Σ *mathematics*

MDPI

Article

Diagnostic Analytics for an Autoregressive Model under the Skew-Normal Distribution

Yonghui Liu [1], Guohua Mao [2], Víctor Leiva [3] (ID), Shuangzhe Liu [4] and Alejandra Tapia [5,*] (ID)

[1]   School of Statistics and Information, Shanghai University of International Business and Economics, Shanghai 201620, China; liuyh@lsec.cc.ac.cn

[2]   School of Mathematics, Shanghai University of Finance and Economics, Shanghai 200433, China; huazai_mao@126.com

[3]   School of Industrial Engineering, Pontificia Universidad Católica de Valparaíso, Valparaíso 2362807, Chile; victorleivasanchez@gmail.com

[4]   Faculty of Science and Technology, University of Canberra, Bruce, ACT 2617, Australia; Shuangzhe.Liu@canberra.edu.au

[5]   School of Engineering in Statistics, Universidad Católica del Maule, Talca 3466706, Chile

*   Correspondence: atapias@ucm.cl

check for updates

Abstract: Autoregressive models have played an important role in time series. In this paper, an autoregressive model based on the skew-normal distribution is considered. The estimation of its parameters is carried out by using the expectation–maximization algorithm, whereas the diagnostic analytics are conducted by means of the local influence method. Normal curvatures for the model under four perturbation schemes are established. Simulation studies are conducted to evaluate the performance of the proposed procedure. In addition, an empirical example involving weekly financial return data are analyzed using the procedure with the proposed diagnostic analytics, which has improved the model fit.

Keywords: AR models; EM algorithm; local influence method; maximum likelihood estimation

1. Introduction

For time series data analysis, autoregressive (AR) modeling is an essential technique and has been applied in many areas including biology, chemistry, earth sciences, economics, education, engineering, finance, health, medicine, and physics; see, for example, [1,2] for recent accounts of time series modeling and applications. Issues related to estimation and testing for AR models are extensive and well-established; see [3,4] for related issues to statistical diagnostics which are of equal importance. Local influence diagnostics is the study of how relevant minor perturbations impact the fit of the model and the results of statistical inference. This has become a useful statistical methodology after [5] introduced the idea of local influence to aid in the identification of potentially influential observations. Diagnostic analytics is used in a number of regression and time series models. Among others, Refs. [6–15] investigated the local influence of linear or nonlinear regression models under non-normal distributional assumptions. In a framework of time series data, Refs. [16–18] considered influence diagnostics for AR and vector AR models under normal or elliptical distributions.

A standard assumption for time series models is that their errors are mutually independent and follow normal or symmetric distributions, as studied in [16–18]. However, it is known that certain financial and other datasets feature errors with skewed distributions. In order to deal with such data, the skew-normal (SN) distribution and its scale-mixtures have provided an appealing alternative and can therefore be adopted. Their properties, extensions, and applications are becoming increasingly popular; see [19–26]. In addition, Refs. [27–29] studied SN linear, linear mixed, and nonlinear

models. Ref. [8] analyzed diagnostics in the nonlinear model with scale mixtures of SN and AR errors. For particular financial applications, Ref. [30] looked at asset pricing issues with return data following SN models. To our knowledge, no study on influence diagnostics in AR time series models under SN distributions have been reported. Therefore, the objective of the present paper is to formulate an AR model under the SN distribution (SN AR model) and to derive diagnostic analytics with applications to financial data. We use the matrix differential calculus pioneered by [31] to establish the mathematical results used in our data analysis. We implement the maximum likelihood (ML) method with the expectation–maximization (EM) algorithm to estimate the SN AR model parameters, whereas the local influence method with four perturbation schemes is used for the diagnostic analytics. The EM algorithm has now become a popular iterative technique for the ML estimation method with incomplete data; see [32,33]. The paper proceeds as follows. In Section 2, the SN AR model is introduced and the ML estimations of the model parameters are derived. Section 3 presents the local influence method and establishes normal curvatures under four perturbation schemes. In Section 4, a simulation study and an empirical example involving an AR model are presented, while the effectiveness of the proposed diagnostics is illustrated and discussed. Our concluding remarks are addressed in Section 5. The derivations of the normal curvatures are presented in the Appendix A.

2. An SN AR Model and its Parameter Estimation

2.1. The SN AR Model

A random variable Y is said to follow an SN distribution with location parameter μ, scale parameter σ^2 and skewness parameter λ, which is denoted by $Y \sim \text{SN}(\mu, \sigma^2, \lambda)$, if its probability density function is given by

$$f(y) = \frac{2}{\sigma} \phi \left(\frac{y - \mu}{\sigma} \right) \Phi \left(\lambda \left(\frac{y - \mu}{\sigma} \right) \right), \quad y \in \mathbb{R}, \mu \in \mathbb{R}, \sigma > 0, \lambda \in \mathbb{R}, \tag{1}$$

where ϕ and Φ are the standard normal probability density and cumulative distribution functions, respectively. We see that, if $\lambda = 0$, then the probability density function of Y defined in (1) reduces to the normal probability density function. If $Y \sim \text{SN}(\mu, \sigma^2, \lambda)$, then $\text{E}(Y) = \mu + \sigma\delta\sqrt{2/\pi}$ and $\text{Var}(Y) = \sigma^2 - (2/\pi)\sigma^2\delta^2$, where $\delta = \lambda/\sqrt{1 + \lambda^2}$.

Let Y_t be generated by a stationary AR(p) process given by

$$Y_t = \beta_1 y_{t-1} + \cdots + \beta_p y_{t-p} + u_t = \boldsymbol{x}_t^\top \boldsymbol{\beta} + u_t, \quad t = 1, \ldots, T, \tag{2}$$

where Y_t is a time series, with $Y_1, \ldots, Y_p$ being the p initial values for y_{t-j}, such that $\boldsymbol{x}_t = (y_{t-1}, \ldots, y_{t-p})^\top$ is the $p \times 1$ vector of SN AR(p) covariates; β_j is a regression coefficient, for $j = 1, \ldots, p$, such that $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^\top$ is a $p \times 1$ regression coefficient vector, and u_t is the scalar error term which follows an SN distribution, that is, $u_t \sim \text{SN}(0, \sigma^2, \lambda)$, where σ^2 is the scale parameter and λ is the skewness parameter. Thus, $\boldsymbol{\theta} = (\boldsymbol{\beta}, \sigma^2, \lambda)^\top$ is the $(p + 2) \times 1$ vector of SN AR(p) model parameters. Note that we do not include an intercept in the SN model above. The expected value of u_t is not zero, due to the assumption that the expected value of the underlying normal distribution is zero. Then, we choose the non-zero expected value of u_t as an approximate to replace an intercept for the SN model.

2.2. ML Estimation

Finding the ML estimate of the parameter vector $\boldsymbol{\theta}$ by direct maximization of the log-likelihood can potentially be a difficult task, so we implement the EM algorithm for this estimation. Let $\boldsymbol{y}_c = (\boldsymbol{y}_o, \boldsymbol{y}_m)^\top$ denote the complete data, with $\boldsymbol{y}_o$ being the observed data and $\boldsymbol{y}_m$ the missing data. We use the notation $\boldsymbol{Y}_c, \boldsymbol{Y}_o, \boldsymbol{Y}_m$ for the random vectors associated with $\boldsymbol{y}_c, \boldsymbol{y}_o$ and $\boldsymbol{y}_m$, respectively. Starting from $\boldsymbol{\theta}^{(0)}$ as the

initial estimate, we can get $\theta^{(0)}, \theta^{(1)}, \ldots$ by running the two steps of the EM algorithm iteratively as defined below.

In order to implement the EM algorithm, consider the stochastic representation given by

$$Y = \mu + \sigma \delta H + \sigma \sqrt{(1 - \delta^2)} H_1, \tag{3}$$

where $H = |H_0| \sim HN(0,1)$, with HN being used to indicate the half normal distribution. In addition, both H_0 and H_1 are independent random variables which follow a standard normal distribution. Note that, from (3), we have $Y|H = h \sim N(\mu + h\lambda\sigma/\sqrt{1 + \lambda^2}, \sigma^2/(1 + \lambda^2))$ and $H \sim HN(0,1)$. Hence, by considering $\boldsymbol{y}_o = (y_{p+1}, \ldots, y_T)^\top$, $\boldsymbol{y}_m = (h_{p+1}, \ldots, h_T)^\top$ and $\boldsymbol{y}_c = (\boldsymbol{y}_o, \boldsymbol{y}_m)^\top$ as the observed, missing and complete data sets, respectively, we have the complete-data log-likelihood function for $\boldsymbol{\theta} = (\boldsymbol{\beta}, \sigma^2, \lambda)^\top$ given by

$$\ell_c(\boldsymbol{\theta}, \boldsymbol{y}_c) = \sum_{t=p+1}^{T} \left(-\frac{1}{2}\log(\sigma^2) + \frac{1}{2}\log(1 + \lambda^2) - \frac{1 + \lambda^2}{2\sigma^2}\left(y_t - \boldsymbol{x}_t^\top \boldsymbol{\beta} - \frac{\lambda\sigma}{\sqrt{1 + \lambda^2}} h_t\right)^2 \right). \tag{4}$$

Therefore, for the E-step of the EM algorithm, given the current estimate $\widehat{\boldsymbol{\theta}}^{(k)}$ and based on (4), we can calculate the function $Q(\boldsymbol{\theta})|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}^{(k)}} = E[\ell_c(\boldsymbol{\theta}, \boldsymbol{Y}_c)|\boldsymbol{Y}_o = \boldsymbol{y}_o]|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}^{(k)}}$ as

$$Q(\boldsymbol{\theta}) = -\frac{(T-p)}{2}\log(\sigma^2) + \frac{(T-p)}{2}\log(1 + \lambda^2) - \frac{(1 + \lambda^2)}{2}\sum_{t=p+1}^{T}\left(\frac{y_t - \boldsymbol{x}_t^\top \boldsymbol{\beta}}{\sigma} - \frac{\lambda}{\sqrt{1 + \lambda^2}}\widehat{c}_t\right)^2 - \frac{\lambda^2}{2}\sum_{t=p+1}^{T}(\widehat{c_t^2} - (\widehat{c}_t)^2), \tag{5}$$

where $\widehat{c}_t = E(H_t|\boldsymbol{Y}_o = \boldsymbol{y}_o)|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}^{(k)}} = \tau_1 + \tau_2\phi(\tau_1/\tau_2)/\Phi(\tau_1/\tau_2)$, $\widehat{c_t^2} = E(H_t^2|\boldsymbol{Y}_o = \boldsymbol{y}_o)|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}^{(k)}} = \tau_1^2 + \tau_2^2 + \tau_1\tau_2\phi(\tau_1/\tau_2)/\Phi(\tau_1/\tau_2)$, and $\tau_1 = (\widehat{\lambda}^{(k)}/\widehat{\sigma}^{(k)})(1 + (\widehat{\lambda^{(k)}})^2)^{1/2}(y_t - \boldsymbol{x}_t^\top \widehat{\boldsymbol{\beta}}^{(k)})$, $\tau_2 = 1/(1 + (\widehat{\lambda^{(k)}})^2)^{1/2}$.

