f \mathbf{W}_f \nabla_f^\top \mathbf{G}^{-\frac{1}{2}}$. By decomposing matrix B, we can obtain non-zero eigenvalues $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_r$ and their corresponding eigenvectors $\boldsymbol{e_1}, \boldsymbol{e_2}, \ldots, \boldsymbol{e_r}$. The appropriate perturbation $\tilde{\omega}$ is reflected by the largest eigenvalue λ_1, which means we can use the corresponding eigenvector e_1 to detect the most significant perturbation. However, Poon and Poon (1999) proposed that evaluating local influence solely by examining e_1 is not sufficient. To address this limitation, we can use the overall contribution vector of all eigenvectors associated with non-zero eigenvalues to estimate local influence: $M(0) = \lambda_1 \boldsymbol{e_1}^2 + \lambda_2 \boldsymbol{e_2}^2 + \ldots + \lambda_n \boldsymbol{e_r}^2$, where $\boldsymbol{e_i}^2 = (e_{i1}^2, e_{i2}^2, \ldots, e_{in}^2)^\top$. It is easy to find that the j-th component of $M(0)$ is equal to $M(0)_j = \sum_{i=1}^r \lambda_i e_{ij}^2 = FIC_{f(\tilde{\boldsymbol{\omega}}^0)} = b_{jj}$, where b_{jj} is the j-th diagonal element of the matrix B, $j = 1, 2, \ldots, n$. Based on the studies of Zhu and Lee (2001) and Lee and Xu (2004), we use $\bar{M}_0 + c^* * SM(0)$ as a benchmark, where $\bar{M}_0$ and $SM(0)$ are the respective mean and standard error of $M(0)$, and c^* is a selected constant, which may depend on the specific application; $c^* = 2$ is suggested by Zhu and Lee (2001).

A. The commonly used objective function used for first-order influence measure: Bayes factor

Modified from Hao et al. (2019), in the measure of the logarithmic Bayes factor, the distance between $\boldsymbol{\omega}$ and $\boldsymbol{\omega}_0$ can be expressed as:

$$BF(\boldsymbol{\omega}) = \ln[p(Y|\boldsymbol{\omega})] - \ln[p(Y|\boldsymbol{\omega}^0)],$$

where $p(y|\boldsymbol{\omega}) = \int p(y, \boldsymbol{\theta}|\boldsymbol{\omega}) d\boldsymbol{\theta}$. If we set $\mathbf{W}_f = \mathbf{I}$, then we can obtain:

$$\nabla_{BF} = E_{p(y,\boldsymbol{\theta}|\boldsymbol{\omega}^0)}\{\partial_{\boldsymbol{\omega}} \ln[p(y, \boldsymbol{\theta}|\boldsymbol{\omega}^0)]\},$$

$$h_{max}^{BF} = \frac{\mathbf{G}(\boldsymbol{\omega}^0)^{-\frac{1}{2}} \nabla_{BF}}{\sqrt{\nabla_{BF}^\top \mathbf{G}(\boldsymbol{\omega}^0)^{-1} \nabla_{BF}}}.$$

However, the vast majority of integrals in the world are difficult to compute. Therefore, in order to compute ∇_{BF}, we use the method mentioned in Sect. 4.3.2, using Gibbs sampling to estimate the statistic

$$\begin{aligned}\nabla_{BF} &= E_{p(y,\boldsymbol{\theta}|\boldsymbol{\omega}^0)}\{\partial_{\boldsymbol{\omega}}\ln[p(y,\boldsymbol{\theta}|\boldsymbol{\omega}^0)]\}\\ &\approx \frac{1}{T}\sum_{t=1}^{T}\partial_{\omega}\ln[p(y,\boldsymbol{\theta}^{(t)}|\boldsymbol{\omega}^0)],\end{aligned}$$

where $\boldsymbol{\theta}^{(t)}$ is obtained by sampling from the posterior distribution using MCMC algorithm, T is the number of samples.

5.2.2 Second-order influence measure

When $\nabla_f = 0$, we consider using the second-order influence measure. For convenience of notation, we assume that the dimension of $f(\omega)$ is 1, then we use Taylor expansion $f[\boldsymbol{\omega}(t)] = f[\boldsymbol{\omega}(0)] + 0 + \frac{1}{2}f''_{\mathbf{h}}(0)t^2 + o(t^3)$, where $f''_{\mathbf{h}}(0) = \mathbf{h}^\top\mathbf{H}_f\mathbf{h}$ and $\mathbf{H}_f = \partial^2_{\omega}f(\omega(0))$. The second-order influence measure (SI) is defined in the $\mathbf{h} \in \mathcal{R}^m$ direction by Zhu et al. (2011):

$$SI_{f,\mathbf{h}} = SI_{f(\boldsymbol{\omega}(0)),\mathbf{h}} = \frac{\mathbf{h}^\top\mathbf{H}_f\mathbf{h}}{\mathbf{h}^\top\mathbf{G}\mathbf{h}}.$$

Specifically, for an appropriate perturbation $\tilde{\boldsymbol{\omega}}$, it can also be written as:

$$SI_{f(\tilde{\boldsymbol{\omega}})),\mathbf{h}|\tilde{\boldsymbol{\omega}}=\boldsymbol{\omega}^0} = \frac{\mathbf{h}^\top\mathbf{G}^{-\frac{1}{2}}\mathbf{H}_f\mathbf{G}^{-\frac{1}{2}}\mathbf{h}}{\mathbf{h}^\top\mathbf{h}}.$$

Similar to the first-order sensitivity measure, for the computation of the second-order sensitivity measure, only the largest eigenvalue of $\mathbf{G}^{-\frac{1}{2}}\mathbf{H}_f\mathbf{G}^{-\frac{1}{2}}$ and its corresponding eigenvector need to be considered. We can also perform the same transformation as before on it:

$$SIC_{f(\omega_0),\mathbf{h}} = \mathbf{h}^\top\mathbf{B}_s\mathbf{h},$$

where $\mathbf{B}_s = \mathbf{Q}_\mathbf{s}/\text{trace}(\mathbf{Q}_\mathbf{s})$ and $\mathbf{Q}_\mathbf{s} = \mathbf{G}^{-\frac{1}{2}}\mathbf{H}_f\mathbf{G}^{-\frac{1}{2}}$. For the selection of benchmarks, we can still use the same approach as in Sect. 5.2.1.

B. The commonly used objective function used for second-order influence measure: ϕ -divergence

Modified from Hao et al. (2019), the ϕ-divergence corresponding to the posterior distributions before and after perturbation ω is defined as follows:

$$D_{\phi}(\boldsymbol{\omega}) = \int \phi[R(\boldsymbol{\theta}|\boldsymbol{\omega})]p(\boldsymbol{\theta}|y)d\boldsymbol{\theta},$$

where $R(\theta|\boldsymbol{\omega}) = \frac{p(\boldsymbol{\theta}|y,\boldsymbol{\omega})}{p(\boldsymbol{\theta}|y)}$, $\phi(\cdot)$ is a convex function, and $\phi(1) = 0$. If $\phi(\cdot)$ is the logarithmic function $\ln(\cdot)$, the ϕ-divergence degenerates into the Kullback–Leibler (K–L) divergence, then

$$\begin{aligned}\mathbf{H}_{\phi} &= \phi''(1)E_{\boldsymbol{\omega}^0}\{\partial_{\boldsymbol{\omega}}\ln[p(\boldsymbol{\theta}|y,\boldsymbol{\omega}^0)]\}^{\otimes 2}\\ &= \phi''(1)E_{\boldsymbol{\omega}^0}\{\partial_{\boldsymbol{\omega}}\ln[p(y,\boldsymbol{\theta}|\boldsymbol{\omega}^0)] - E_{\boldsymbol{\omega}^0}\{\partial_{\boldsymbol{\omega}}\ln[p(y,\boldsymbol{\theta}|\boldsymbol{\omega}^0)]\}\}^{\otimes 2}\\ &= \phi''(1)\big[E_{\boldsymbol{\omega}^0}\{\partial_{\boldsymbol{\omega}}\ln[p(y,\boldsymbol{\theta}|\boldsymbol{\omega}^0)]\}^{\otimes 2} - \{E_{\boldsymbol{\omega}^0}\{\partial_{\boldsymbol{\omega}}\ln[p(y,\boldsymbol{\theta}|\boldsymbol{\omega}^0)]\}\}^{\otimes 2}\big].\end{aligned}$$

where $\mathbf{a}^{\otimes 2} = \mathbf{a}\mathbf{a}^{\top}, \mathbf{a} = [a_1, a_2, \cdots, a_n]^{\top}$. Similar to Sect. 5.2.1, H_{ϕ} can be approximated by MCMC sampling:

$$\begin{aligned}\mathbf{H}_{\phi} &= \phi''(1)\big[E_{\boldsymbol{\omega}^0}\{\partial_{\boldsymbol{\omega}}\ln[p(y,\boldsymbol{\theta}|\boldsymbol{\omega}^0))]\}^{\otimes 2} - \{E_{\omega^0}\{\partial_{\omega}\ln[p(y,\boldsymbol{\theta}|\boldsymbol{\omega}^0))]\}\}^{\otimes 2}\big]\\ &\approx \phi''(1)\Big[\frac{1}{T}\sum_{t=1}^{T}\{\partial_{\boldsymbol{\omega}}\ln[p(y,\boldsymbol{\theta}^{(t)}|\boldsymbol{\omega}^{\mathbf{0}})]\}^{\otimes 2} - \{\frac{1}{T}\sum_{t=1}^{T}\{\partial_{\boldsymbol{\omega}}\ln[p(y,\boldsymbol{\theta}^{(t)}|\boldsymbol{\omega}^{\mathbf{0}})]\}^{\otimes 2}\Big].\end{aligned}$$

