Journal of Multivariate Analysis 100 (2009) 1761–1776

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

Estimation of autoregressive models with epsilon-skew-normal innovations

Pascal Bondon

CNRS UMR 8506, 3 rue Joliot–Curie, 91192 Gif-sur-Yvette, France

ARTICLE INFO

Article history:
Received 1 February 2008
Available online 20 February 2009

AMS 2000 subject classification:
primary 62M10
secondary 62F12

Keywords:
Non-Gaussian time series
Autoregression
Maximum likelihood estimation
Skewness
Skew-normal distribution

ABSTRACT

A non-Gaussian autoregressive model with epsilon-skew-normal innovations is introduced. Moments and maximum likelihood estimators of the parameters are proposed and their limit distributions are derived. Monte Carlo simulation results are analysed and the model is fitted to a real time series.

1. Introduction

In the modelling of non-Gaussian time series, two strategies may be adopted. We may either retain the general autoregressive moving average (ARMA) framework and allow the white noise to be non-Gaussian, or we may completely abandon the linearity assumption, see e.g. [1,2]. In the former case, the difficulty is to choose the distribution of the white noise appropriately so that the ARMA time series exhibits a specified non-Gaussian feature. In the latter case, one has to find an adequate explicit model among infinitely many nonlinear forms that typically express the time series as a nonlinear function of its lagged values.

In this paper, we are interested in correlated data exhibiting asymmetry and we follow the first strategy. The data are short-range dependent in the sense that their autocorrelations decay to zero exponentially, and their distributions are near-Gaussian. We study the problem of fitting an AR model to these data. Many non-Gaussian AR models were proposed in the literature, see e.g. [3] and references therein. In particular, Jacobs and Lewis [4] considered the construction of models for stationary sequences of discrete random variables with given first-order marginal probability mass functions and nonnegative autocorrelation structures. Anděl [5] studied AR(1) models with exponentially distributed innovations. Li and McLeod [6] addressed the problem of ARMA modelling with non-Gaussian innovations. They established general results on maximum likelihood estimates (MLE) and as real examples, they fitted ARMA models with log-normal and gamma innovations to the sunspot and the Canadian lynx data respectively, demonstrating that linear time series model with non-Gaussian innovations can be a useful tool in time series modelling. Tiku, Wong and Bian [7] and Tiku, Wong, Vaughan and Bian [8] considered the estimation of AR models with symmetric innovations that follow a shift-scaled Student's t distribution, and Tiku, Wong and Bian [9], Akkaya and Tiku [10] and Wong and Bian [11] studied AR models with asymmetric innovations distributed according to gamma and generalised logistic distributions. These authors derived modified MLE of the parameters that are easy to compute. On the other hand, Janacek and Swift [12] proposed a different approach that

E-mail address: bondon@lss.supelec.fr.

0047-259X/$ – see front matter
doi:10.1016/j.jmva.2009.02.006

consists in modelling a non-Gaussian time series as a nonlinear instantaneous transformation of a Gaussian ARMA time series, the nonlinear transformation being determined from the first-order marginal distribution of the data. Recently, Pourahmadi [13] considered the construction of stationary ARMA models with multivariate skew-normal distributions. When these distributions belong to the class defined by Azzalini and Dalla Valle [14] and Azzalini and Capitanio [15], the innovations are correlated and the predictors are nonlinear and heteroscedastic. Unfortunately, the autocorrelations of the ARMA model differ from their Gaussian ARMA counterparts in that they do not converge to zero for large lags, which is a limitation for modelling real time series. When the multivariate distributions of the ARMA model lie in the family of closed skew-normal distributions introduced by González-Farías, Domínguez-Molina and Gupta [16] and re-parametrised and generalised by Arellano-Valle and Azzalini [17], it is possible and natural to define the innovations as a sequence of iid random variables with a univariate distribution in this family. In this case, the autocorrelations of the ARMA model decay to zero exponentially and the predictors are linear and homoscedastic as in the Gaussian case. Nevertheless, as mentioned by Pourahmadi [13], the maximum likelihood estimation of the parameters of the ARMA model might be computationally intensive and the asymptotic properties of these estimates are not established. Here, we do not investigate the construction of ARMA models with given multivariate skew-normal distributions, but we consider the statistical estimation of an AR model with iid epsilon-skew-normal (ESN) innovations. The ESN distribution was introduced by Mudholkar and Hutson [18] and has been used recently in regression problems by Hutson [19]. Its main advantage is its flexibility since it is analytically tractable, it accommodates practical values of skewness and kurtosis, and it strictly includes the Gaussian distribution. As far as we know, the idea of using an AR model with ESN innovations to represent correlated asymmetric data has not been explored before, despite the flexibility of the ESN distribution.