Note that equations in (4) and (5) are obtained from the probability density function given in (1) and after calculating the expected value of (4). In addition, $\widehat{c_t^2}$ is different from $(\widehat{c}_t)^2$. For the M-step, we update $\widehat{\boldsymbol{\theta}}^{(k)}$ by the Newton–Raphson iteration $\dot{Q}(\widehat{\boldsymbol{\theta}}^{(k+1)}) = \dot{Q}(\widehat{\boldsymbol{\theta}}^{(k)}) + \ddot{Q}(\widehat{\boldsymbol{\theta}}^{(k)})(\widehat{\boldsymbol{\theta}}^{(k+1)} - \widehat{\boldsymbol{\theta}}^{(k)}) + o(|\widehat{\boldsymbol{\theta}}^{(k+1)} - \widehat{\boldsymbol{\theta}}^{(k)}|)$, where $\dot{Q}$ denotes the gradient vector and $\ddot{Q}$ denotes the Hessian matrix. As it is known, a suitable initial value $\widehat{\boldsymbol{\theta}}^{(0)}$ is important and difficult to find in numerical computation. Thus, we can get $\widehat{\boldsymbol{\theta}}^{(0)} = (\widehat{\boldsymbol{\beta}}^{(0)}, \widehat{\sigma^2}^{(0)}, \widehat{\lambda}^{(0)})$ considering $\widehat{\boldsymbol{\beta}}^{(0)}$ as the ordinary least squares (OLS) estimate and then $\widehat{\sigma^2}^{(0)}, \widehat{\lambda}^{(0)}$ can be computed as $\widehat{\boldsymbol{\theta}} = (\widehat{\boldsymbol{\beta}}, \widehat{\sigma^2}, \widehat{\lambda})$ until $|\widehat{\boldsymbol{\theta}}^{(k+1)} - \widehat{\boldsymbol{\theta}}^{(k)}| < 10^{-5}$.

2.3. The Hessian Matrix

Next, we compute the Hessian matrix $\ddot{Q}(\boldsymbol{\theta})$ evaluated at the ML estimate $\widehat{\boldsymbol{\theta}}$ using

$$\ell_c(\boldsymbol{\theta}, \boldsymbol{y}_c) = \sum_{t=p+1}^{T} \left(-\frac{1}{2}\log(\sigma^2) + \frac{1}{2}\log(1 + \lambda^2) - \frac{1 + \lambda^2}{2\sigma^2}\left(u_t - \frac{\lambda\sigma}{\sqrt{1 + \lambda^2}} h_t\right)^2 \right). \tag{6}$$

Then, based on (6), the Q function is given by

$$
\begin{aligned}
Q(\boldsymbol{\theta})|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}} &= E[\ell_c(\boldsymbol{\theta}, \boldsymbol{Y}_c)|\boldsymbol{Y}_o = \boldsymbol{y}_o]|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}} \\
&= -\frac{(T-p)}{2}\log(\sigma^2) + \frac{(T-p)}{2}\log(1 + \lambda^2) \\
&\quad -\frac{(1 + \lambda^2)}{2}\sum_{t=p+1}^{T}\left(\frac{u_t}{\sigma} - \frac{\lambda}{\sqrt{1 + \lambda^2}}\widehat{c}_t\right)^2 - \frac{\lambda^2}{2}\sum_{t=p+1}^{T}(\widehat{c_t^2} - (\widehat{c}_t)^2).
\end{aligned}
$$

For the SN AR(p) model, we have the $(p+2) \times (p+2)$ Hessian matrix $\ddot{Q}(\widehat{\boldsymbol{\theta}})$ given by

$$
\ddot{Q}(\widehat{\boldsymbol{\theta}}) = \begin{pmatrix} \frac{\partial^2 Q(\boldsymbol{\theta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^\top} & \frac{\partial^2 Q(\boldsymbol{\theta})}{\partial \boldsymbol{\beta} \partial \sigma^2} & \frac{\partial^2 Q(\boldsymbol{\theta})}{\partial \boldsymbol{\beta} \partial \lambda} \\[2mm] \frac{\partial^2 Q(\boldsymbol{\theta})}{\partial \sigma^2 \partial \boldsymbol{\beta}} & \frac{\partial^2 Q(\boldsymbol{\theta})}{\partial (\sigma^2)^2} & \frac{\partial^2 Q(\boldsymbol{\theta})}{\partial \sigma^2 \partial \lambda} \\[2mm] \frac{\partial^2 Q(\boldsymbol{\theta})}{\partial \lambda \partial \boldsymbol{\beta}} & \frac{\partial^2 Q(\boldsymbol{\theta})}{\partial \lambda \partial \sigma^2} & \frac{\partial^2 Q(\boldsymbol{\theta})}{\partial \lambda^2} \end{pmatrix} \Bigg|_{\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}'} \tag{7}
$$

where $\widehat{\boldsymbol{\theta}} = (\widehat{\boldsymbol{\beta}}, \widehat{\sigma}^2, \widehat{\lambda})$ is the ML estimate of $\boldsymbol{\theta}$. The expression for $\partial^2 Q(\boldsymbol{\theta})/\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^\top$ and the other submatrices are presented in Appendix A.1.

3. Influence Analysis for the SN AR Model

3.1. Local Influence

Let $\ell(\boldsymbol{\theta})$ denote the log-likelihood function for the model given in (2), which is the postulated model, where $\boldsymbol{\theta}$ is a $(p+2) \times 1$ vector of unknown parameters with its ML estimate $\widehat{\boldsymbol{\theta}}$. Let $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_q)^\top$ denote a $q \times 1$ vector of perturbations confined to some open subset of $\mathbb{R}^q$ and let $\boldsymbol{\omega}_0$ denote a $q \times 1$ non-perturbation vector, with q as a suitable number of dimensions and $\boldsymbol{\omega}_0 = (0, \ldots, 0)$, or $\boldsymbol{\omega}_0 = (1, \ldots, 1)$, or a third choice, depending on the context. Then, $\ell(\boldsymbol{\theta})$ and $\ell(\boldsymbol{\theta}|\boldsymbol{\omega})$ represent the log-likelihood functions of the postulated and perturbed models, respectively. Note that $\ell(\boldsymbol{\theta}) = \ell(\boldsymbol{\theta}|\boldsymbol{\omega}_0)$.

We suppose that $\ell(\boldsymbol{\theta}|\boldsymbol{\omega})$ is twice continuously differentiable in a neighborhood of $(\widehat{\boldsymbol{\theta}}, \boldsymbol{\omega}_0)$. We are interested in the comparison of $\widehat{\boldsymbol{\theta}}$ and $\widehat{\boldsymbol{\theta}}_{\boldsymbol{\omega}}$ using the local influence method idea, which is to investigate the degree of inference affected by those minor changes in the corresponding perturbations. Ref. [5] used likelihood displacement (LD) to assess the influence of the perturbation $\boldsymbol{\omega}$ defined as $\mathrm{LD}(\boldsymbol{\omega}) = 2(\ell(\widehat{\boldsymbol{\theta}}) - \ell(\widehat{\boldsymbol{\theta}}_{\boldsymbol{\omega}}))$. Here, large values of $\mathrm{LD}(\boldsymbol{\omega})$ indicate that $\widehat{\boldsymbol{\theta}}$ and $\widehat{\boldsymbol{\theta}}_{\boldsymbol{\omega}}$ differ considerably in relation to the contours of the non-perturbed log-likelihood function $\ell(\boldsymbol{\theta})$. The idea is based on studying the local behavior of $\mathrm{LD}(\boldsymbol{\omega})$ and the normal curvature $C_l(\boldsymbol{\theta})$ in a unit-length direction vector l, where $||l|| = 1$. According to [5], the normal curvature used to examine the local influence of the perturbation vector at $\boldsymbol{\omega} = \boldsymbol{\omega}_0$ is $C_l(\boldsymbol{\theta}) = 2|l^\top \ddot{F} l| = 2|l^\top (\boldsymbol{\Delta}^\top \ddot{\ell}^{-1} \boldsymbol{\Delta}) l|$, with

$$
\ddot{F} = \frac{\partial^2 \ell(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial \boldsymbol{\omega} \partial \boldsymbol{\omega}^\top}, \quad \boldsymbol{\Delta} = \frac{\partial^2 \ell(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\omega}^\top}, \quad \ddot{\ell} = \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top},
$$

where l is a $q \times 1$ vector of unit length, $-\ddot{\ell}$ is the $(p+2) \times (p+2)$ observed information matrix for the postulated model, $\boldsymbol{\Delta}$ is the $(p+2) \times q$ perturbation matrix for the perturbed model, and $-\ddot{\ell}$ and $\boldsymbol{\Delta}$ are evaluated at $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}$ and $\boldsymbol{\omega} = \boldsymbol{\omega}_0$. The suggestion is to make the local influence diagnostic analytics by finding the maximum curvature $C_{\max} = \max_{||l||=1} C_l$, where $C_{\max}$ corresponds to the largest absolute eigenvalue $\lambda_{\max}$ and its associated eigenvector $l_{\max}$ of the matrix $\ddot{F} = \boldsymbol{\Delta}^\top \ddot{\ell}^{-1} \boldsymbol{\Delta}$. If the absolute value of the ith element of $l_{\max}$ is the largest, then the ith observation in the data may be the most influential. To examine the magnitude of influence, it is useful to have a benchmark value for $C_{\max}$ and for the elements of $l_{\max}$, see [10,18,34].

3.2. Local Influence for the SN AR Model

Next, we conduct a local influence diagnostic analytics for the SN AR(p) model. Due to the complexity of the SN distribution, we obtain the ML estimates based on the EM algorithm. As suggested by [29,34], the Q function and a Q function obtained similarly as LD can be used to

replacethe log-likelihood function and likelihood displacement in the method of [5], in order to assess the influence of the perturbation. Thus, the normal curvature should be changed to be

$$C_l(\boldsymbol{\theta}) = 2|l^\top \ddot{F} l| = 2|l^\top (\boldsymbol{\Delta}^\top \ddot{Q}^{-1} \boldsymbol{\Delta}) l|, \tag{8}$$

with

$$\ddot{F} = \frac{\partial^2 Q(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial \boldsymbol{\omega} \partial \boldsymbol{\omega}^\top}, \quad \boldsymbol{\Delta} = \frac{\partial^2 Q(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\omega}^\top}, \quad \ddot{Q} = \frac{\partial^2 Q(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top},$$

where l is a $q \times 1$ vector of unit length, and $\ddot{F}$, $\ddot{Q}$ and $\boldsymbol{\Delta}$ are $q \times q$, $(p+2) \times (p+2)$ and $(p+2) \times q$ matrices, respectively. In addition, $\ddot{Q}$ and $\boldsymbol{\Delta}$ need to be evaluated at $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}$ and $\boldsymbol{\omega} = \boldsymbol{\omega}_0$.

The method examines the total local influence $C_t = C_{l_t}(\boldsymbol{\theta})$, where l_t is a $q \times 1$ unit-length vector with one at the tth position and zeros elsewhere. We denote $S = -\boldsymbol{\Delta}^\top \ddot{Q}^{-1} \boldsymbol{\Delta}$. Since $C_l(\boldsymbol{\theta})$ is not invariant under a uniform change of scale, Ref. [34] proposed the conformal normal curvature $B_l(\boldsymbol{\theta}) = C_l(\boldsymbol{\theta})/(2\mathrm{tr}(S))$. An interesting property of the conformal normal curvature is that, for any unit-length direction l, we have $0 \leq B_l(\boldsymbol{\theta}) \leq 1$, which allows comparison of curvatures among different models.