C. The commonly used objective function used for second-order influence measure: posterior mean

Modified from Hao et al. (2019), the posterior mean of the function $\mathbf{h}(\boldsymbol{\theta})$ is defined as follows:

$$\mathbf{M}_{\mathbf{h}}(\boldsymbol{\omega}) = \int h(\boldsymbol{\theta})p(\boldsymbol{\theta}|y,\boldsymbol{\omega})d\boldsymbol{\theta}.$$

In this paper, we adopt the Cook's posterior mean distance to characterize the effect of $\boldsymbol{\omega}$ on the posterior mean distance of $\mathbf{h}(\boldsymbol{\theta})$:

$$\mathbf{CM}_{\mathbf{h}}(\boldsymbol{\omega}) = \big(\mathbf{M}_{\mathbf{h}}(\boldsymbol{\omega}) - \mathbf{M}_{\mathbf{h}}(\boldsymbol{\omega}_0)\big)^{\top}\mathbf{G}_{\mathbf{h}}\big(\mathbf{M}_{\mathbf{h}}(\boldsymbol{\omega}) - \mathbf{M}_{\mathbf{h}}(\boldsymbol{\omega}_0)\big),$$

where $\mathbf{G}_{\mathbf{h}} = \big[\mathrm{Var}(\mathbf{h}(\boldsymbol{\theta}|\mathbf{y}))\big]^{-1}$ is the inverse of the posterior covariance matrix of $h(\boldsymbol{\theta})$ with respect to $p(\boldsymbol{\theta}|y,\boldsymbol{\omega}_0)$. When $f(\boldsymbol{\omega}) = \mathbf{CM}_{\mathbf{h}}(\boldsymbol{\omega})$, we can get

$$\begin{aligned}\nabla_h &= \mathbf{0},\\ \mathbf{H}_{\mathbf{h}} &= \mathbf{M}_{\mathbf{h}}^{*\top}\mathbf{G}_{\mathbf{h}}\mathbf{M}_{\mathbf{h}}^{*},\\ \mathbf{M}_{\mathbf{h}}{}^{*} &= \mathrm{Cov}_{\boldsymbol{\omega}^{\mathbf{0}}}\{\mathbf{h}(\boldsymbol{\theta}), \partial_{\boldsymbol{\omega}}\ln[p(y,\boldsymbol{\theta}^{(n)}|\boldsymbol{\omega}^0)]\}\\ &\approx \frac{1}{T}\sum_{t=1}^{T}\{\mathbf{h}(\boldsymbol{\theta}^{(t)})\partial_{\boldsymbol{\omega}}\ln[p(y,\boldsymbol{\theta}^{(t)}|\boldsymbol{\omega}^0)]\} - \Big[\frac{1}{T}\sum_{t=1}^{T}\mathbf{h}(\boldsymbol{\theta}^{(t)})\Big]\Big[\frac{1}{T}\sum_{t=1}^{T}\partial_{\boldsymbol{\omega}}\ln[p(y,\boldsymbol{\theta}^{(t)}|\boldsymbol{\omega}^0)]\Big].\end{aligned}$$

The operational steps of the proposed Bayesian local sensitivity analysis method for solving the objective function above can be summarized as follows:

Step 1: Construct a Bayesian perturbation manifold.

Step 2: Calculate $\nabla_f = \partial_{\boldsymbol{\omega}} f(\boldsymbol{\omega}(0)), \mathbf{H}_{\mathbf{f}} = \partial^2_{\boldsymbol{\omega}} f(\boldsymbol{\omega}(0))$, and $\mathbf{G} = \mathbf{G}(\boldsymbol{\omega}^0)$.

Step 3: If ∇_f is not equal to zero, then calculate the first-order influence measure. Otherwise, calculate the second-order influence measure.

Step 4: Determine whether the k-th point is an influential point for a given objective function $f(\boldsymbol{\omega})$.

Note: The selection of the objective function

Indeed, there is no one-size-fits-all criterion for selecting different objective functions or local sensitivity measures in Bayesian local influence analysis. The choice of objective function depends on the specific context and the target of the assessment.

For example, when evaluating the impact of perturbations on the posterior distribution, ϕ-divergences are often preferred. On the other hand, if the focus is on assessing the impact of perturbations on model parameters, the posterior mean distance is commonly considered first.

Using different objective functions may lead to varying results in the final analysis. This variation is natural since different objective functions capture different aspects of the model or the perturbation's influence. Researchers should carefully select the most appropriate objective function based on the specific research question and the desired insights. The flexibility in choosing objective functions provides a valuable tool in Bayesian local influence analysis, enabling a comprehensive understanding of the model's sensitivity and robustness to perturbations, see Tang and Duan (2014).

6 Simulation study

To investigate the effectiveness of the proposed method for diagnostics of PLM-SNAR(1), we conduct a simulation study here.

We generate a sample of $n = 200$ based on the following model:

$$\begin{aligned} y_i &= 2x_{1i} + 4x_{2i} + \cos(4\pi t_i)\exp\left(-\frac{1}{2}{t_i}^2\right) + \epsilon_i, \\ \epsilon_i &= (2 * 0.9 - 1)\epsilon_{i-1} + e_i, \\ e_i &\overset{iid}{\sim} \mathrm{SN}(-\sqrt{2/\pi} * 0.22, 0.97^2, 0.22), \quad i = 1, 2, 3 \ldots 200, \end{aligned} \tag{6}$$

where $x_{1i} \sim U(0, 1)$, $x_{2i} \sim U(1, 2)$ and $t_i \sim U(0.6, 1.6)$ are used, with U indicating a uniform distribution.

To perform spline interpolation, we calculate the sample quantiles of t. The simulation includes two parts, namely parameter estimation and statistical diagnostics. Our focus is on statistical diagnostics in this section.

6.1 Parameter estimation

For these parameters, the following prior distributions are set, with the intercept β_0 not taken into consideration as in Ferreira et al. (2022b):

Table 2 Parameter estimates with the corresponding values used in the simulation study

Parameter	True value	Mean	SD
β_1	2.00	1.97	0.19
β_2	4.00	4.01	0.20
σ	0.97	0.89	0.10
λ	0.22	0.24	0.68
ρ_0	0.90	0.91	0.02

$$\begin{aligned} &\beta_1 \sim \mathrm{N}(0, 100), \\ &\beta_2 \sim \mathrm{N}(0, 100), \\ &\sigma^2 \sim \mathrm{IGamma}(0.1, 10), \\ &\lambda \sim \mathrm{N}(0.2, 400), \\ &\rho_0 \sim \mathrm{Beta}(20, 1.5). \end{aligned} \tag{7}$$

For each sample in our analysis, we utilize the MCMC algorithm and the posterior distribution to calculate the parameter estimates. The resulting estimates for the parameters are as described in Table 2 and Fig. 2 (the sampling process is presented in Appendix):

These estimates provide valuable insights into the model's behavior and help us understand the relationships between the variables under consideration.

In summary, our sampling process has yielded favorable results. The close agreement between the sample mean and the true parameter value demonstrates the effectiveness of our sampling approach. Additionally, the small standard error indicates the stability of our sampling process, particularly for moderate-sized samples. This confirms the effectiveness of Bayesian estimation in our analysis.