Specifically, the ESN$(\theta, \sigma, \epsilon)$ distribution with location $\theta \in \mathbb{R}$, scale $\sigma > 0$ and skewness $\epsilon \in (-1, 1)$ is characterised by the density

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma}\left[\exp\left(-\frac{(x-\theta)^2}{2\sigma^2(1+\epsilon)^2}\right)\mathbb{1}_{(-\infty,\theta)}(x) + \exp\left(-\frac{(x-\theta)^2}{2\sigma^2(1-\epsilon)^2}\right)\mathbb{1}_{(\theta,+\infty)}(x)\right], \tag{1}$$

where, for any set S, $\mathbb{1}_S$ denotes the indicator function of S. This distribution is unimodal with mode at θ and it has probability mass $(1+\epsilon)/2$ below the mode. If Z has an ESN$(\theta, \sigma, \epsilon)$ distribution, $\mathrm{E}\,Z = \theta - 4\sigma\epsilon/\sqrt{2\pi}$ and the kth central moment of Z is

$$\begin{aligned}\mathrm{E}(Z-\mathrm{E}\,Z)^k &= \int_{\mathbb{R}}(x-\mathrm{E}\,Z)^k f(x)\mathrm{d}x = \sigma^{k+1}\int_{\mathbb{R}}\left(x+\frac{4\epsilon}{\sqrt{2\pi}}\right)^k f(\sigma x+\theta)\mathrm{d}x \\ &= \frac{\sigma^k}{\sqrt{2\pi}}\sum_{l=0}^{k}\binom{k}{l}\left(\frac{4\epsilon}{\sqrt{2\pi}}\right)^{k-l}\left((-1)^l I_l(-\epsilon)+I_l(\epsilon)\right),\end{aligned}$$

where

$$I_l(\epsilon) = \int_0^\infty x^l \exp\left(-\frac{x^2}{2(1-\epsilon)^2}\right)\mathrm{d}x = \begin{cases}\sqrt{\dfrac{\pi}{2}}(1-\epsilon)^{l+1}\displaystyle\prod_{i=1}^{m}(2i-1) & \text{if } l=2m,\\ 2^m(1-\epsilon)^{l+1}m! & \text{if } l=2m+1.\end{cases} \tag{2}$$

Therefore, $(-1)^l I_l(-\epsilon)+I_l(\epsilon)$ is a polynomial of degree l and the kth central moment of Z takes the form $\sigma^k P_k(\epsilon)$ where P_k is a polynomial of degree k. For $k > 1$, the kth cumulant $c_{k,Z}$ of Z is obtained from the lth central moments for $l \le k$ by means of well-known polynomial relations, see for instance [20, eqn (3.43)]. It follows from these relations that $c_{k,Z}$, $k > 1$, takes also the form $\sigma^k P'_k(\epsilon)$ where P'_k is a polynomial of degree k. The four firsts cumulants of Z are

$$\begin{aligned}c_{1,Z} &= \theta - 4\sigma\epsilon/\sqrt{2\pi},\\ c_{2,Z} &= \frac{\sigma^2}{\pi}[(3\pi-8)\epsilon^2+\pi],\\ c_{3,Z} &= \frac{2\sqrt{2}\sigma^3\epsilon}{\pi^{3/2}}[(5\pi-16)\epsilon^2-\pi],\\ c_{4,Z} &= \frac{4\sigma^4\epsilon^2}{\pi^2}[(-3\pi^2+40\pi-96)\epsilon^2+\pi(3\pi-8)].\end{aligned} \tag{3}$$

Since $\epsilon \in (-1, 1)$, $c_{3,Z}/c_{2,Z}^{3/2} \in (-c_0, c_0)$ where $c_0 = \sqrt{2}(4-\pi)(\pi-2)^{-3/2} = 0.995$, and $c_{4,Z}/c_{2,Z}^2 \in (0, 0.870)$. The ESN distribution is therefore useful for modelling asymmetric data with slight leptokurticity. Of course, the ESN$(\theta, \sigma, \epsilon)$ distribution reduces to the Gaussian distribution with mean θ and variance σ^2 when $\epsilon = 0$.