In order to determine if the tth observation is influential, Ref. [34] proposed to classify the tth observation as possibly influential if $M(0)_t = B_{l_t}$ is greater than the benchmark $1/q + c^* \mathrm{SM}(0)$, where $\mathrm{SM}(0)$ is the sample standard error of $M(0)_k$, for $k = 1, \ldots, q$, and c^* is a certain constant. Depending on the specific application, c^* may be taken to be a suitably selected positive value.

The forms given in (8) are used to obtain our normal curvature results under the four perturbation schemes, namely the case-weights, data, variance parameter, and skewness parameter schemes. The matrices $\ddot{Q}$ and $\boldsymbol{\Delta}$ need to be established for each scheme.

We employ the matrix differential calculus proposed by [31] to establish these algebraic results in the following sections. We present their derivations in their respective Appendix A.

3.3. Perturbation Matrices

3.3.1. Perturbation of Case-Weights

Assume that a minor perturbation is made on the SN AR(p) model, with $y_t = \boldsymbol{x}_t^\top \boldsymbol{\beta} + u_t$ being replaced by $\omega_t y_t = \omega_t \boldsymbol{x}_t^\top \boldsymbol{\beta} + \omega_t u_t$, where ω_t is the weight. Let $\boldsymbol{\omega} = (\omega_{p+1}, \ldots, \omega_T)^\top$ denote the $(T-p) \times 1$ perturbation vector and $\boldsymbol{\omega}_0 = (1, \ldots, 1)^\top$ denote the $(T-p) \times 1$ non-perturbation vector. Then, the complete-data log-likelihood function of the perturbed model is given by

$$\ell_c(\boldsymbol{\theta}, \boldsymbol{\omega}, \boldsymbol{y}_c) = \sum_{t=p+1}^{T} \left(-\frac{1}{2} \log(\sigma^2) + \frac{1}{2} \log(1+\lambda^2) - \frac{1+\lambda^2}{2\sigma^2} \left(\omega_t u_t - \frac{\lambda\sigma}{\sqrt{1+\lambda^2}} h_t \right)^2 \right). \tag{9}$$

Thus, the perturbed Q function obtained from (9) is expressed as

$$\begin{aligned}
Q(\boldsymbol{\theta}, \boldsymbol{\omega})|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}} &= \mathrm{E}[\ell_c(\boldsymbol{\theta}, \boldsymbol{\omega}, \boldsymbol{Y}_c)|\boldsymbol{Y}_o = \boldsymbol{y}_o]|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}} \\
&= -\frac{(T-p)}{2} \log(\sigma^2) + \frac{(T-p)}{2} \log(1+\lambda^2) \\
&\quad -\frac{1+\lambda^2}{2} \sum_{t=p+1}^{T} \left(\frac{\omega_t u_t}{\sigma} - \frac{\lambda}{\sqrt{1+\lambda^2}} \widehat{c}_t \right)^2 - \frac{\lambda^2}{2} \sum_{t=p+1}^{T} (\widehat{c_t^2} - (\widehat{c}_t)^2).
\end{aligned} \tag{10}$$

For the SN AR(p) model in the perturbation of case-weights, the $(p+2) \times (T-p)$ perturbation matrix $\boldsymbol{\Delta}$ must be evaluated at $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}$ and $\boldsymbol{\omega} = \boldsymbol{\omega}_0$ after taking derivatives, obtaining

$$\boldsymbol{\Delta} = \frac{\partial^2 Q(\boldsymbol{\theta}, \boldsymbol{\omega})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\omega}}\bigg|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}, \boldsymbol{\omega}=\boldsymbol{\omega}_0} = \begin{pmatrix} \frac{1+\widehat{\lambda}^2}{\widehat{\sigma}^2} \widehat{u}_t \boldsymbol{x}_t^\top \\ \frac{1+\widehat{\lambda}^2}{2\widehat{\sigma}^3} \left(\frac{2\widehat{u}_t^2}{\widehat{\sigma}} - \frac{\widehat{\lambda}\widehat{c}_t}{\sqrt{1+\widehat{\lambda}^2}} \widehat{u}_t \right) \\ -\frac{2\widehat{\lambda}}{\widehat{\sigma}^2} \widehat{u}_t^2 + \frac{2\widehat{\lambda}+1}{\sqrt{1+\widehat{\lambda}^2}} \frac{\widehat{c}_t}{\widehat{\sigma}} \widehat{u}_t \end{pmatrix}, \tag{11}$$

where

$$\widehat{u}_t = y_t - \boldsymbol{x}_t^\top \widehat{\boldsymbol{\beta}},$$

$$\widehat{c}_t = \mathrm{E}(H_t|Y_0, \widehat{\boldsymbol{\theta}}) = \tau_1 + \frac{\phi(\frac{\tau_1}{\tau_2})}{\Phi(\frac{\tau_1}{\tau_2})}\tau_2,$$

$$\widehat{c_t^2} = \mathrm{E}(H_t{}^2|Y_0, \widehat{\boldsymbol{\theta}}) = \tau_1^2 + \tau_2^2 + \frac{\phi(\frac{\tau_1}{\tau_2})}{\Phi(\frac{\tau_1}{\tau_2})}\tau_1\tau_2,$$

$$\tau_1 = \frac{\widehat{\lambda}}{\widehat{\sigma}\sqrt{1+\widehat{\lambda}^2}}\widehat{u}_t, \quad \tau_2 = \frac{1}{\sqrt{1+\widehat{\lambda}^2}}.$$

3.3.2. Perturbation of Data

Assume that a perturbation replaces y_t by $\omega_t + y_t$. Let $\boldsymbol{\omega} = (\omega_{p+1}, \ldots, \omega_T)^\top$ denote the $(T-p) \times 1$ perturbation vector and $\boldsymbol{\omega}_0 = (0, \ldots, 0)^\top$ the $(T-p) \times 1$ non-perturbation vector. The perturbed AR(p) model can be written as $y_t + \omega_t = \beta_1(y_{t-1} + \omega_{t-1}) + \cdots + \beta_p(y_{t-p} + \omega_{t-p}) + u_t$, where $u_t = y_t - \boldsymbol{x}_t^\top \boldsymbol{\beta} + \mu(\omega_t)$ and $\mu(\omega_t) = \omega_t - \beta_1\omega_{t-1} - \cdots - \beta_p\omega_{t-p}$. Then, the complete-data log-likelihood function of the perturbed model is given by

$$\ell_c(\boldsymbol{\theta}, \boldsymbol{\omega}, \boldsymbol{y}_c) = \sum_{t=p+1}^{T}\left(-\frac{1}{2}\log(\sigma^2) + \frac{1}{2}\log(1+\lambda^2) - \frac{1+\lambda^2}{2\sigma^2}\left(u_t + \mu(\omega_t) - \frac{\lambda\sigma}{\sqrt{1+\lambda^2}}h_t\right)^2\right).$$

Thus, the perturbed Q function is expressed as

$$
\begin{aligned}
Q(\boldsymbol{\theta}, \boldsymbol{\omega})|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}} &= -\frac{(T-p)}{2}\log(\sigma^2) + \frac{(T-p)}{2}\log(1+\lambda^2) \\
&\quad -\frac{1+\lambda^2}{2}\sum_{t=p+1}^{T}\left(\frac{u_t + \mu(\omega_t)}{\sigma} - \frac{\lambda}{\sqrt{1+\lambda^2}}\widehat{c}_t\right)^2 - \frac{\lambda^2}{2}\sum_{t=p+1}^{T}(\widehat{c_t^2} - (\widehat{c}_t)^2).
\end{aligned}
\tag{12}
$$

For the SN AR(p) model in the perturbation of data, we have

$$
\boldsymbol{\Delta} = \frac{\partial^2 Q(\boldsymbol{\theta}, \boldsymbol{\omega})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\omega}}\bigg|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}, \boldsymbol{\omega}=\boldsymbol{\omega}_0} = \begin{pmatrix} \frac{1+\widehat{\lambda}^2}{\widehat{\sigma}^2}\boldsymbol{x}_t^\top \\ \frac{1+\widehat{\lambda}^2}{2\widehat{\sigma}^3}\left(\frac{2\widehat{u}_t^2}{\widehat{\sigma}} - \frac{\widehat{\lambda}}{\sqrt{1+\widehat{\lambda}^2}}\widehat{c}_t\right) \\ -\frac{2\widehat{\lambda}}{\widehat{\sigma}^2}\widehat{u}_t^2 + \frac{2\widehat{\lambda}+1}{\sqrt{1+\widehat{\lambda}^2}}\frac{\widehat{c}_t}{\widehat{\sigma}} \end{pmatrix},
\tag{13}
$$

where $\widehat{u}_t = y_t - \boldsymbol{x}_t^\top\widehat{\boldsymbol{\beta}}$, and $\widehat{h}_t$ is as in the data perturbation.

3.3.3. Perturbation of Scale

Assume that the variance parameter σ^2 in the model is replaced by $\omega_t^{-1}\sigma^2$, meaning that $u_t \sim \mathrm{SN}(0, \omega_t^{-1}\sigma^2, \lambda)$. Let $\boldsymbol{\omega} = (\omega_{p+1}, \ldots, \omega_T)^\top$ denote the $(T-p) \times 1$ perturbation vector and $\boldsymbol{\omega}_0 = (1, \ldots, 1)^\top$ denote the $(T-p) \times 1$ non-perturbation vector. Then, the complete-data log-likelihood function of the perturbed model is given by

$$\ell_c(\boldsymbol{\theta}, \boldsymbol{\omega}, \boldsymbol{y}_c) = \sum_{t=p+1}^{T}\left(-\frac{1}{2}\log(\sigma^2) + \frac{1}{2}\log(\omega_t) + \frac{1}{2}\log(1+\lambda^2) - \frac{\omega_t(1+\lambda^2)}{2\sigma^2}\left(u_t - \frac{\lambda\sigma}{\sqrt{\omega_t(1+\lambda^2)}}h_t\right)^2\right).$$

Thus, the perturbed Q function is expressed as

$$
\begin{aligned}
Q(\boldsymbol{\theta}, \boldsymbol{\omega})|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}} &= -\frac{(T-p)}{2}\log(\sigma^2) + \sum_{t=p+1}^{\top}\frac{1}{2}\log(\omega_t) + \frac{(T-p)}{2}\log(1+\lambda^2) \\
&\quad - \frac{1+\lambda^2}{2}\sum_{t=p+1}^{\top}\left(\frac{\sqrt{\omega_t}u_t}{\sigma} - \frac{\lambda}{\sqrt{1+\lambda^2}}\widehat{c}_t\right)^2 - \frac{\lambda^2}{2}\sum_{t=p+1}^{\top}(\widehat{c_t^2} - (\widehat{c}_t)^2).
\end{aligned}
\tag{14}
$$

For the SN AR(p) model with the perturbation of variance parameter, we have

$$
\boldsymbol{\Delta} = \frac{\partial^2 Q(\boldsymbol{\theta}, \boldsymbol{\omega})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\omega}}\bigg|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}},\boldsymbol{\omega}=\boldsymbol{\omega}_0} = \begin{pmatrix} \frac{1+\widehat{\lambda}^2}{\widehat{\sigma}}\left(\frac{\widehat{u}_t}{\widehat{\sigma}} - \frac{\widehat{\lambda}}{2\sqrt{1+\widehat{\lambda}^2}}\widehat{c}_t\right)x_t^{\top} \\ \frac{(1+\widehat{\lambda}^2)\widehat{u}_t^2}{2\widehat{\sigma}^4} - \frac{\widehat{\lambda}\sqrt{1+\widehat{\lambda}^2}}{4\widehat{\sigma}^3}\widehat{c}_t\widehat{u}_t \\ -\frac{\widehat{\lambda}}{\widehat{\sigma}^2}\widehat{u}_t^2 + \frac{2\widehat{\lambda}^2+1}{\sqrt{1+\widehat{\lambda}^2}}\frac{\widehat{c}_t}{2\widehat{\sigma}}\widehat{u}_t \end{pmatrix},
\tag{15}
$$

where $\widehat{u}_t = y_t - x_t^{\top}\widehat{\boldsymbol{\beta}}$, and $\widehat{h}_t$ is as in data perturbation.