6.2 Local influence analysis

Incorporating the MCMC algorithm, we conduct diagnostic work for the PLM-SNAR(1) model using Bayesian local influence analysis. Following a similar procedure as described in Sect. 5, we subject the model to prior perturbation, variance perturbation, and data perturbation. Moreover, we use three objective functions (Bayes factor, ϕ-divergence and posterior mean) to evaluate the influence measures. This comprehensive analysis allows us to assess the impact of small perturbations in the model and gain valuable insights into its behavior and sensitivity.

6.2.1 Perturbation of variances (and priors)

Firstly, we perform the perturbation of variances. We choose to perturb at $num = 50$ and $num = 150$ and the perturbation involves changing the corresponding y_{num} to $y_{num} + 5\sigma_y$, where σ_y is the standard error in generating y. The diagnostic results are displayed in Fig. 3.

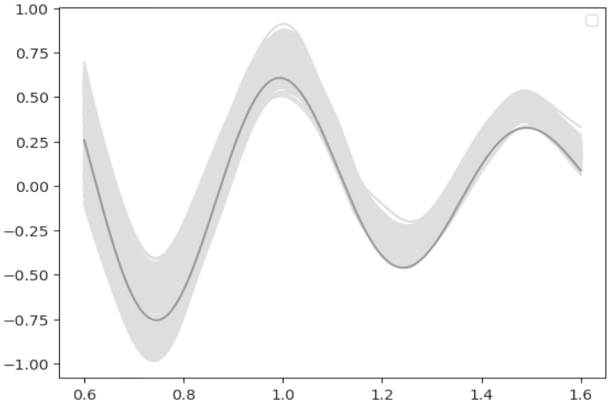
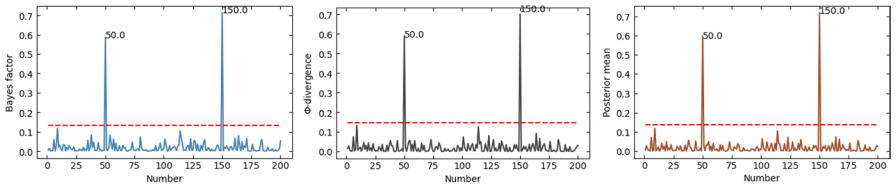


Fig. 2 Graph of non-parametric thin-plate spline fit showing adjusted curves (light gray lines) and true curves (dark gray lines). Y-axis represents the non-parametric component, and X-axis represents 't'



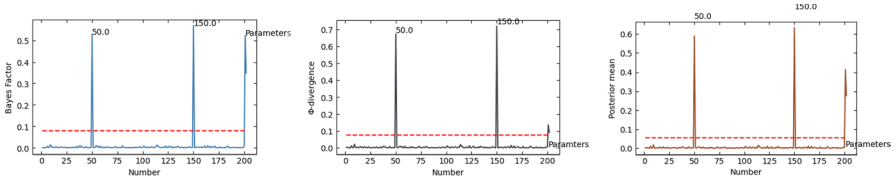
(a) Bayes factor (b) ϕ-divergence (c) Posterior mean

Fig. 3 Schematic diagrams of the Bayesian local influence analysis for the three measures under the perturbation of variances

From Fig. 3, it is evident that we accurately diagnose the predetermined perturbation points for all three measures, particularly at $num = 50$ and $num = 150$, where the diagnostic results stand out significantly compared to other points.

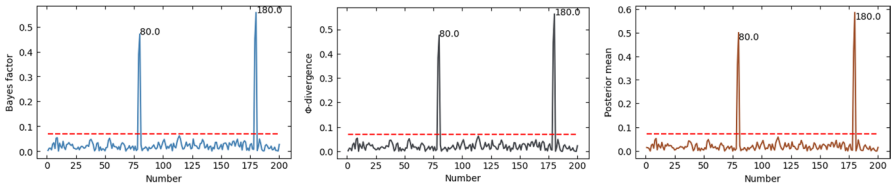
Next, we delve into the simultaneous perturbation of variances and priors. We perturb the variances and, concurrently, adjust the linear parameter β_1 and the shape parameter σ^2 (as the parameters of the first-order autoregressive assume a similar role to the linear parameters, we solely perturb the prior distribution of the linear parameters). The perturbation process entails two steps. First, we change the corresponding y_{num} to $y_{num} + 5\sigma_y$ (where σ_y is the standard error in generating y). Second, we modify the prior distribution of β_1 to $\mathrm{N}(0, N(0, 3 \times 10^7)$, the prior distribution of σ^2 to IGamma $(0.1, IGamma\,(0.1, 10^{-6})$. The diagnostic results are presented in Fig. 4.

Upon examination, it is observed that the measurement method of the Bayesian factor is not currently effective in diagnosing the prior perturbation part. However, with an increase in the perturbation of the prior part, the Bayesian factor continues to be effective in diagnosing the prior perturbation part. Figure 4 illustrates our accurate diagnosis of predetermined perturbation points for all three measures. Particularly noteworthy are the two locations at $num = 50$ and $num = 150$, as well as



(a) Bayes factor (b) ϕ-divergence (c) Posterior mean

Fig. 4 Schematic diagrams of the Bayesian local influence analysis for the three measures under simultaneous perturbation of variances and priors, with larger perturbation on parameters in the picture (a)



(a) Bayes factor (b) ϕ-divergence (c) Posterior mean

Fig. 5 Schematic diagrams of the Bayesian local influence analysis for the three measures under the perturbation of data on the response variable

for the parameter β_1, where diagnostic results stand out significantly compared to other points. Additionally, it is evident that the ϕ-divergence and Bayes factor measure may not be the optimal choices for diagnosing prior perturbation; the Posterior mean appears to be the most effective choice.

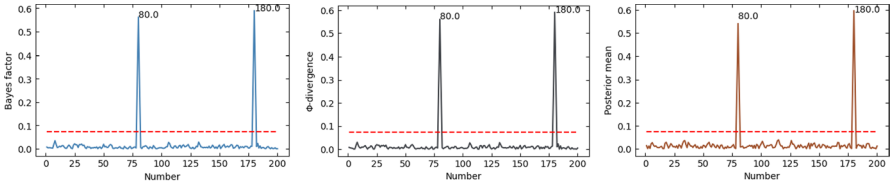
6.2.2 Perturbation of data

Then, we perform a data perturbation on the response variable. We choose to perturb at $num = 80$ and $num = 180$ and the perturbation method involves changing the corresponding y_{num} to $y_{num} + 10y_{num}$. The diagnostic results are shown in Fig. 5.

From Fig. 5, we can observe that the perturbation of the data (response variable y) can be accurately diagnosed. Specifically, the two points where $num = 80$ and $num = 180$ were successfully diagnosed by all three methods. For other points, the diagnostic values are generally small, indicating the effectiveness of our diagnostic work.

Finally, we perform data perturbation on the explanatory variable (previous perturbation on y has been canceled). The perturbation involves changing the corresponding x_{num} to $x_{num} + 8x_{num}$ at $num = 80$ and $num = 180$. The diagnostic results are shown in Fig. 6.

In Fig. 6, we can clearly see that the diagnostic values obtained at the two locations where the explanatory variable was perturbed ($num = 80$ and $num = 180$) are significantly higher than the values at other locations. This indicates that we have successfully diagnosed these two points. It appears that the diagnostic of perturbations in the explanatory variable is equally effective as the previous perturbations we analyzed.



(a) Bayes factor (b) ϕ-divergence (c) Posterior mean

Fig. 6 Schematic diagrams of the Bayesian local influence analysis for the three measures under the perturbation of data on the explanatory variable

7 Application

To validate the effectiveness of statistical diagnosis using Bayesian local influence analysis on the PLM-SNAR(1) model, we conducted practical applications on the 1976 US California high-altitude ozone pollution data. The dataset consists of $n = 366$ observations, with each representing a response variable (daily ozone concentration) and an explanatory variable (temperature of the day).

To process the data, we follow the following steps:

Step 1: Extract the desired variable column data.

Step 2: Record the corresponding order of the data and denoted it as "t".

Step 3: Delete the sets of data with missing values in the response variable.

Step 4: Delete the sets of data with missing values in the explanatory variable.

Next, we conduct a partial descriptive analysis on the dataset shown in Fig. 7.

Figure 7 provides valuable insights into the dataset. Figure 7a, b reveal significant skewness in the data, suggesting that a SN distribution could be a suitable fit. Figure 7c illustrates a clear linear relationship between ozone concentration and temperature, while Fig. 7d indicates a notable nonlinear relationship between ozone concentration and the corresponding day. These findings suggest that partial linear models can effectively capture the dataset's characteristics.