The paper is organised as follows. In Section 2, we propose moments estimates (ME) and conditional MLE of the parameters $(\phi_1, \ldots, \phi_p, \theta, \sigma, \epsilon)$ of the AR(p) model (X_t) defined by the difference equation

$$X_t = \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} + Z_t, \tag{4}$$

where the polynomial $\phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p$ has no zeros in the closed unit disk $\{z \in \mathbb{C} : |z| \le 1\}$, and (Z_t) is a sequence of iid random variables with an ESN$(\theta, \sigma, \epsilon)$ distribution. Strong consistency and asymptotic normality are established. In Section 3, the behaviour of the estimators for finite samples is studied via simulation, and we fit an AR model with ESN innovations to a real time series.

2. Parameter estimation

In all the following, the parameter vector $\eta = (\phi', \theta, \epsilon, \sigma^2)'$ where $\phi = (\phi_1, \dots, \phi_p)'$ is assumed to be lying in the open set $S = C \times \mathbb{R} \times (-1, 1) \times (0, \infty)$, where C is the interior of the domain of vectors ϕ such that $\phi(z)$ has no zeros in the closed unit disk. We shall denote by $\|\cdot\|$ the Euclidean norm in $\mathbb{R}^{p+3}$, so that $\|\eta\| = (\eta'\eta)^{1/2}$.

2.1. Method of moments

The difference equation (4) has the unique stationary solution

$$X_t = \sum_{i=0}^{\infty} \psi_i Z_{t-i}, \tag{5}$$

where the series converges absolutely almost surely (a.s.) and in the mean square sense, and where $(\psi_i)_{i\in\mathbb{N}}$ are the coefficients in the Taylor series expansion of $1/\phi(z)$ for $|z| \le 1$. Since $\sum |\psi_i| < \infty$, finiteness of $\mathrm{E}\,|Z_t|^k$ imply finiteness of $\mathrm{E}\,|X_t|^k$ for all $k \ge 1$, see e.g. [21, Lemma 2.7.3].

Let $m = \mathrm{E}\,X_t$, $\widetilde{X}_t = X_t - m$, $m_{2,k} = \mathrm{E}\,\widetilde{X}_t\widetilde{X}_{t+k}$, and $m_{3,k} = \mathrm{E}\,\widetilde{X}_t\widetilde{X}_{t+k}^2$. It results from (4) that

$$m(1 - e'\phi) = \mathrm{E}\,Z_t = \theta - 4\sigma\epsilon/\sqrt{2\pi}, \tag{6}$$

where $e = (1, \dots, 1)'$. The time series $(\widetilde{X}_t)$ satisfies the causal AR(p) model

$$\widetilde{X}_t = \phi_1\widetilde{X}_{t-1} + \cdots + \phi_p\widetilde{X}_{t-p} + \widetilde{Z}_t, \tag{7}$$

where $\widetilde{Z}_t = Z_t - \mathrm{E}\,Z_t$. The standard Yule–Walker equations for model (7) are

$$M_2\phi = m_2, \tag{8}$$

and

$$m_{2,0} - \phi' m_2 = \mathrm{E}\,\widetilde{Z}_t^2 = \frac{\sigma^2}{\pi}[(3\pi - 8)\epsilon^2 + \pi], \tag{9}$$

where M_2 is the invertible covariance matrix $[m_{2,i-j}]_{i,j=1}^{p}$ and $m_2 = (m_{2,1}, \dots, m_{2,p})'$. Eqs. (6), (8) and (9) allow to estimate the autoregressive parameters as well as the mean and variance of (Z_t). To adjust the model to the skewness of the series, we introduce the following third-order moment equations. According to (7),

$$\begin{aligned} m_{3,0} - \phi' m_3 &= \mathrm{E}\,\widetilde{Z}_t\widetilde{X}_t^2 = \mathrm{E}\,\widetilde{Z}_t(\phi_1\widetilde{X}_{t-1} + \cdots + \phi_p\widetilde{X}_{t-p} + \widetilde{Z}_t)^2 \\ &= \mathrm{E}\,\widetilde{Z}_t^3 = \frac{2\sqrt{2}\sigma^3\epsilon}{\pi^{3/2}}[(5\pi - 16)\epsilon^2 - \pi], \end{aligned} \tag{10}$$

where $m_3 = (m_{3,1}, \dots, m_{3,p})'$. We deduce from (9) and (10) that

$$g(\epsilon) = \frac{m_{3,0} - \phi' m_3}{(m_{2,0} - \phi' m_2)^{3/2}}, \tag{11}$$

where $g : (-1, 1) \to (-c_0, c_0)$ is defined by

$$g(x) = 2\sqrt{2}x\frac{(5\pi - 16)x^2 - \pi}{[(3\pi - 8)x^2 + \pi]^{3/2}}. \tag{12}$$

Function g is continuously differentiable on $(-1, 1)$ with derivative

$$g'(x) = 2\sqrt{2}\pi\frac{(21\pi - 64)x^2 - \pi}{[(3\pi - 8)x^2 + \pi]^{5/2}}.$$

On the interval $(-1, 1)$, $g' < 0$ and then g is strictly monotone which implies that g is an homeomorphism from $(-1, 1)$ onto $(-c_0, c_0)$.