3.3.4. Perturbation of Skewness

Considering the particular skewed feature of the distribution, we may investigate the effect on the model fit by making a minor change of the skewness parameter λ. In our perturbed model, we propose to replace λ by $\sqrt{\omega_i}\lambda$. Let $\boldsymbol{\omega} = (\omega_{p+1}, \ldots, \omega_T)^{\top}$ denote the $(T-p) \times 1$ perturbation vector and $\boldsymbol{\omega}_0 = (1, \ldots, 1)^{\top}$ denote the $(T-p) \times 1$ non-perturbation vector. Then, the complete-data log-likelihood function of the perturbed model is given by

$$
\ell_c(\boldsymbol{\theta}, \boldsymbol{\omega}, \boldsymbol{y}_c) = \sum_{t=p+1}^{T}\left(-\frac{1}{2}\log(\sigma^2) + \frac{1}{2}\log(1+\omega_t\lambda^2) - \frac{1+\omega_t\lambda^2}{2\sigma^2}\left(u_t - \frac{\sqrt{\omega_t}\lambda\sigma}{\sqrt{1+\omega_t\lambda^2}}h_t\right)^2\right).
$$

Thus, the perturbed Q function is expressed as

$$
\begin{aligned}
Q(\boldsymbol{\theta}, \boldsymbol{\omega})|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}} &= -\frac{(T-p)}{2}\log(\sigma^2) + \frac{(T-p)}{2}\log(1+\omega_t\lambda^2) \\
&\quad - \sum_{t=p+1}^{\top}\frac{1+\omega_t\lambda^2}{2}\left(\frac{u_t}{\sigma} - \frac{\sqrt{\omega_t}\lambda}{\sqrt{1+\omega_t\lambda^2}}\widehat{c}_t\right)^2 - \sum_{t=p+1}^{\top}\frac{\omega_t\lambda^2}{2}(\widehat{c_t^2} - (\widehat{c}_t)^2).
\end{aligned}
$$

For the SN AR(p) model with the perturbation of the skewness parameter, we have

$$
\boldsymbol{\Delta} = \frac{\partial^2 Q(\boldsymbol{\theta}, \boldsymbol{\omega})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\omega}}\bigg|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}},\boldsymbol{\omega}=\boldsymbol{\omega}_0} = \begin{pmatrix} \left(\frac{\widehat{u}_t}{\widehat{\sigma}^2}\widehat{\lambda}^2 - \frac{\widehat{\lambda}(2\widehat{\lambda}^2+1)}{2\sqrt{1+\widehat{\lambda}^2}}\frac{\widehat{c}_t}{\widehat{\sigma}}\right)x_t^{\top} \\ \frac{\widehat{\lambda}^2\widehat{u}_t^2}{2\widehat{\sigma}^4} - \frac{\widehat{\lambda}}{2\widehat{\sigma}^3}\frac{1+2\widehat{\lambda}^2}{2\sqrt{1+\widehat{\lambda}^2}}\widehat{c}_t\widehat{u}_t \\ \frac{\widehat{\lambda}}{(\widehat{\lambda}^2+1)^2} - \frac{\widehat{\lambda}}{\widehat{\sigma}^2}\widehat{u}_t^2 - \widehat{\lambda}\widehat{c}_t^2 + \frac{(6\widehat{\lambda}^2+1)(\widehat{\lambda}^2+1)-\widehat{\lambda}^2(2\widehat{\lambda}^2+1)}{2(\widehat{\lambda}^2+1)^{3/2}}\frac{\widehat{c}_t}{\widehat{\sigma}}\widehat{u}_t \end{pmatrix},
\tag{16}
$$

where $\widehat{u}_t = y_t - x_t^{\top}\widehat{\boldsymbol{\beta}}$ and $\widehat{h}_t$, $\widehat{c_t^2}$ are as in the data perturbation.

4. Numerical Results

4.1. Simulation Study I: Effectiveness of the Diagnostics

For our simulation, we consider an AR(1) model as specified in Section 2 given by

$$
y_t = \beta y_{t-1} + u_t, \quad u_t \sim \text{SN}(0, \sigma^2, \lambda),
$$

where we choose $\beta = 0.12$, $\sigma^2 = 0.003$ and $\lambda = 0.1$. We generate $T = 400$ observations. Now, we compare the performance of the ML estimates in the presence of five perturbed or shifted observations among three different scenarios with $\lambda = 0.1, 0.2, 0.3$.

We perturb the value y_t by $y_t{}^* = y_t + \beta y_{t-1}d$, where $t = 200, 201, 202, 203, 204$ and $d = 5, 10, \ldots, 50$ in order to guarantee the presence of outliers. We calculate the ML estimate of β under the perturbed and non-perturbed data, by fitting both data sets under the AR(1) model with $\lambda = 0.1, 0.2, 0.3$, respectively. Then, we compute the relative changes of the estimates as $|(E_{st(i)} - E_{st})/E_{st}|$, where $E_{st(i)}$ is the estimate of β under the perturbed data and E_{st} is the estimate of β under the non-perturbed data. The above computation in MATLAB (version 9.3, MathWorks, US) runs five iterations to converge in less than 60 s on a PC. Our experience indicates that the computation runs up to 10 iterations to converge in less than 120 s for even $T = 1000$. Figure 1 shows the effectiveness of the influence diagnostics.
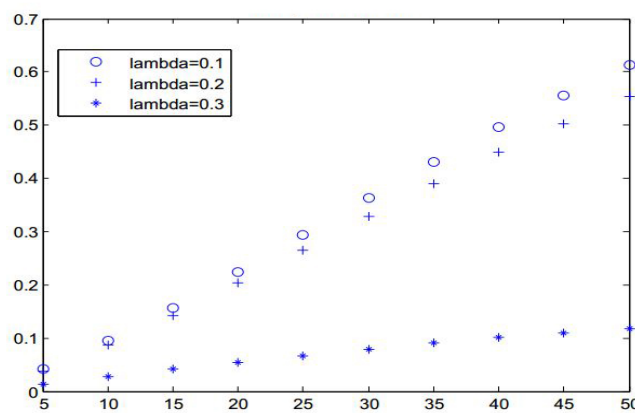
Figure 1. Relative change of estimate of β against d.

4.2. Simulation Study II: Comparing SN and Normal Distributions

We perform a numerical simulation to examine the effectiveness of our method under SN distributions. The results under the SN and normal distributions are compared as follows:

Step 1. We use the simulated data ($\lambda = 0.1$) with y_t perturbed by $y_t{}^* = y_t + \beta y_{t-1}d$, where $d = 5$ and $t = 200, 201, 202, 203, 204$. We fit an AR(1) model under normality to the data, and obtain the fitted AR(1) model given by $Y_t = 0.1549 y_{t-1} + u_t$, where $u_t \sim N(0, 0.0151)$.

Step 2. We conduct a local influence diagnostic analytics under the normal distribution using the diagnostic results given by [18].

Step 3. We compare the local influence results for the normal distribution in Step 2. In Figure 2, twenty-four influence observations are detected under the SN distribution. These results are summarized in Table 1.

As presented, twenty-four influential observations are detected by the local influence diagnostic analytics under the SN distribution, which is more than the twenty influential values detected by the local influence analytics under the normal distribution. These results are informative and as expected. We see that the influential values (that is, cases #62, #153, #201 and #301) are less than zero.

As we understand, when λ for the SN distribution is larger than zero, it is easier to find possible influential values less than zero due to the difference in patterns between the SN distribution and the normal distribution. This indicates that the diagnostic results under the SN distribution established in Section 3 work well.
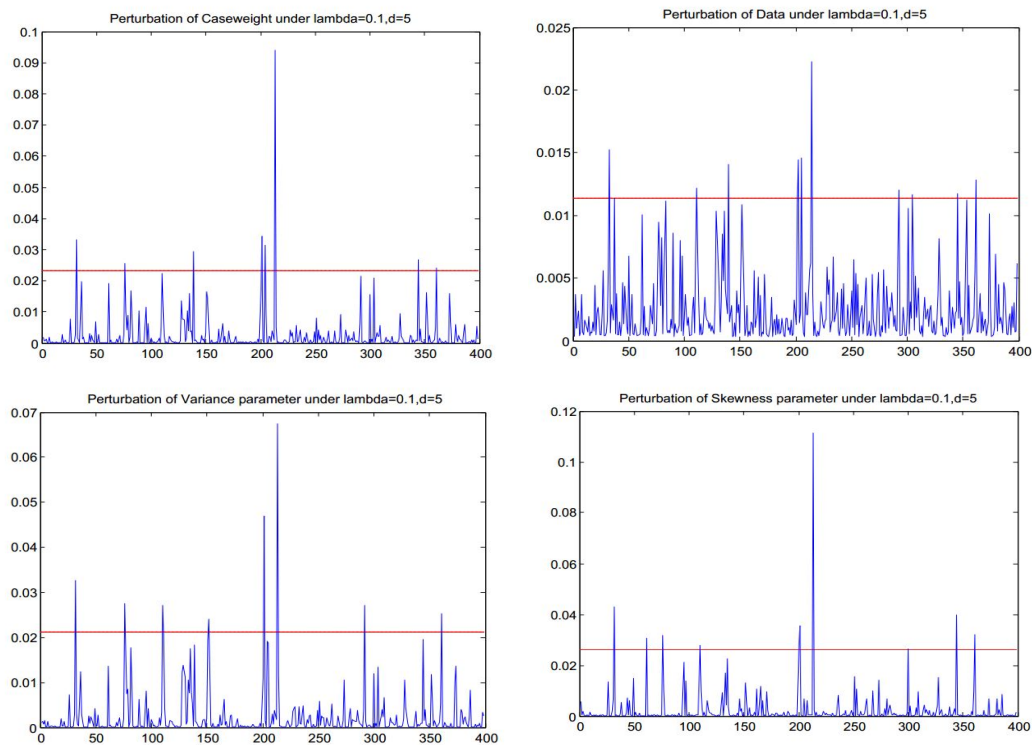
Figure 2. Diagnostics for the perturbations of case-weights, data, variance and skewness.