Furthermore, Ferreira et al. (2022b) used the maximum likelihood estimation method and demonstrated that the partial residuals in the data exhibit characteristics of an autoregressive process of order 1, i.e., AR(1). This evidence further supports the use of the PLM-SNAR(1) model for analyzing the dataset and validates the effectiveness of our statistical diagnostics using Bayesian local influence analysis.

Therefore, we construct the following special case of model (3) for the data:

$$\begin{aligned} &y_i = \beta_0 + \beta_1 * x_i + g(t_i) + \epsilon_i, \\ &\epsilon_i = \rho\epsilon_{i-1} + e_i, \qquad -1 < \rho < 1, \\ &e_i \overset{iid}{\sim} SN(-\sqrt{2/\pi}\lambda, \sigma^2, \lambda), \quad i = 1, 2, 3 \ldots 330. \end{aligned} \tag{8}$$

Before proceeding with the subsequent statistical diagnostic work, we first need to estimate the parameters of the model. The prior distributions will remain the same as Eq. (7), in addition to the intercept with $\beta_0 \sim N(-5, 100)$ (Table 3).

Indeed, the comparison between Bayesian and frequentist parameter estimation methods shows that they perform similarly in this context. However, the main

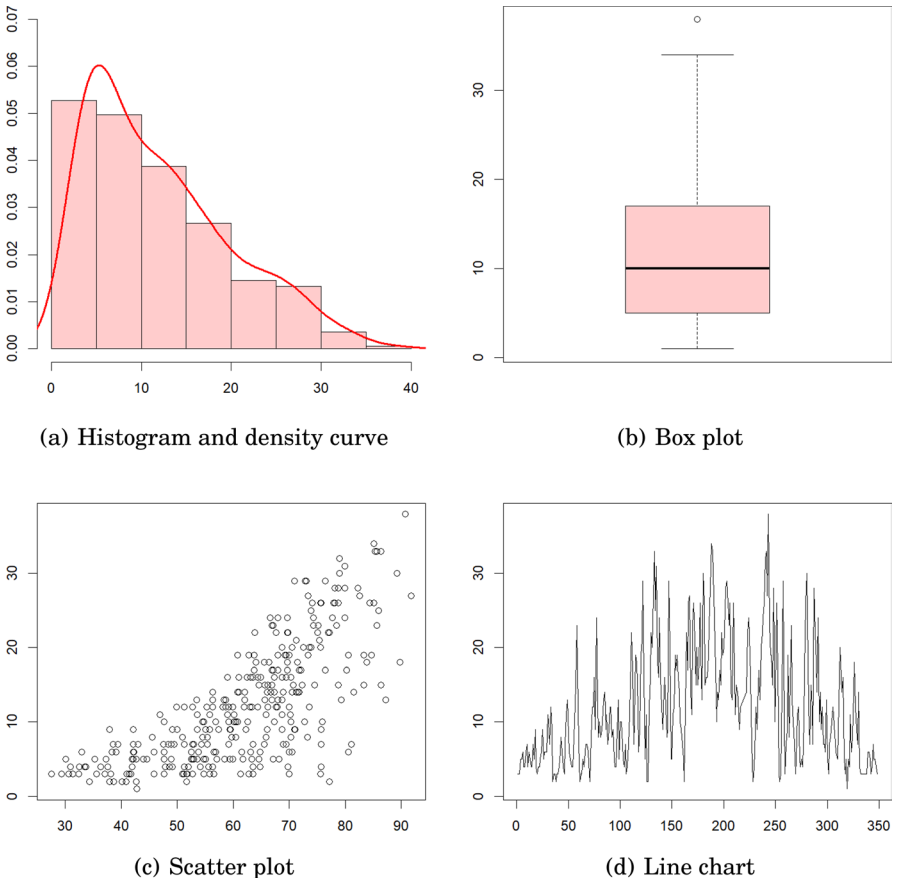


Fig. 7 Partial descriptive analysis of Ozone data: **a** Histogram of Ozone, **b** Boxp lot of ozone, **c** scatter plot of ozone versus temperature, **d** line chart of ozone versus day

emphasis of this paper lies in the statistical diagnostics, with parameter estimation being only an integral component. As a result, we will not delve further into the discussion of parameter estimation in this section.

Next, we proceed with the local influence analysis under two scenarios: model perturbation and data perturbation. We utilize three measures, namely Bayes

Table 3 MCMC and maximum penalized likelihood (MPL) estimates (Ferreira et al. 2022b)

Parameter	MCMC	MPL
β_0	−5.73	−7.32
β_1	0.29	0.31
σ	2.66	1.91
λ	6.13	6.80
ρ_0	0.46	0.33

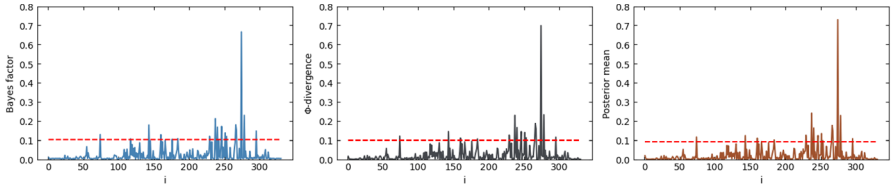


Fig. 8 Perturbation of variances

factor, ϕ-divergence and posterior mean, to conduct the analysis. The results of the analysis are presented below (Figs. 8, 9 and 10).

Note that in Table 4 B indicates Bayes-factor, ϕ indicates ϕ-divergence, P indicates posterior mean, Data y indicates perturbation of data (Ozone) and Data x indicates perturbation of data (Temperature). As multiple methods detected an issue with observation #278, it is necessary to review the dataset and verify the accuracy of this result (Table 5).

Our assessment holds significance as the results of the Bayesian local influence analysis align with the frequentist local influence analysis conducted by Ferreira et al. (2022b), albeit with a slight difference. While their diagnostic results were commendable, ours are comparable, and notably, the Bayesian approach identifies certain observations (such as #74, #119, #184, #278 and #295) that were not previously detected. This newly discovered information adds a valuable dimension to our diagnostic findings (Table 6).

Using observation #74 as an example, we conduct a detailed analysis of the causes to showcase the effectiveness of our diagnostic method. We see that #74 corresponds to March 17, 1976. To understand the reason for such an influential observation, we need to examine the weather information for a few days before that day. This analysis will provide insights into any unusual weather conditions

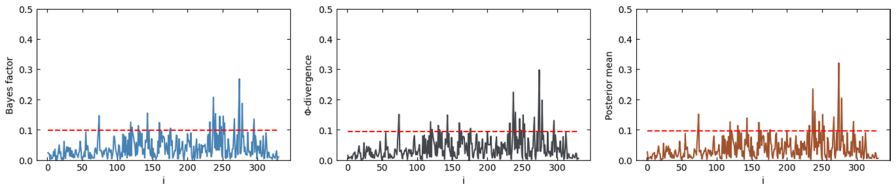


Fig. 9 Perturbation of data (Ozone)

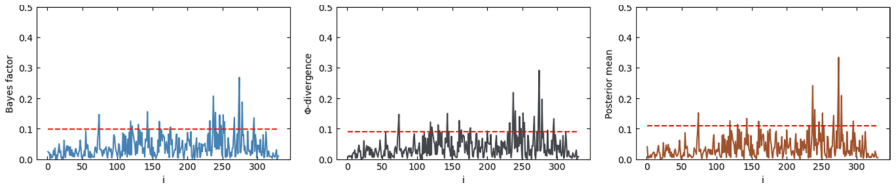


Fig. 10 Perturbation of data (Temperature)

Table 4 Summary of the diagnostic results

i	Variance-B	Variance-ϕ	Variance-P	Data y-B	Data y-ϕ	Data y-P	Data x-B	Data x-ϕ	Data x-P
74	*	*	*	*	*	*	*	*	*
117	*								
119				*	*	*	*	*	*
130				*	*	*	*	*	*
143	*	*	*	*	*	*	*	*	*
160	*	*	*	*	*	*	*	*	*
163					*	*		*	
166			*						
176	*			*	*	*	*	*	
184	*	*	*						
200								*	
229	*	*	*						
232				*	*	*	*	*	*
237	*	*	*	*	*	*	*	*	*
240	*	*	*	*	*	*	*	*	*
246	*	*	*	*	*	*	*	*	*
251	*	*	*	*	*	*	*	*	*
253	*	*	*	*	*	*	*	*	
266	*	*	*	*	*	*	*	*	
274	*	*	*	*	*	*	*	*	*
278	*	*	*	*	*	*	*	*	*
286								*	
295	*	*	*	*	*	*	*	*	*

Table 5 Observation #278 and its neighboring observations

i	Ozone concentration
277	14
278	24
279	10
280	14

Table 6 Observation #74 and its neighboring observations

i	Ozone concentration
71	12
72	16
73	9
74	24

or other factors that may have influenced the ozone concentration on March 17, 1976. By examining the weather data leading up to that date, we can gain a better understanding of the underlying cause of the influential observation and its impact on the ozone concentration.