Table 1. Local influence results for normal and SN models

ID	Index under the Normal Model	Index under the SN Model	Observed Value
1	33	33	−0.334
2	34	34	0.030
3	-	62	−0.271
4	77	77	−0.260
5	111	111	−0.297
6	112	112	−0.225
7	140	140	0.319
8	141	141	0.089
9	-	153	−0.175
10	-	201	−0.269
11	202	202	−0.326
12	203	203	−0.090
13	205	205	0.328
14	206	206	−0.124
15	214	214	−0.406
16	215	215	−0.159
17	293	293	0.296
18	294	294	−0.012
19	-	301	−0.278
20	306	306	0.043
21	345	345	−0.293
22	346	346	−0.092
23	362	362	−0.307
24	363	363	−0.048

4.3. Real-World Data Analysis

We conduct an empirical example of financial data based on our results presented in Sections 2 and 3. We choose Chevron shares (hereinafter referred to as CVX weekly financial return data), which were

collected from 12 January 2007 to 1 August 2014 to construct the AR models. Figure 3 shows the log transformation of this data set (a total of 395 observations).
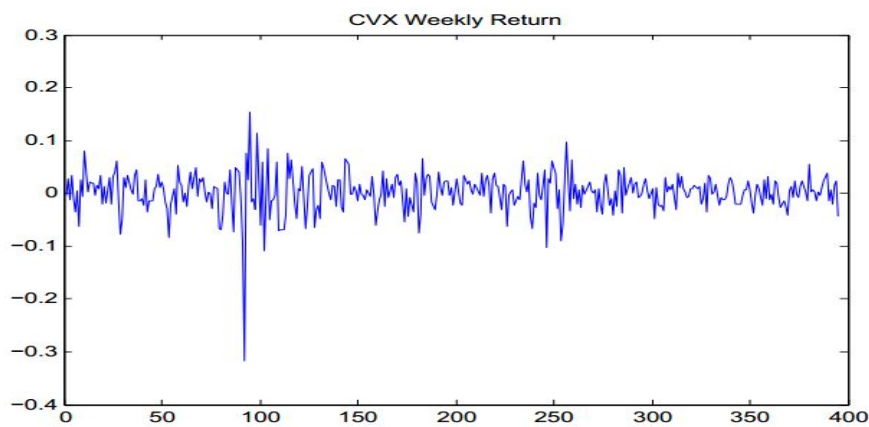
Figure 3. CVX weekly log-returns.

We first determine the order of the AR model, using the following steps:

Step 1. We assume the data are subject to an AR(p) model, for $p = 1, 2, \ldots$, given by

$$
\begin{aligned}
Y_t &= \beta_1 y_{t-1} + u_t \\
&\vdots \\
Y_t &= \beta_1 y_{t-1} + \beta_2 y_{t-2} + \ldots + \beta_p y_{t-p} + u_t \\
&\vdots
\end{aligned}
\tag{17}
$$

Step 2. For the ith equation given in (17), we let $\widehat{\beta}_j^{(i)}$ be the OLS estimate of β_j, where the superscript (i) denotes the estimates for the AR(i) model. Then, the residual is defined as $\widehat{u}_t^{(i)} = y_t - \widehat{\beta}_1^{(i)} y_{t-1} - \widehat{\beta}_2^{(i)} y_{t-2} - \cdots - \widehat{\beta}_i^{(i)} y_{t-i}$. The estimated variance for the AR(i) model is expressed as $\widehat{\sigma}_i^2 = (1/T - 2i - 1) \sum_{t=i+1}^{\top} (\widehat{u}_t^{(i)})^2$.

Step 3. We use the ith and $(i-1)$th equations given in (17) to test $H_0 : \beta_i = 0$ versus $H_1 : \beta_i \neq 0$, that is, we test the AR(i) model versus the AR($i-1$) model. The test statistic is defined as

$$
M(i) = -(T - i - 2.5) \log \left(\frac{\widehat{\sigma}_i^2}{\widehat{\sigma}_{i-1}^2} \right).
\tag{18}
$$

For our model, $M(i)$ follows an asymptotically chi-square distribution with one degree of freedom, that is, $M(i) \sim \chi^2(1)$.

We calculate $M(i)$ by (18), for $i = 1, \ldots, 7$, and present the result in Table 2. As the 99th percentile of the chi-square distribution with one degree of freedom is 6.635 ($\chi^2_{0.99}(1) = 6.635$ from Table 2), we select the order of the AR(p) model to be $p = 3$.

Table 2. Test statistic $M(i)$, for $i = 1, \ldots, 7$ using CVX data.

Order	1	2	3	4	5	6	7
$M(i)$	-0.3826	-1.001	16.9551	2.9495	-1.6675	-1.9619	8.9958

From the EM algorithm detailed in Section 2, we can find the estimate of parameter $\widehat{\theta}$ given by

$$(\widehat{\beta_1}, \widehat{\beta_2}, \widehat{\beta_3}, \widehat{\sigma^2}, \widehat{\lambda}) = (0.0153, -0.0333, -0.2198, 0.0013, 0.0661).$$

Since the absolute values of β_1, β_2 and β_3 all are less than 1, we can accept that the CVX time series is stationary for the AR(3) model. Thus, we obtain a predictive model in the fitted SN AR(3) structure given by

$$\widehat{y}_t = 0.0153 y_{t-1} - 0.0333 y_{t-2} - 0.2198 y_{t-3},$$

with $\widehat{\mu} = 0$, $\widehat{\sigma}^2 = 0.0013$, and $\widehat{\lambda} = 0.0661$.

By applying the method in Section 3, we can conduct influence analysis for the SN AR(3) model. After calculating the observed Hessian matrix and the Δ matrices for the four perturbation schemes of case-weights, data, variance parameter, and skewness parameter, we obtain the diagnostics matrices S_1, S_2, S_3, and S_4, respectively. In this case, we select the benchmark as $1/392 + 3\mathrm{SM}(0)$, with the values of 0.1144, 0.0178, 0.0347, and 0.0791 established in the simulation study for the four perturbation schemes, respectively.

In Figure 4, the straight line represents the benchmark value which determines whether an observation is potentially influential. The observation is justified to be potentially influential when its diagnostic value exceeds the benchmark.
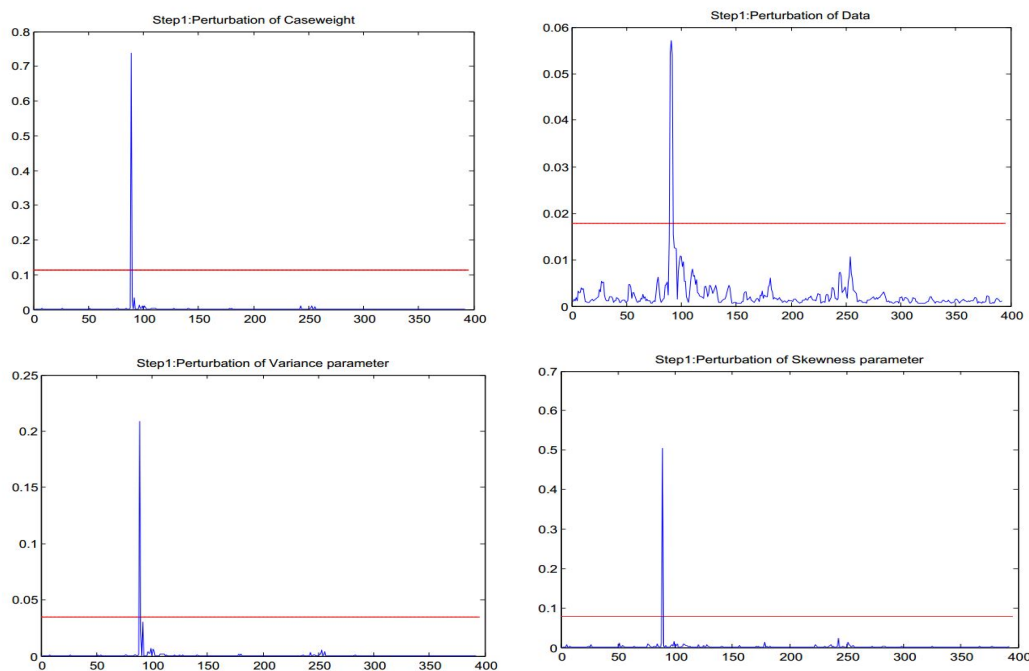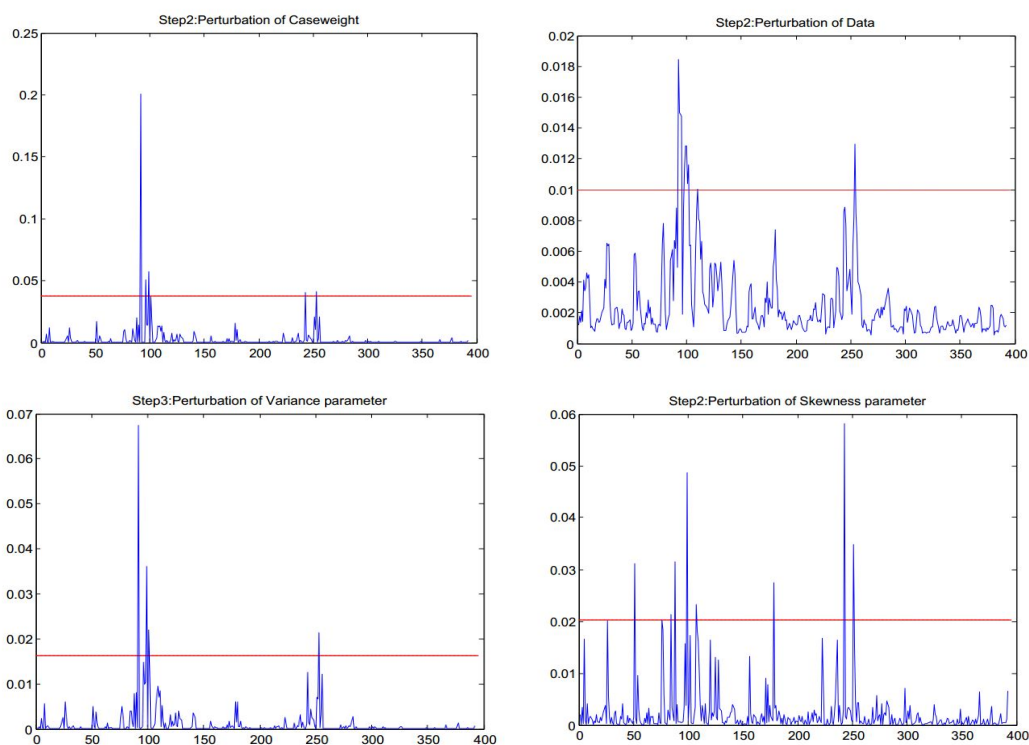

Figure 4. Diagnostics for the perturbations of case-weights, data, variance, and skewness in the SN AR(3) model [step 1].

Firstly, we only find case #92 to be potentially influential. The other potential influential observations may be masked by case #92. Similar to a step-wise diagnostic procedure proposed by [6], a second step of identification of influential observations is conducted. In the second rstep, we replace the value of case #92 by the average of cases #91 and #93 to get a new time series.

We re-fit an AR model in the same manner as done previously in the first step. For the new time series, the parameters in the SN AR model are again estimated by applying the EM-algorithm with the order also selected to be three. Thus, we present the new SN AR(3) model as

$$\widehat{y}_t = 0.0092 y_{t-1} - 0.0374 y_{t-2} - 0.1328 y_{t-3},$$

with $\widehat{\mu} = 0$, $\widehat{\sigma}^2 = 0.0011$, $\widehat{\lambda} = 0.1015$. Since the absolute values of $\widehat{\beta}_1$, $\widehat{\beta}_2$, and $\widehat{\beta}_3$ are all less than one, we can accept the CVX time series to be stationary for the AR(3) model. For the new AR(3) model, we conduct influence analysis. We select the benchmark as $1/392 + 3SM(0)$ again, with the four values of 0.0377, 0.0100, 0.0162, and 0.0203 for the corresponding perturbation schemes, respectively. Thus, twenty-two influence observations are detected in Figure 5, summarized in Table 3. It tallies the observations which are identified to be potentially influential for CVX, and an "*" in Table 3 indicates that the observation has been identified via the assigned perturbation scheme. It is worth noting that the points shown in Table 3 correspond to a number of material historical events. Many of these points relate to events around the 2008 global financial crisis. For example, on September 7 2008, the US Treasury Department announced to take over Fannie Mae and Freddie Mac. On 3 October 2008, the Bush administration signed a total of up to 700 billion dollars in a financial rescue plan. This shows the effectiveness of our procedure in identifying potentially influential observations to improve modeling outcomes.

Figure 5. Diagnostics for the perturbations of case-weights, data, variance, and skewness in the SN AR(3) model [step 2].

Table 3. Summary of the curvature-based diagnostic analytics.

ID	Time	CVXs Return	Case-Weights	Data	Variance	Skewness
54	2008-01-18	−8.286%				*
79	2008-07-11	−6.687%				*
87	2008-09-05	−7.329%				*
91	2008-10-03	−9.109%				*
92	2008-10-10	−31.674%	*	*	*	*
95	2008-10-31	15.4665%	*		*	
96	2008-11-07	−1.54%		*		
97	2008-11-14	−1.067%		*		
98	2008-11-21	−3.06%		*		
99	2008-11-28	11.4104%	*			
101	2008-12-12	5.9723%		*		
102	2008-12-19	−10.888%	*	*	*	*
103	2008-12-26	−0.708%		*		
104	2009-01-02	8.4069%			*	
105	2009-01-09	−4.956%		*		*
110	2009-02-13	−7.152%				*
113	2009-03-06	−4.102%		*		
181	2010-06-25	−7.505%				*
246	2011-09-23	−10.154%	*			*
254	2011-11-18	−8.955%				*
256	2011-12-02	9.6993%			*	
257	2011-12-09	2.4863%		*		

5. Conclusions

In this paper, we have researched the influence diagnostics in the AR(p) model under skew-normal distributions. We have obtained the normal curvatures for the model under four perturbation schemes, including the newly proposed perturbation of skewness. We have conducted a Monte Carlo simulation study to obtain approximate benchmark values for determining those possible influential points, and use them to analyze the weekly log-returns of Chevron. The findings outlined in this paper suggest that our local influence approach in the AR(p) model effectively identifies potentially influential observations to improve the fit of the model.

Note that we have made a new application of the matrix differential calculus developed by [31] in mathematical optimization and data analysis. In particular, this paper focuses on the detection of local (rather than global) influential observations. The extension of the current study to identify global influential observations will be considered in the future.

Author Contributions: Data curation, Y.L.; formal analysis, G.M. and S.L.; investigation, G.M., V.L. and A.T.; methodology, V.L., S.L. and A.T.; writing–original draft, V.L.; writing–review and editing, S.L. and A.T. All authors have read and agreed to the published version of the manuscript.

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

Appendix A

Based on the matrix differential calculus presented by [31], we establish the derivatives involved in our calculations.

Appendix A.1. The Hessian Matrix

For the function $Q(\boldsymbol{\theta})|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}} = \mathrm{E}[\ell_c(\boldsymbol{\theta},\boldsymbol{Y}_c)|\boldsymbol{Y}_o=\boldsymbol{y}_o]|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}}$, we have

$$
\begin{aligned}
Q(\boldsymbol{\theta}) &= -\frac{(T-p)}{2}\log(\sigma^2) + \frac{(T-p)}{2}\log(1+\lambda^2) - \frac{1+\lambda^2}{2}\sum_{t=p+1}^{T}\left(\left(\frac{u_t}{\sigma}\right)^2 + \frac{\lambda^2}{1+\lambda^2}\widehat{c_t^2} - \frac{u_t}{\sigma}\frac{2\lambda}{\sqrt{1+\lambda^2}}\widehat{c}_t\right) \\
&= -\frac{(T-p)}{2}\log(\sigma^2) + \frac{(T-p)}{2}\log(1+\lambda^2) - \frac{1+\lambda^2}{2}\sum_{t=p+1}^{T}\left(\frac{u_t}{\sigma} - \frac{\lambda}{\sqrt{1+\lambda^2}}\widehat{c}_t\right)^2 - \frac{\lambda^2}{2}\sum_{t=p+1}^{T}(\widehat{c_t^2} - (\widehat{c}_t)^2),
\end{aligned}
$$

where

$$
\begin{aligned}
\widehat{c}_t &= \mathrm{E}(H_t|Y_0,\widehat{\boldsymbol{\theta}}) = \tau_1 + \frac{\phi(\frac{\tau_1}{\tau_2})}{\Phi(\frac{\tau_1}{\tau_2})}\tau_2,\quad \widehat{c_t^2} = \mathrm{E}(H_t^2|Y_0,\widehat{\boldsymbol{\theta}}) = \tau_1^2 + \tau_2^2 + \frac{\phi(\frac{\tau_1}{\tau_2})}{\Phi(\frac{\tau_1}{\tau_2})}\tau_1\tau_2, \\
\tau_1 &= \frac{\widehat{\lambda}}{\widehat{\sigma}\sqrt{1+\widehat{\lambda}^2}}\widehat{u}_t,\quad \tau_2 = \frac{1}{\sqrt{1+\widehat{\lambda}^2}}.
\end{aligned}
$$

The first-order derivatives related to $\ddot{Q}(\widehat{\boldsymbol{\theta}})$ in (7) are given by

$$
\begin{aligned}
\frac{\partial Q(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}} &= \sum_{t=p+1}^{T}\frac{1+\lambda^2}{\sigma}\left(\frac{u_t}{\sigma} - \frac{\lambda}{\sqrt{1+\lambda^2}}\widehat{c}_t\right)x_t^\top, \\
\frac{\partial Q(\boldsymbol{\theta})}{\partial \sigma^2} &= -\frac{(T-p)}{2\sigma^2} + \sum_{t=p+1}^{T}\frac{1+\lambda^2}{2\sigma^3}\left(\frac{u_t}{\sigma} - \frac{\lambda}{\sqrt{1+\lambda^2}}\widehat{c}_t\right)u_t, \\
\frac{\partial Q(\boldsymbol{\theta})}{\partial \lambda} &= \sum_{t=p+1}^{T}\left(\frac{\lambda}{1+\lambda^2} - \left(\frac{u_t}{\sigma}\right)^2\lambda - \lambda\widehat{c_t^2} + \frac{2\lambda^2+1}{\sqrt{1+\lambda^2}}\frac{u_t}{\sigma}\widehat{c}_t\right).
\end{aligned}
$$

The second-order derivatives related to $\ddot{Q}(\widehat{\boldsymbol{\theta}})$ in (7) are given by

$$
\begin{aligned}
\frac{\partial^2 Q(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}\partial \boldsymbol{\beta}^\top} &= -\sum_{t=p+1}^{T}\frac{1+\lambda^2}{\sigma^2}x_t^\top x_t, \\
\frac{\partial^2 Q(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}\partial \sigma^2} &= \sum_{t=p+1}^{T}\frac{1+\lambda^2}{2\sigma^3}\left(\frac{\lambda}{\sqrt{1+\lambda^2}}\widehat{c}_t - \frac{2u_t}{\sigma}\right)x_t^\top, \\
\frac{\partial^2 Q(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}\partial \lambda} &= \sum_{t=p+1}^{T}\left(\frac{2\lambda}{\sigma^2}u_t - \frac{2\lambda^2+1}{\sigma\sqrt{1+\lambda^2}}\widehat{c}_t\right)x_t^\top, \\
\frac{\partial^2 Q(\boldsymbol{\theta})}{\partial (\sigma^2)^2} &= \frac{(T-p)}{2\sigma^4} - \sum_{t=p+1}^{T}\frac{1+\lambda^2}{4\sigma^5}\left(\frac{4u_t}{\sigma} - \frac{3\lambda}{\sqrt{1+\lambda^2}}\widehat{c}_t\right)u_t, \\
\frac{\partial^2 Q(\boldsymbol{\theta})}{\partial \sigma^2\partial \lambda} &= \sum_{t=p+1}^{T}\left(\frac{\lambda}{\sigma^4}u_t^2 - \frac{2\lambda^2+1}{\sqrt{1+\lambda^2}}\frac{u_t}{2\sigma^3}\widehat{c}_t\right), \\
\frac{\partial^2 Q(\boldsymbol{\theta})}{\partial \lambda^2} &= \frac{(T-p)(1-\lambda^2)}{(1+\lambda^2)^2} + \sum_{t=p+1}^{T}\left(-\frac{u_t^2}{\sigma^2} - \widehat{c_t^2} + \frac{2\lambda^3+3\lambda}{\sqrt{(1+\lambda^2)^3}}\frac{u_t}{\sigma}\widehat{c}_t\right).
\end{aligned}
$$

Appendix A.2. Perturbation Matrix for Case-Weights

For the function $Q(\boldsymbol{\theta}, \boldsymbol{\omega})|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} = \mathrm{E}[\ell_c(\boldsymbol{\theta}, \boldsymbol{\omega}, \boldsymbol{Y}_c)|\boldsymbol{Y}_o = \boldsymbol{y}_o]|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}$ given in (10), we have

$$
\begin{aligned}
Q(\boldsymbol{\theta}, \boldsymbol{\omega}) & = -\frac{(T-p)}{2}\log(\sigma^2) + \frac{(T-p)}{2}\log(1+\lambda^2) \\
& \quad - \frac{1+\lambda^2}{2}\sum_{t=p+1}^{\top}\left(\left(\frac{\omega_t u_t}{\sigma}\right)^2 + \frac{\lambda^2}{1+\lambda^2}\widehat{c_t^2} - \frac{\omega_t u_t}{\sigma}\frac{2\lambda}{\sqrt{1+\lambda^2}}\widehat{c}_t\right) \\
& = -\frac{(T-p)}{2}\log(\sigma^2) + \frac{(T-p)}{2}\log(1+\lambda^2) \\
& \quad - \frac{1+\lambda^2}{2}\sum_{t=p+1}^{\top}\left(\frac{\omega_t u_t}{\sigma} - \frac{\lambda}{\sqrt{1+\lambda^2}}\widehat{c}_t\right)^2 - \frac{\lambda^2}{2}\sum_{t=p+1}^{\top}(\widehat{c_t^2} - (\widehat{c}_t)^2).
\end{aligned}
$$

The first-order derivatives related to $\boldsymbol{\Delta}$ in (11) are given by

$$
\begin{aligned}
\frac{\partial Q(\boldsymbol{\theta}, \boldsymbol{\omega})}{\partial \boldsymbol{\beta}} & = \sum_{t=p+1}^{\top}\frac{1+\lambda^2}{\sigma}\left(\frac{\omega_t u_t}{\sigma} - \frac{\lambda}{\sqrt{1+\lambda^2}}\widehat{c}_t\right)\boldsymbol{x}_t^{\top}, \\
\frac{\partial Q(\boldsymbol{\theta}, \boldsymbol{\omega})}{\partial \sigma^2} & = -\frac{(T-p)}{2\sigma^2} + \sum_{t=p+1}^{\top}\frac{1+\lambda^2}{2\sigma^3}\left(\frac{\omega_t u_t}{\sigma} - \frac{\lambda}{\sqrt{1+\lambda^2}}\widehat{c}_t\right)\omega_t u_t, \\
\frac{\partial Q(\boldsymbol{\theta}, \boldsymbol{\omega})}{\partial \lambda} & = \sum_{t=p+1}^{\top}\left(\frac{\lambda}{1+\lambda^2} - \left(\frac{\omega_t u_t}{\sigma}\right)^2\lambda - \lambda\widehat{c_t^2} + \frac{2\lambda^2+1}{\sqrt{1+\lambda^2}}\frac{\omega_t u_t}{\sigma}\widehat{c}_t\right),
\end{aligned}
$$

The second-order derivatives related to $\boldsymbol{\Delta}$ in (11) are given by

$$
\begin{aligned}
\frac{\partial^2 Q(\boldsymbol{\theta}, \boldsymbol{\omega})}{\partial \boldsymbol{\beta} \partial \omega_t} & = \frac{1+\lambda^2}{\sigma^2}u_t \boldsymbol{x}_t^{\top}, \\
\frac{\partial^2 Q(\boldsymbol{\theta}, \boldsymbol{\omega})}{\partial \sigma^2 \partial \omega_t} & = \frac{1+\lambda^2}{2\sigma^3}\left(\frac{2u_t^2}{\sigma}\omega_t - \frac{\lambda\widehat{c}_t}{\sqrt{1+\lambda^2}}u_t\right), \\
\frac{\partial^2 Q(\boldsymbol{\theta}, \boldsymbol{\omega})}{\partial \lambda \partial \omega_t} & = -\frac{2\lambda}{\sigma}u_t^2\omega_t + \frac{2\lambda^2+1}{\sqrt{1+\lambda^2}}\frac{u_t}{\sigma}\widehat{c}_t.
\end{aligned}
$$

Noting that $\omega_0 = (1, \dots, 1)^{\top}$, we obtain (11).