Under normal circumstances, the temperature of the troposphere decreases with increasing altitude, leading to air convection where cold air sinks and hot air rises, carrying pollutants to higher altitudes for release. However, in certain conditions, such as the inversion phenomenon, the temperature of the troposphere may increase with altitude. In Los Angeles, the inversion phenomenon is more severe due to its location on the west coast with nearby cold ocean currents and surrounded by deserts.

The inversion phenomenon traps pollutants at ground level, leading to atmospheric pollution and severe ozone concentration. Table 7 confirms that Los Angeles experienced the inversion phenomenon on several days, causing ozone to accumulate at the surface.

Other factors, such as wind speed, humidity, and air visibility, also contribute to ozone pollution. Low wind speed and high humidity can lead to ozone accumulation, while good air visibility reduces the reaction of ozone with suspended particles. The data in Table 8 indicates that on the day of abnormal ozone concentration, wind speed was low, humidity was at 60% (favorable for ozone storage), and air visibility was good with fewer suspended particles.

Tables 6 and 7 reveal that observation #74 (March 17, 1976) is indeed influential. Table 8 further indicates that the ozone accumulated on the preceding day was not effectively cleared, and the favorable conditions on March 17, 1976, exacerbated

Table 7 Inversion phenomenon that occurred from March 12 to March 17, 1976

Date	Surface temp	Temp at the bottom of the inversion layer	Height of the bottom of the inversion layer
12/03/1976	46.04	59.72	613
13/03/1976		64.40	334
14/03/1976		61.88	567
15/03/1976	57.92	64.94	488
16/03/1976	57.02	71.06	531
17/03/1976	58.64	66.56	508

Table 8 Some weather indicators from March 12, 1976 (the first day when the inversion occurred during this period) to March 17, 1976 (the units of measurement for each column are the same)

Date	Wind speed	Humidity (%)	Air visibility
12/03/1976	0	60	300
13/03/1976	4	31	300
14/03/1976	3	66	150
15/03/1976	5	53	2
16/03/1976	2	42	50
17/03/1976	3	60	70

ozone pollution. Consequently, there is no misdiagnosis in the Bayesian local influence analysis, affirming the validity of the conclusion regarding the influential observation.

8 Concluding remarks

This paper presented the use of Bayesian local influence analysis for statistical diagnosis in the PLM-SNAR(1) model, addressing the limitations of maximum likelihood estimation in frequentist statistics for moderate-sized data. By employing the Gibbs sampling and Metropolis-Hastings sampling, we successfully obtained parameter estimation results. The method's feasibility was demonstrated through simulation experiments and real-world applications, employing different objective functions and applying variance and data perturbation in the analysis. In the empirical analysis, we applied the PLM-SNAR(1) model to the 1976 Los Angeles ozone concentration dataset (moderate-sized data) and achieved superior results compared to frequentist statistics, confirming the effectiveness of our improvements. Our future research may include extending the method to study PLM-SNAR(p) models and PLMs with errors following a skew-t autoregressive structure of order p, i.e. PLM-STAR(p) models.

Appendix

A1: Deviation of model reduction

Proof By Lemma 2, model (3) can be further expressed as:

$$\begin{aligned} y_i &= \mathbf{x}_i{}^\top \boldsymbol{\beta} + g(t_i) + \rho(y_{i-1} - \mathbf{x}_{i-1}{}^\top \boldsymbol{\beta} - g(t_{i-1})) + e_i, \\ e_i &= -\sqrt{2/\pi}\lambda + \lambda\left|h_{0i}\right| + \sigma h_{1i}, \quad i = 1, 2, 3 \ldots n. \end{aligned} \tag{9}$$

In model (9), h_{0i} and h_{1i} are mutually independent standard normal random variables. Let $h_i = \left|h_{0i}\right|$ and $\eta_i = \sigma h_{1i}$. Then, we can further simplify the model to:

$$\begin{aligned} &y_i = \mathbf{x}_i{}^\top \boldsymbol{\beta} + g(t_i) + \rho(y_{i-1} - \mathbf{x}_{i-1}{}^\top \boldsymbol{\beta} - g(t_{i-1})) - \sqrt{2/\pi}\lambda + \lambda h_i + \eta_i, \\ &\eta_i \overset{iid}{\sim} N(0, \sigma^2), \quad i = 1, 2, 3 \ldots n. \end{aligned} \tag{10}$$

Ruppert et al. (2003) addressed the semiparametric modeling problem by demonstrating the equivalence of penalized splines and additive mixed models. In this paper, similar to Ruppert, we aim to use the low-rank thin-plate spline to represent the non-parametric part of the model in a different manner. We will expand the smooth function g(t) as follows:

$$g(t) = \alpha_0 + \alpha_1 t + \sum_{s=1}^{K} u_s |t - \kappa_s|^3, \tag{11}$$

where $t_{min} \le \kappa_1 \le \kappa_2 \ldots \le \kappa_K \le t_{max}$, and they are fixed knots. In spline expansion, the selection of knots is crucial. Typically, these knots are determined as the sample quantiles of the covariate t's corresponding to probability $k/(K+1)$, where K is the number of knots. Having too many or too few knots can lead to poor spline estimation. In this paper, we adopt the knots selection method proposed by Ruppert et al. (2003) and choose 20 knots ($K = 20$) for the spline expansion of the non-parametric part. After a series of matrix transformations (Crainiceanu et al. 2005), we can express Eq. (7) in the following form:

$$\begin{aligned} \boldsymbol{g(t)} &= \boldsymbol{T\alpha} + \boldsymbol{Z}_K \boldsymbol{u} \\ &= \boldsymbol{T\alpha} + (\boldsymbol{Z}_K \boldsymbol{\Omega}_K^{-\frac{1}{2}})(\boldsymbol{\Omega}_K^{\frac{1}{2}} \boldsymbol{u}) \\ &= \boldsymbol{T\alpha} + \boldsymbol{Zb}, \end{aligned}$$

where $\boldsymbol{g(t)} = (g(t_1), g(t_2), \ldots, g(t_n))^\top$, $\mathbf{t} = (t_1, t_2, \ldots, t_n)^\top$, $\boldsymbol{\alpha} = (\alpha_0, \alpha_1)^\top$, $\boldsymbol{T} = (\mathbf{1}_{n\times 1}, \boldsymbol{t})$, $\boldsymbol{u} = (u_1, u_2, \ldots, u_K)^\top$, $\boldsymbol{Z}_K$ is a matrix with i-th row $\boldsymbol{Z}_{Ki} = \{|t_i - \kappa_1|^3, \ldots, |t_i - \kappa_K|^3\}$, $\boldsymbol{\Omega}_K$ is a penalty coefficient matrix with k-th row $\boldsymbol{\Omega}_{Kk} = \{|\kappa_k - \kappa_1|^3, \ldots, |\kappa_k - \kappa_K|^3\}$ (to avoid overfitting), $\boldsymbol{Z} = \boldsymbol{Z_K}\boldsymbol{\Omega}_K^{-\frac{1}{2}}$, $\boldsymbol{b} = \boldsymbol{\Omega}_K^{\frac{1}{2}}\boldsymbol{u}$ is assumed to be normally distributed with mean zero and variance σ_b^2.