Appendix A.3. Perturbation Matrix for Data

For the function $Q(\boldsymbol{\theta}, \boldsymbol{\omega})|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} = \mathrm{E}[\ell_c(\boldsymbol{\theta}, \boldsymbol{\omega}, \boldsymbol{Y}_c)|\boldsymbol{Y}_o = \boldsymbol{y}_o]|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}$ given in (12), we have

$$
\begin{aligned}
Q(\boldsymbol{\theta}, \boldsymbol{\omega}) & = -\frac{(T-p)}{2}\log(\sigma^2) + \frac{(T-p)}{2}\log(1+\lambda^2) \\
& \quad - \frac{1+\lambda^2}{2}\sum_{t=p+1}^{\top}\left(\left(\frac{u_t + \mu\left(\omega_t\right)}{\sigma}\right)^2 + \frac{\lambda^2}{1+\lambda^2}\widehat{c_t^2} - \frac{u_t + \mu\left(\omega_t\right)}{\sigma}\frac{2\lambda}{\sqrt{1+\lambda^2}}\widehat{c}_t\right) \\
& = -\frac{(T-p)}{2}\log(\sigma^2) + \frac{(T-p)}{2}\log(1+\lambda^2) \\
& \quad - \frac{1+\lambda^2}{2}\sum_{t=p+1}^{\top}\left(\frac{u_t + \mu\left(\omega_t\right)}{\sigma} - \frac{\lambda}{\sqrt{1+\lambda^2}}\widehat{c}_t\right)^2 - \frac{\lambda^2}{2}\sum_{t=p+1}^{\top}(\widehat{c_t^2} - (\widehat{c}_t)^2).
\end{aligned}
$$

The first-order derivatives related to Δ in (13) are given by

$$
\frac{\partial Q(\boldsymbol{\theta}, \boldsymbol{\omega})}{\partial \boldsymbol{\beta}} = \sum_{t=p+1}^{\top} \frac{1+\lambda^2}{\sigma} \left(\frac{u_t + \mu(\omega_t)}{\sigma} - \frac{\lambda}{\sqrt{1+\lambda^2}} \widehat{c}_t \right) \left(\boldsymbol{x}_t^{\top} + (\omega_{t-1}, \ldots, \omega_{t-p})^{\top} \right)
$$

$$
\frac{\partial Q(\boldsymbol{\theta}, \boldsymbol{\omega})}{\partial \sigma^2} = -\frac{(T-p)}{2\sigma^2} + \sum_{t=p+1}^{\top} \frac{1+\lambda^2}{2\sigma^3} \left(\frac{(u_t + \mu(\omega_t))^2}{\sigma} - \frac{\lambda(u_t + \mu(\omega_t))}{\sqrt{1+\lambda^2}} \widehat{c}_t \right)
$$

$$
\frac{\partial Q(\boldsymbol{\theta}, \boldsymbol{\omega})}{\partial \lambda} = \sum_{t=p+1}^{\top} \left(\frac{\lambda}{1+\lambda^2} - \left(\frac{u_t + \mu(\omega_t)}{\sigma} \right)^2 \lambda - \lambda \widehat{c}_t^2 + \frac{2\lambda^2+1}{\sqrt{1+\lambda^2}} \frac{u_t + \mu(\omega_t)}{\sigma} \widehat{c}_t \right)
$$

The second-order derivatives related to Δ in (13) are given by

$$
\frac{\partial^2 Q(\boldsymbol{\theta}, \boldsymbol{\omega})}{\partial \boldsymbol{\beta} \partial \omega_t} = \frac{1+\lambda^2}{\sigma^2} \left(\boldsymbol{x}_t^{\top} + (\omega_{t-1}, \ldots, \omega_{t-p})^{\top} \right)
$$

$$
\frac{\partial^2 Q(\boldsymbol{\theta}, \boldsymbol{\omega})}{\partial \sigma^2 \partial \omega_t} = \frac{1+\lambda^2}{2\sigma^3} \left(\frac{2(u_t + \mu(\omega_t))}{\sigma} - \frac{\lambda \widehat{c}_t}{\sqrt{1+\lambda^2}} \right)
$$

$$
\frac{\partial^2 Q(\boldsymbol{\theta}, \boldsymbol{\omega})}{\partial \lambda \partial \omega_t} = \left(-\frac{2\lambda}{\sigma^2} (u_t + \mu(\omega_t)) + \frac{2\lambda^2+1}{\sqrt{1+\lambda^2}} \frac{\widehat{c}_t}{\sigma} \right)
$$

Noting that $\omega_0 = (0, \ldots, 0)^{\top}$, we obtain (13).

Appendix A.4. Perturbation Matrix for Scale

For the function $Q(\boldsymbol{\theta}, \boldsymbol{\omega})|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}} = \mathrm{E}[\ell_c(\boldsymbol{\theta}, \boldsymbol{\omega}, \boldsymbol{Y}_c)|\boldsymbol{Y}_o = \boldsymbol{y}_o]|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}}$ given in (14), we have

$$
\begin{aligned}
Q(\boldsymbol{\theta}, \boldsymbol{\omega}) &= -\frac{(T-p)}{2} \log(\sigma^2) + \frac{(T-p)}{2} \log(\omega_t) + \frac{(T-p)}{2} \log(1+\lambda^2) \\
&\quad - \frac{1+\lambda^2}{2} \sum_{t=p+1}^{\top} \left(\frac{\omega_t u_t^2}{\sigma^2} + \frac{\lambda^2}{1+\lambda^2} \widehat{c}_t^2 - \frac{\sqrt{\omega_t} u_t}{\sigma} \frac{2\lambda}{\sqrt{1+\lambda^2}} \widehat{c}_t \right) \\
&= -\frac{(T-p)}{2} \log(\sigma^2) + \sum_{t=p+1}^{\top} \frac{1}{2} \log(\omega_t) + \frac{(T-p)}{2} \log(1+\lambda^2) \\
&\quad - \frac{1+\lambda^2}{2} \sum_{t=p+1}^{\top} \left(\frac{\sqrt{\omega_t} u_t}{\sigma} - \frac{\lambda}{\sqrt{1+\lambda^2}} \widehat{c}_t \right)^2 - \frac{\lambda^2}{2} \sum_{t=p+1}^{T} (\widehat{c}_t^2 - (\widehat{c}_t)^2).
\end{aligned}
$$

The first-order derivatives related to Δ in (15) are given by

$$
\frac{\partial Q(\boldsymbol{\theta}, \boldsymbol{\omega})}{\partial \boldsymbol{\beta}} = \sum_{t=p+1}^{\top} \left(\frac{(1+\lambda^2)\omega_t u_t}{\sigma^2} - \frac{\lambda\sqrt{\omega_t(1+\lambda^2)}}{\sigma} \widehat{c}_t \right) \boldsymbol{x}_t^{\top}
$$

$$
\frac{\partial Q(\boldsymbol{\theta}, \boldsymbol{\omega})}{\partial \sigma^2} = \sum_{t=p+1}^{\top} \left(-\frac{1}{2\sigma^2} + \frac{(1+\lambda^2)\omega_t u_t^2}{2\sigma^4} - \frac{\lambda\sqrt{\omega_t(1+\lambda^2)}}{2\sigma^3} \widehat{c}_t u_t \right)
$$

$$
\frac{\partial Q(\boldsymbol{\theta}, \boldsymbol{\omega})}{\partial \lambda} = \sum_{t=p+1}^{\top} \left(\frac{\lambda}{1+\lambda^2} - \frac{\omega_t u_t^2}{\sigma^2} \lambda - \lambda \widehat{c}_t^2 + \frac{2\lambda^2+1}{\sqrt{1+\lambda^2}} \frac{\sqrt{\omega_t} u_t}{\sigma} \widehat{c}_t \right)
$$

The second-order derivatives related to $\mathbf{\Delta}$ in (15) are given by

$$
\frac{\partial^2 Q(\boldsymbol{\theta}, \boldsymbol{\omega})}{\partial \boldsymbol{\beta} \partial \omega_t} = \left(\frac{(1+\lambda^2)}{\sigma^2} u_t - \frac{\lambda \sqrt{1+\lambda^2}}{\sigma} \frac{1}{2\sqrt{\omega_t}} \widehat{c}_t \right) {x_t}^\top
$$

$$
\frac{\partial^2 Q(\boldsymbol{\theta}, \boldsymbol{\omega})}{\partial \sigma^2 \partial \omega_t} = \frac{(1+\lambda^2) u_t^2}{2\sigma^4} - \frac{\lambda \sqrt{1+\lambda^2}}{4\sigma^3 \sqrt{\omega_t}} \widehat{c}_t u_t
$$

$$
\frac{\partial^2 Q(\boldsymbol{\theta}, \boldsymbol{\omega})}{\partial \lambda \partial \omega_t} = -\frac{u_t^2 \lambda}{\sigma^2} + \frac{2\lambda^2+1}{\sqrt{1+\lambda^2}} \frac{u_t}{2\sqrt{\omega_t}} \frac{\widehat{c}_t}{\sigma}
$$

Noting that $\omega_0 = (1, \ldots, 1)^\top$, we obtain (15).