Then we get

$$\begin{aligned} &Y_i = \mathbf{w}_i{}^\top \boldsymbol{\Lambda} + z_i^\top \boldsymbol{b} + \rho(y_{i-1} - \mathbf{w}_{i-1}{}^\top \boldsymbol{\Lambda} - z_{i-1}\boldsymbol{b}) - \sqrt{2/\pi}\lambda + \lambda h_i + \eta_i, \\ &\eta_i \overset{iid}{\sim} N(0, \sigma^2), \quad i = 1, 2, 3 \ldots n, \end{aligned} \tag{12}$$

which gives model (4). □

A2: Perturbation of priors

We established perturbations of the parameters.

$$\begin{aligned} p(y, \boldsymbol{\theta}|\boldsymbol{\omega}_{p+5+20}) \propto & \prod_{i=1}^{p} \varphi\Big(\frac{\beta_i - \beta_{0i} - \omega_{\beta_i}}{\sigma_{\beta_i}}\Big) \times \varphi\Big(\frac{\alpha_0 - \omega_{\alpha_0}}{\sigma_{\alpha_0}}\Big) \times \varphi\Big(\frac{\alpha_1 - \omega_{\alpha_1}}{\sigma_{\alpha_1}}\Big) \times \varphi\Big(\frac{\lambda - \mu_\lambda - \omega_\lambda}{\sigma_\lambda}\Big) \\ & \times \frac{1}{B(a_{\rho 0} + \omega_{\rho 0}, b_{\rho 0})} \rho_0^{a_{\rho 0} + \omega_{\rho 0} - 1} (1 - \rho_0)^{b_{\rho 0} - 1} \\ & \times \frac{(\frac{1}{2}q_1 + \omega_{\sigma^2})^{\frac{1}{2}q_0}}{\Gamma(\frac{1}{2}q_0)} \sigma^{-q_0 - 2} \exp\Big\{-\frac{q_1 + \omega_{\sigma^2}}{2\sigma^2}\Big\} \times \prod_{j=1}^{20} \varphi\Big(\frac{b_j - \omega_{b_j}}{\sigma_{bj}}\Big), \end{aligned}$$

where $\boldsymbol{\omega}_{p+5+20} = (\omega_{\boldsymbol{\beta}}, \omega_{\boldsymbol{\alpha}}, \omega_\lambda, \omega_\rho, \omega_{\sigma^2}, \omega_{\mathbf{b}})$, and $\boldsymbol{\omega}^0_{p+5+20} = \mathbf{0}$ represents the scenario without any perturbation, $B(\cdot)$ means beta function, $\Gamma(\cdot)$ means gamma function. The

perturbation model $\mathcal{M} = \{p(y, \boldsymbol{\theta}|\boldsymbol{\omega}) : \boldsymbol{\omega} \in R^{p+5+20}\}$ yields a Riemannian manifold and the tangent space $T_{\boldsymbol{\omega}}$ spanned by $\mathcal{M}$ is given by

$$\begin{aligned}\dot{l}(y, \boldsymbol{\theta}|\boldsymbol{\omega}) = \Big(& \frac{\beta_1 - \beta_{01} - \omega_{\beta_1}}{\sigma^2_{\beta_1}}, \ldots, \frac{\beta_p - \beta_{0p} - \omega_{\beta_p}}{\sigma^2_{\beta_p}}, \frac{\alpha_0 - \omega_{\alpha_0}}{\sigma^2_{\alpha_0}}, \frac{\alpha_1 - \omega_{\alpha_1}}{\sigma^2_{\alpha_1}}, \frac{\lambda - \mu_\lambda - \omega_\lambda}{\sigma^2_\lambda}, \\ & - \frac{\int_0^1 \ln\rho_0 (1-\rho_0)^{b_{\rho 0}-1} \rho_0^{a_{\rho 0}+\omega_{\rho 0}-1} d\rho_0}{B(a_{\rho 0}+\omega_{\rho 0}, b_{\rho 0})} + \ln\rho_0, \frac{q_0}{q_1 + 2\omega_{\sigma^2}} - \frac{1}{2\sigma^2}, \frac{b_1 - \omega_{b1}}{\sigma^2_{\beta_1}}, \\ & \ldots, \frac{b_{20} - \omega_{b_{20}}}{\sigma^2_{b_{20}}} \Big).\end{aligned}$$

Then, we obtain

$$\mathbf{G}(\boldsymbol{\omega}^0) = \mathrm{diag}\Big(\frac{1}{\sigma^2_{\beta_1}}, \ldots, \frac{1}{\sigma^2_{\beta_p}}, \frac{1}{\sigma^2_{\alpha_0}}, \frac{1}{\sigma^2_{\alpha_1}}, \frac{1}{\sigma^2_\lambda}, D(\ln\rho_0), \frac{2q_0}{q_1}, \frac{1}{\sigma^2_b}, \ldots, \frac{1}{\sigma^2_{b_{20}}} \Big),$$

where

$$\begin{aligned}D(\ln\rho_0) = & \int_0^1 \ln^2\rho_0 \frac{1}{B(a_{\rho 0}, b_{\rho 0})} (1-\rho_0)^{b_{\rho 0}-1} \rho_0^{a_{\rho 0}-1} d\rho_0 \\ & - \Big(\int_0^1 \ln\rho_0 \frac{1}{B(a_{\rho 0}, b_{\rho 0})} (1-\rho_0)^{b_{\rho 0}-1} \rho_0^{a_{\rho 0}-1} d\rho_0 \Big)^2.\end{aligned}$$

A3: Perturbation of variances

By incorporating the perturbations, we obtain a refined posterior distribution that helps in better understanding the model's behavior under different variance scenarios. We have that:

$$\begin{aligned}p(y, \boldsymbol{\theta}|\boldsymbol{\omega}) &= p(\boldsymbol{\theta}) \cdot L(\boldsymbol{\theta}|\boldsymbol{\omega}) \\ &\propto L(\boldsymbol{\theta}|\boldsymbol{\omega}) \\ &= \exp[-\frac{1}{2}\ln\frac{\sigma^2}{\omega_1} - \frac{\omega_1}{2\sigma^2}(y_1 - m_1 + \sqrt{2/\pi}\lambda - \lambda h_1)^2 - \sum_{i=2}^n \frac{1}{2}\ln\frac{\sigma^2}{\omega_i} \\ &\quad - \sum_{i=2}^n \frac{\omega_i}{2\sigma^2}(y_i - \mu_i + \sqrt{2/\pi}\lambda - \lambda h_i)^2], \\ \dot{l}(y, \boldsymbol{\theta}|\boldsymbol{\omega}) &= \Big(\frac{1}{2\omega_1} - \frac{(y_1 - m_1 + \sqrt{2/\pi}\lambda - \lambda h_1)^2}{2\sigma^2}, \frac{1}{2\omega_2} - \frac{(y_2 - \mu_2 + \sqrt{2/\pi}\lambda - \lambda h_2)^2}{2\sigma^2}, \ldots, \\ &\quad \frac{1}{2\omega_n} - \frac{(y_n - \mu_n + \sqrt{2/\pi}\lambda - \lambda h_n)^2}{2\sigma^2} \Big),\end{aligned}$$

Then, we obtain:

$$\mathbf{G}(\boldsymbol{\omega}^0) = \frac{1}{2}\mathbf{I}_{n\times n}.$$

A4: Perturbation of data

A4A: Response variable

Establishing perturbations on the response variable $\mathbf{y}$

$$\begin{bmatrix} y_1+\omega_1 \\ y_2+\omega_2 \\ y_3+\omega_3 \\ \vdots \\ y_n+\omega_n \end{bmatrix} = \begin{bmatrix} m_1 \\ m_2 \\ m_3 \\ \vdots \\ m_n \end{bmatrix} + \rho \begin{bmatrix} 0 \\ y_1+\omega_1-m_1 \\ y_2+\omega_2-m_2 \\ \vdots \\ y_{n-1}+\omega_{n-1}-m_{n-1} \end{bmatrix} - \lambda \begin{bmatrix} \sqrt{2/\pi}-h_1 \\ \sqrt{2/\pi}-h_2 \\ \sqrt{2/\pi}-h_3 \\ \vdots \\ \sqrt{2/\pi}-h_n \end{bmatrix} + \begin{bmatrix} \eta_1 \\ \eta_2 \\ \eta_3 \\ \vdots \\ \eta_n \end{bmatrix}. \tag{13}$$

We have that

$$\begin{aligned} p(y,\boldsymbol{\theta}|\boldsymbol{\omega}) &= p(\boldsymbol{\theta})\cdot L(\boldsymbol{\theta}|\boldsymbol{\omega}) \\ &\propto L(\boldsymbol{\theta}|\boldsymbol{\omega}) \\ &\propto \exp[-\frac{1}{2\sigma^2}(R_1^2+\sum_{i=2}^{n} R_i^2)], \\ \dot{l}(y,\boldsymbol{\theta}|\boldsymbol{\omega}) &= \Big(\frac{\rho R_2 - R_1}{\sigma^2}, \ldots, \frac{\rho R_{i+1}-R_i}{\sigma^2}, \ldots, \frac{\rho R_n - R_{n-1}}{\sigma^2}, -\frac{R_n}{\sigma^2}\Big). \end{aligned}$$

Then, we obtain:

$$\mathbf{G}(\boldsymbol{\omega}^0) = \mathrm{diag}\Big(\mathbf{E}_{\boldsymbol{\omega}^0}[\frac{\rho^2+1}{\sigma^2}], \ldots, \mathbf{E}_{\boldsymbol{\omega}^0}[\frac{\rho^2+1}{\sigma^2}], \mathbf{E}_{\boldsymbol{\omega}^0}[\frac{1}{\sigma^2}]\Big),$$

where

$$\begin{aligned} R_1 &= (y_1+\omega_1) - m_1 + \sqrt{2/\pi}\lambda - \lambda h_1, \\ R_i &= (y_i+\omega_i) - m_i - \rho[(y_{i-1}+\omega_{i-1}) - m_i - 1)] + \sqrt{2/\pi}\lambda - \lambda h_i, \qquad i = 2,3,\ldots,n. \end{aligned}$$

A4B: Explanatory variable

Establishing perturbations on the explanatory variable $\mathbf{x_k}$ (assumed continuous)

$$\mathbf{X_k}(\omega) = \begin{bmatrix} x_{11} & x_{21} & \cdots & x_{n1} \\ x_{12} & x_{22} & \cdots & x_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ x_{1k}+\omega_1 & x_{2k}+\omega_2 & \cdots & x_{nk}+\omega_n \\ \vdots & \vdots & \ddots & \vdots \\ x_{1p} & x_{2p} & \cdots & x_{np} \end{bmatrix}. \tag{14}$$

As a result, we obtain

$$\begin{aligned} p(y,\boldsymbol{\theta}|\boldsymbol{\omega}) &= p(\boldsymbol{\theta})\cdot L(\boldsymbol{\theta}|\boldsymbol{\omega}) \\ &\propto L(\boldsymbol{\theta}|\boldsymbol{\omega}) \\ &\propto \exp\left[-\frac{1}{2\sigma^2}(U_1^2+\sum_{i=2}^{n}U_i^2)\right], \\ \dot{l}(y,\boldsymbol{\theta}|\boldsymbol{\omega}) &= \left(\frac{\beta_k}{\sigma^2}(U_1-\rho U_2),\ \ldots,\ \frac{\beta_k}{\sigma^2}(U_i-\rho U_{i+1}),\ \ldots,\ \frac{\beta_k}{\sigma^2}(U_{n-1}-\rho U_n),\ \frac{\beta_k}{\sigma^2}U_n\right). \end{aligned}$$

Then, it follows that

$$\mathbf{G}(\omega^0) = \text{diag}\left(\mathbf{E}_{\boldsymbol{\omega}^0}\left[\beta_k\frac{\rho^2+1}{\sigma^2}\right],\ldots,\mathbf{E}_{\boldsymbol{\omega}^0}\left[\beta_k\frac{\rho^2+1}{\sigma^2}\right],\mathbf{E}_{\boldsymbol{\omega}^0}\left[\beta_k\frac{1}{\sigma^2}\right]\right),$$

where

$$\begin{aligned} U_1 &= y_1-(x_{1k}+\omega_1)\beta_k - x_{1(-k)}^{\top}\beta_{(-k)} - t_1^{\top}\alpha - z_1^{\top}b + \sqrt{2/\pi}\lambda - \lambda h_1 \\ U_i &= y_i-(x_{ik}+\omega_i)\beta_k - x_{i(-k)}^{\top}\beta_{(-k)} - t_i^{\top}\alpha - z_i^{\top}b - \rho[y_{i-1}-(x_{i-1,k}+\omega_{i-1})\beta_k \\ &\quad - x_{i-1,(-k)}^{\top}\beta_{(-k)} - t_{i-1}^{\top}\alpha - z_{i-1}^{\top}b] + \sqrt{2/\pi}\lambda - \lambda h_i, \qquad i=2,3,\ldots,n. \end{aligned}$$

A5: Metropolis-Hastings algorithm

As a widely used MCMC method, the Metropolis-Hastings (MH) algorithm generates samples from complex probability distributions where direct sampling is challenging. For detailed introductions, we refer to two books Albert (2007) and Gelman et al. (2013), which provide a comprehensive overview of Bayesian statistical methods, including WinBUGS, covering both theory and practical applications. Here's how the MH algorithm works:

Step 1. Initialization: Start with an initial sample from the target distribution, preferably drawn from a distribution similar to the target.

Step 2. Proposal Distribution: Choose a proposal distribution for suggesting new samples, often selected for convenience and ease of sampling.

Step 3. Proposing a Candidate: Generate a new sample (candidate sample) by drawing from the proposal distribution.

Step 4. Acceptance Probability: Calculate the acceptance probability for the candidate sample based on the ratio of the target distribution at the candidate sample and the current sample.

Step 5. Acceptance or Rejection: Accept the candidate sample with a probability determined by the acceptance probability; otherwise, reject it and remain at the current sample.

Step 6. Iteration: Repeat Steps 3-5 for numerous iterations until convergence criteria are met.

The algorithm constructs a Markov chain with the desired target distribution as its stationary distribution, leading to convergence of the generated samples over iterations.

To address low acceptance rates in MCMC sampling, careful selection of the proposed distribution $g(x)$ is crucial. This distribution should encompass the support set of the target distribution; be easily samplable, often chosen from known distributions like the normal or student distributions; facilitate easy calculation of the acceptance probability; have a thicker tail compared to the target distribution; and minimise the frequency of rejecting new candidate points. Adhering to these conditions ensures that the resulting Markov chain satisfies normalisation conditions and possesses a stationary distribution $f(x)$, crucial for effective sampling from complex target distributions.

Given $f(x)$ as the target distribution (posterior distribution), $g(x)$ as the proposed distribution, and $h(x)$ as the acceptance rate, we iteratively generate a sample sequence $\rho_0^0, \rho_0^1, \cdots, \rho_0^N$ from the target probability distribution $f(x)$. Below is our pseudo-code outlining the MH algorithm:

Algorithm 1 Metropolis-Hastings algorithm

Data: ρ_0^0 #Select a random initial sample ρ_0^0
Result: Sampling sequence $\{\rho_0^0, \rho_0^1, \rho_0^2, \cdots, \rho_0^N\}$
for $t = 1$ *to* N **do**
  $y \sim g\,(\cdot|\rho_0^{t-1})$ #For the t-th sample, repeat the following steps.
  $h\,(\rho_0^{t-1}|y) \leftarrow \min\{1, \frac{f(y)g(\rho_0^{t-1}|y)}{f(\rho_0^{t-1})g(y|\rho_0^{t-1})}\}$ #Calculate the acceptance rate
  if $r \sim U(0,1) \leq h(\rho_0^{t-1}|y) \quad (where \;\; r \sim U(0,1))$ **then**
    $\rho_0^t \leftarrow y$;
  else
    $\rho_0^t \leftarrow \rho_0^{t-1}$;
  end
end

A6: The sampling process of parameters in parameter estimation (simulation)

In the iterative process of parameter estimation, it is evident that the vast majority of parameters have reached a convergent state. The primary emphasis of our paper lies in statistical diagnostics rather than extensive parameter estimation. Therefore, our requirement may be limited to obtaining a parameter estimate that closely approximates the true value.

Figures 11, 12, 13, 14 and 15 depict the iterative process of three Markov chains through curves represented in three distinct colors. These curves illustrate three independent parameter estimation sampling processes, offering valuable insights into the fluctuation range of each sampling procedure.