Appendix A.5. Perturbation Matrix for Skewness

For the function $Q(\boldsymbol{\theta}, \boldsymbol{\omega})|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}} = \mathrm{E}[\ell_c(\boldsymbol{\theta}, \boldsymbol{\omega}, \boldsymbol{Y}_c)|\boldsymbol{Y}_o = \boldsymbol{y}_o]|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}}$ given in (16), we have

$$
\begin{aligned}
Q(\boldsymbol{\theta}, \boldsymbol{\omega}) &= -\frac{(T-p)}{2} \log(\sigma^2) + \frac{(T-p)}{2} \log(1+\omega_t \lambda^2) \\
&\quad - \frac{1+\omega_t \lambda^2}{2} \sum_{t=p+1}^{\top} \left(\frac{u_t^2}{\sigma^2} + \frac{\omega_t \lambda^2}{1+\omega_t \lambda^2} \widehat{c}_t^2 - \frac{u_t}{\sigma} \frac{2\sqrt{\omega_t}\lambda}{\sqrt{1+\omega_t \lambda^2}} \widehat{c}_t \right) \\
&= -\frac{(T-p)}{2} \log(\sigma^2) + \frac{(T-p)}{2} \log(1+\omega_t \lambda^2) \\
&\quad - \frac{1+\omega_t \lambda^2}{2} \sum_{t=p+1}^{\top} \left(\frac{u_t}{\sigma} - \frac{\sqrt{\omega_t}\lambda}{\sqrt{1+\omega_t \lambda^2}} \widehat{c}_t \right)^2 - \frac{\omega_t \lambda^2}{2} \sum_{t=p+1}^{\top} (\widehat{c_t^2} - (\widehat{c}_t)^2).
\end{aligned}
$$

The first-order derivatives related to $\mathbf{\Delta}$ in (16) are given by

$$
\frac{\partial Q(\boldsymbol{\theta}, \boldsymbol{\omega})}{\partial \boldsymbol{\beta}} = \sum_{t=p+1}^{\top} \frac{1+\omega_t \lambda^2}{\sigma} \left(\frac{u_t}{\sigma} - \frac{\lambda \sqrt{\omega_t}}{\sqrt{1+\omega_t \lambda^2}} \widehat{c}_t \right) {x_t}^\top
$$

$$
\frac{\partial Q(\boldsymbol{\theta}, \boldsymbol{\omega})}{\partial \sigma^2} = \sum_{t=p+1}^{\top} \left(-\frac{1}{2\sigma^2} + \frac{1+\omega_t \lambda^2}{2\sigma^3} \left(\frac{u_t}{\sigma} - \frac{\lambda \sqrt{\omega_t}}{\sqrt{1+\omega_t \lambda^2}} \widehat{c}_t \right) u_t \right)
$$

$$
\frac{\partial Q(\boldsymbol{\theta}, \boldsymbol{\omega})}{\partial \lambda} = \sum_{t=p+1}^{\top} \left(\frac{\omega_t \lambda}{1+\omega_t \lambda^2} - \frac{u_t^2}{\sigma^2} \omega_t \lambda - \omega_t \lambda \widehat{c}_t^2 + \frac{\sqrt{\omega_t} \left(2\omega_t \lambda^2 + 1\right)}{\sqrt{1+\omega_t \lambda^2}} \frac{u_t}{\sigma} \widehat{c}_t \right).
$$

The second-order derivatives related to $\mathbf{\Delta}$ in (16) are given by

$$
\frac{\partial^2 Q(\boldsymbol{\theta}, \boldsymbol{\omega})}{\partial \boldsymbol{\beta} \partial \omega_t} = \left(\frac{\lambda^2}{\sigma^2} u_t - \frac{\lambda \left(1 + 2\lambda^2 \omega_t\right)}{2\sqrt{\omega_t \left(1+\lambda^2 \omega_t\right)}} \frac{\widehat{c}_t}{\sigma} \right) {x_t}^\top
$$

$$
\frac{\partial^2 Q(\boldsymbol{\theta}, \boldsymbol{\omega})}{\partial \sigma^2 \partial \omega_t} = \left(\frac{\lambda^2 u_t}{2\sigma^4} - \frac{\lambda}{4\sigma^3} \frac{1 + 2\lambda^2 \omega_t}{\sqrt{\omega_t \left(1+\lambda^2 \omega_t\right)}} \widehat{c}_t \right) u_t
$$

$$
\begin{aligned}
\frac{\partial^2 Q(\boldsymbol{\theta}, \boldsymbol{\omega})}{\partial \lambda \partial \omega_t} &= \frac{\lambda}{\left(1+\lambda^2 \omega_t\right)^2} - \frac{u_t^2 \lambda}{\sigma^2} - \lambda \widehat{c}_t^2 \\
&\quad + \frac{\left(6\lambda^2 \omega_t + 1\right) \left(\lambda^2 \omega_t + 1\right) - \lambda^2 \omega_t \left(2\lambda^2 \omega_t + 1\right)}{2\sqrt{\omega_t}\left(1+\lambda^2 \omega_t\right)^{3/2}} \frac{u_t}{\sigma} \widehat{c}_t
\end{aligned}
$$

Noting that $\omega_0 = (1, \ldots, 1)^\top$, we obtain (16).

References

1. Hyndman, R.J.; Athanasopoulos, G. *Forecasting: Principles and Practice*; OTexts: Melbourne, Australia, 2018.
2. Liu, T.; Liu, S.; Shi, L. *Time Series Analysis Using SAS Enterprise Guide*; Springer: Singapore, 2020.
3. Li, W.K. *Diagnostic Checks in Time Series*; Chapman and Hall/CRC: Boca Raton, FL, USA, 2004.
4. Liu, S.; Welsh, A.H. Regression diagnostics. In *International Encyclopedia of Statistical Science*; Lovric, M., Ed.; Springer: Berlin/Heidelberg, Germany, 2011; pp. 1206–1208.
5. Cook, D. Assessment of local influence. *J. R. Stat. Soc. B* **1986**, *48*, 133–169. [CrossRef]
6. Shi, L.; Huang, M. Stepwise local influence analysis. *Comput. Stat. Data Anal.* **2011**, *55*, 973–982. [CrossRef]
7. Lu, J.; Shi, L.; Chen, F. Outlier detection in time series models using local influence method. *Comm. Stat. Theor. Meth.* **2012**, *41*, 2202–2220. [CrossRef]
8. Cao, C.Z.; Lin, J.G.; Shi, J.Q. Diagnostics on nonlinear model with scale mixtures of skew-normal and first-order autoregressive errors. *Statistics* **2014**, *48*, 1033–1047. [CrossRef]
9. Leiva, V.; Liu, S.; Shi, L.; Cysneiros, F.J.A. Diagnostics in elliptical regression models with stochastic restrictions applied to econometrics. *J. Appl. Stat.* **2016**, *43*, 627–642. [CrossRef]
10. Liu, S.; Leiva, V.; Ma, T.; Welsh, A.H. Influence diagnostic analysis in the possibly heteroskedastic linear model with exact restrictions. *Stat. Meth. Appl.* **2016**, *25*, 227–249. [CrossRef]
11. Liu, S.; Ma, T.; SenGupta, A.; Shimizu, K.; Wang, M.Z. Influence diagnostics in possibly asymmetric circular-linear multivariate regression models. *Sankhya B* **2017**, *79*, 76–93. [CrossRef]
12. Garcia-Papani, F.; Leiva, V.; Ruggeri, F.; Uribe-Opazo, M.A. Kriging with external drift in a Birnbaum-Saunders geostatistical model. *Stoch. Environ. Res. Risk Assess.* **2019**, *32*, 1517–1530. [CrossRef]
13. Leao, J.; Leiva, V.; Saulo, H.; Tomazella, V. Incorporation of frailties into a cure rate regression model and its diagnostics and application to melanoma data. *Stat. Med.* **2018**, *37*, 4421–4440. [CrossRef]
14. Tapia, A.; Leiva, V.; Diaz, M.P.; Giampaoli, V. Influence diagnostics in mixed effects logistic regression models. *TEST* **2019**, *28*, 920–942. [CrossRef]
15. Tapia, H.; Giampaoli, V.; Diaz, M.P.; Leiva, V. Sensitivity analysis of longitudinal count responses: A local influence approach and application to medical data. *J. Appl. Stat.* **2019**, *46*, 1021–1042. [CrossRef]
16. Liu, S. On diagnostics in conditionally heteroskedastic time series models under elliptical distributions. *J. Appl. Prob.* **2004**, *41*, 393–405. [CrossRef]
17. Zevallos, M.; Santos, B.; Hotta, L.K. A note on influence diagnostics in AR(1) time series models. *J. Stat. Plan. Inference* **2012**, *142*, 2999–3007. [CrossRef]
18. Liu, Y.; Ji, G.; Liu, S. Influence diagnostics in a vector autoregressive model. *J. Stat. Comput. Simul.* **2015**, *85*, 2632–2655. [CrossRef]
19. Azzalini, A. A class of distribution which includes the normal ones. *Scand. J. Stat.* **1985**, *12*, 171–178.
20. Eling, M. Fitting insurance claims to skewed distributions: Are the skew-normal and skew-student good models? *Insur. Math. Econ.* **2012**, *51*, 239–248. [CrossRef]
21. Azzalini, A. *The Skew-Normal and Related Families*; Cambridge University Press: Cambridge, UK, 2014.
22. Arnold, B.C.; Gomez, H.W.; Varela, H.; Vidal, I. Univariate and bivariate models related to the generalized epsilon-skew-Cauchy distribution. *Symmetry* **2019**, *11*, 794. [CrossRef]
23. Cysneiros, F.J.A.; Leiva, V.; Liu, S.; Marchant, C.; Scalco, P. A Cobb-Douglas type model with stochastic restrictions: Formulation, local influence diagnostics and data analytics in economics. *Qual. Quant.* **2019**, *53*, 1693–1719. [CrossRef]
24. Huerta, M.; Leiva, V.; Liu, S.; Rodriguez, M.; Villegas, D. On a partial least squares regression model for asymmetric data with a chemical application in mining. *Chem. Intell. Lab. Syst.* **2019**, *190*, 55–68. [CrossRef]
25. Ventura, M.; Saulo, H.; Leiva, V.; Monsueto, S. Log-symmetric regression models: Information criteria, application to movie business and industry data with economic implications. *Appl. Stoch. Models Bus. Ind.* **2019**, *35*, 963–977. [CrossRef]
26. Maleki, M.; Arellano-Valle, R.V. Maximum a-posteriori estimation of autoregressive processes based on finite mixtures of scale-mixtures of skew-normal distributions. *J. Stat. Comput. Simul.* **2017**, *87*, 1061–108. [CrossRef]
27. Bolfarine, H.; Montenegro, L.C.; Lachos, V.H. Influence diagnostics for skew-normal linear mixed models. *Indian J. Stat.* **2007**, *69*, 648–670.

28.	Xie, F.C.; Lin, J.G.; Wei, B.C. Diagnostics for skew-normal nonlinear regression models with AR(1) errors. *Comput. Stat. Data Anal.* **2009**, *53*, 4403–4416. [CrossRef]

29.	Garay, A.M.; Lachos, V.H.; Labra, F.V.; Ortega, E.M.M. Statistical diagnostics for nonlinear regression models based on scale mixtures of skew-normal distributions. *J. Stat. Comput. Simul.* **2014**, *84*, 1761–1778. [CrossRef]

30.	Carmichael, B.; Coën, A. Asset pricing with skewed-normal return. *Finance Res. Lett.* **2013**, *10*, 50–57. [CrossRef]

31.	Magnus, J.R.; Neudecker, H. *Matrix Differential Calculus with Applications in Statistics and Econometrics*; Wiley: Chichester, UK, 2019.

32.	Dempster, A.P.; Laird, N.M.; Rubin, D.B. Maximum likelihood from incomplete data via the EM algorithm. *J. R. Stat. Soc. B* **1977**, *39*, 1–38.

33.	McLachlan, G.; Krishnan, T. *The EM Algorithm and Extensions*; Wiley: New York, NY, USA, 1997.

34.	Poon, W.Y.; Poon, Y.S. Conformal normal curvature and assessment of local influence. *J. R. Stat. Soc. B* **1999**, *61*, 51–61. [CrossRef]