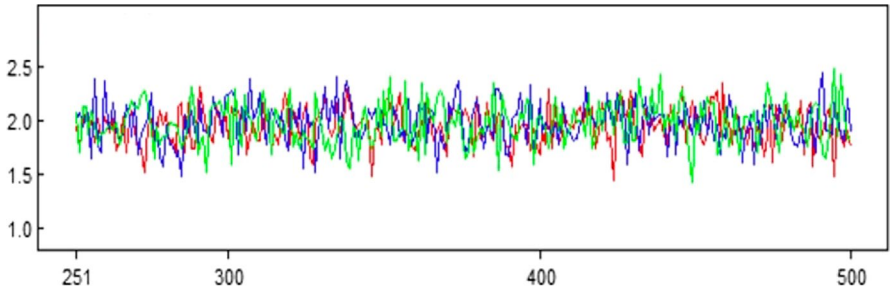


Fig. 11 The iterative process of parameter β_1

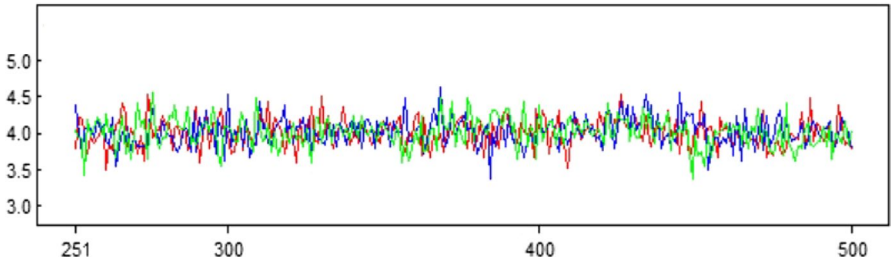


Fig. 12 The iterative process of parameter β_2

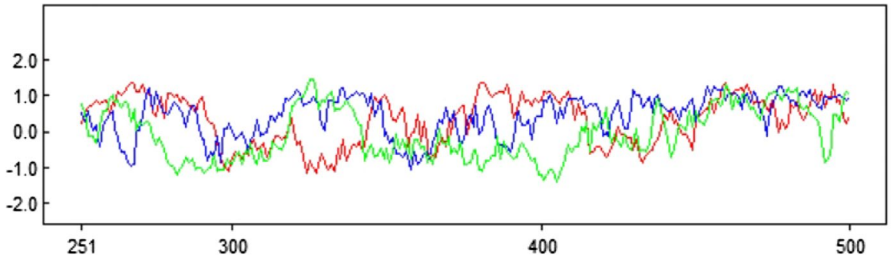


Fig. 13 The iterative process of parameter λ

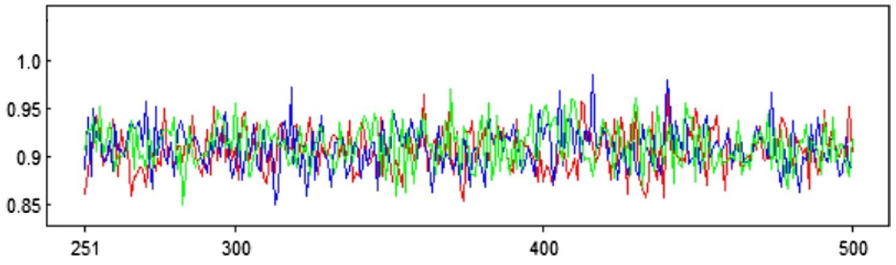


Fig. 14 The iterative process of parameter ρ

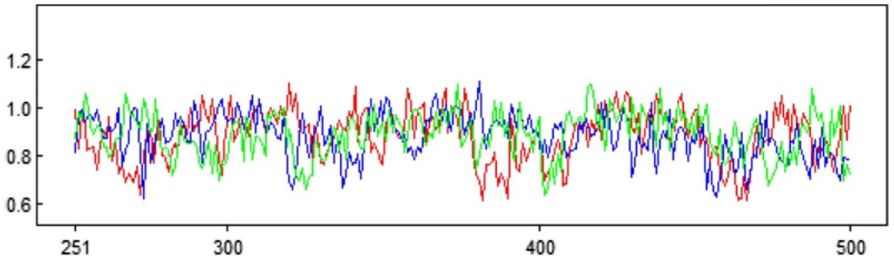


Fig. 15 The iterative process of parameter σ

Acknowledgements We extend our sincere gratitude to the Reviewers and Editors for their constructive and insightful comments, which have significantly enhanced the presentation of our manuscript.

Declarations

Conflict of interest There are no conflict of interest.

References

Albert J (2007) Bayesian Computation with R. Springer
Azzalini A (1985) A class of distribution which includes the normal ones. Scandinavian Journal of Statistics 12:171–178
Cardozo CA, Paula GA, Vanegas LH (2022) Generalized log-gamma additive partial linear models with P-spline smoothing. Statistical Papers 63(6):1953–1978
Cook RD (1986) Assessment of local influence. Journal of the Royal Statistical Society B 48:133–169
Crainiceanu C, Ruppert D, Wand M (2005) Bayesian analysis for penalized spline regression using WinBUGS. Journal of Statistical Software 14(14):1–24
Dai XW, Jin LB, Tian MZ, Shi L (2019) Bayesian local influence for spatial autoregressive models with heteroscedasticity. Statistical Papers 60(5):1423–1446
Dominici F, McDermott A, Hastie T (2004) Improved semiparametric time series models of air pollution and mortality. Journal of the American Statistical Association 9(468):938–948
Ferreira CS, Montoril MH, Paula GA (2022) Partially linear models with p-order autoregressive skew-normal errors. Brazilian Journal of Probability and Statistics 36(4):792–806
Ferreira CS, Paula GA (2017) Estimation and diagnostic for skew-normal partially linear models. Journal of Applied Statistics 44(16):3033–3053
Ferreira CS, Paula GA, Lana GC (2022) Estimation and diagnostic for partially linear models with first-order autoregressive skew-normal errors. Computational Statistics 37:445–468
Ferreira G, Castro LM, Lachos VH, Dias R (2013) Bayesian modeling of autoregressive partial linear models with scale mixture of normal errors. Journal of Applied Statistics 8:1976–1816
Galea MP, Paula GA, Bolfarine H (1997) Local influence in elliptical linear regression models. Journal of the Royal Statistical Society: Series D (The Statistician) 46:71–79
Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A, Rubin DB (2013) Bayesian Data Analysis, Third Edition. Chapman and Hall/CRC
Hao HX, Lin JG (2019) Wang HX (2019) Bayesian local influence analysis and its application of GARCH model. Journal of Mathematical Statistics and Management 4:602–618
Ju YY, Yang Y, Hu MX, Dai L, Wu LC (2022) Bayesian influence analysis of the skew-normal spatial autoregression models. Mathematics 10(8):1306
Lee SY, Xu L (2004) Influence analysis of nonlinear mixed-effects models. Computational Statistics & Data Analysis 45:321–441
Liu S (2000) On local influence for elliptical linear models. Statistical Papers 41:211–224

Liu S (2004) On diagnostics in conditionally heteroscedastic time series models under elliptical distributions. Journal of Applied Probability 41A:394–405
Liu S, Leiva V, Zhuang D, Ma T, Figueroa-Zúñiga JI (2022) Matrix differential calculus with applications in the multivariate linear model and its diagnostics. Journal of Multivariate Analysis 188:104849
Liu YH, Mao GH, Leiva V, Liu S, Tapia A (2020) Diagnostic analytics for an autoregressive model under the skew-normal distribution. Mathematics 8(5):693
Liu YH, Sang RC, Liu S (2017) Diagnostic analysis for a vector autoregressive model under Student's t -distributions. Statistica Neerlandica 71(2):86–114
Liu YH, Wang J, Leiva V, Tapia A, Tan W, Liu S (2024) Robust autoregressive modeling and its diagnostic analytics with a COVID-19 related application. Journal of Applied Statistics 51(7):1318–1343
Liu YH, Wang J, Shi DW, Leiva V, Liu S (2023) A score test for detecting extreme values in a vector autoregressive model. Journal of Statistical Computation and Simulation 93(15):2751–2779
Liu YH, Wang J, Yao Z, Liu C, Liu S (2022) Diagnostic analytics for a GARCH model under skew-normal distributions. Communication in Statistics -Simulation and Computation. https://doi.org/10.1080/03610918.2022.2157015
Marriott J, Newbold P (1998) Bayesian comparison of ARIMA and stationary ARMA models. International Statistical Review 66(3):323–336
Oliveira RA, Paula GA (2021) Additive models with autoregressive symmetric errors based on penalized regression splines. Computational Statistics 36(4):2435–2466
Paula GA, Leiva V, Barros M, Liu S (2012) Robust statistical modeling using the Birnbaum-Saunders-t distribution applied to insurance. Applied Stochastic Models in Business & Industry 28(1):16–34
Poon WY, Poon YS (1999) Conformal normal curvature and assessment of local influence. Journal of the Royal Statistical Society B 61:51–61
Ruppert D, Wand M, Carroll R (2003) Semiparametric Regression. Cambridge University Press, Cambridge, UK
Sahu SK, Dey DK, Branco MD (2003) A new class of multivariate distributions with applications to Bayesian regression models. Canadian Journal of Statistics 31:129–150
Seber GAF, Wild CL (1989) Nonlinear Regression. Wiley, New York
Sturtz S, Ligges U, Gelman A (2005) R2WinBUGS: A package for running WinBUGS from R. Journal of Statistical Software 12(3):1–16
Tang N, Duan XD (2014) Bayesian influence analysis of generalized partial linear mixed models for longitudinal data. Journal of Multivariate Analysis 126:86–99
Zeger S, Diggle P (1994) Semiparametric models for longitudinal data with application to CD4 cell numbers in HIV seroconverters. Biometrics 50(3):689–699
Zhu H, Ibrahim JG, Tang N (2011) Bayesian influence analysis: a grometric approach. Biometrika 98(2):307–323
Zhu H, Lee SY (2001) Local influence for incomplete data models. Journal of the Royal Statistical Society Series B 63(1):111–126