The ME $\widehat{\eta}_n = (\widehat{\phi}_n', \widehat{\theta}_n, \widehat{\epsilon}_n, \widehat{\sigma}_n^2)'$ of η is obtained by replacing in (6), (8), (9) and (11) the moments m, $m_{2,k}$ and $m_{3,k}$ by the sample moments

$$\widehat{m} = \frac{1}{n}\sum_{t=1}^{n} X_t, \qquad \widehat{m}_{2,k} = \frac{1}{n}\sum_{t=1}^{n-k}\widehat{X}_t\widehat{X}_{t+k}, \qquad \widehat{m}_{3,k} = \frac{1}{n}\sum_{t=1}^{n-k}\widehat{X}_t\widehat{X}_{t+k}^2, \tag{13}$$

where $\widehat{X}_t = X_t - \widehat{m}$. Therefore, $\widehat{\eta}_n$ satisfies the equations

$$\begin{aligned}
\widehat{\phi}_n &= \widehat{M}_2^{-1}\widehat{m}_2,\\
\widehat{\epsilon}_n &= g^{-1}\left(\frac{\widehat{m}_{3,0} - \widehat{\phi}_n'\widehat{m}_3}{(\widehat{m}_{2,0} - \widehat{\phi}_n'\widehat{m}_2)^{3/2}}\right),\\
\widehat{\sigma}_n^2 &= \frac{(\widehat{m}_{2,0} - \widehat{\phi}_n'\widehat{m}_2)\pi}{(3\pi - 8)\widehat{\epsilon}_n^2 + \pi},\\
\widehat{\theta}_n &= \widehat{m}(1 - e'\widehat{\phi}_n) + 4\widehat{\sigma}_n\widehat{\epsilon}_n/\sqrt{2\pi}.
\end{aligned} \tag{14}$$

We have $\widehat{m}_{2,0} - \widehat{\phi}_n'\widehat{m}_2 = n^{-1}\sum_{t=1}^{n+p}\widehat{Z}_t^2 > 0$, where

$$\widehat{Z}_t = \widehat{X}_t - \widehat{\phi}_{n,1}\widehat{X}_{t-1} - \cdots - \widehat{\phi}_{n,p}\widehat{X}_{t-p}$$

and $\widehat{X}_t = 0$ if $t < 1$ or $t > n$. When $(\widehat{m}_{3,0} - \widehat{\phi}_n'\widehat{m}_3)(\widehat{m}_{2,0} - \widehat{\phi}_n'\widehat{m}_2)^{-3/2} \notin (-c_0, c_0)$, the ME $(\widehat{\theta}_n, \widehat{\epsilon}_n, \widehat{\sigma}_n^2)$ are not defined.

Since Z_t has an ESN$(\theta, \sigma, \epsilon)$ distribution, it follows from (1) that $Z_t = \theta + \sigma W_t$ where W_t has an ESN$(0, 1, \epsilon)$ distribution. Then, according to (4), $X_t = \theta(1 - e'\phi)^{-1} + \sigma Y_t$ where Y_t does not depend on the location θ and the scale σ. Therefore, $\widehat{X}_t/\sigma$, $\widehat{m}_{2,k}/\sigma^2$, $\widehat{m}_{3,k}/\sigma^3$, the Yule–Walker estimator $\widehat{\phi}_n$, and $(\widehat{m}_{3,0} - \widehat{\phi}_n'\widehat{m}_3)(\widehat{m}_{2,0} - \widehat{\phi}_n'\widehat{m}_2)^{-3/2}$ do not depend on θ and σ.

Another ME $\bar{\eta}_n = (\widehat{\phi}_n', \bar{\theta}_n, \bar{\epsilon}_n, \bar{\sigma}_n^2)'$ of η is given by

$$\begin{aligned}
\widehat{\phi}_n &= \widehat{M}_2^{-1}\widehat{m}_2,\\
\bar{\epsilon}_n &= g^{-1}(\widehat{c}_{3,Z}/\widehat{c}_{2,Z}^{3/2}),\\
\bar{\sigma}_n^2 &= \frac{\pi\widehat{c}_{2,Z}}{(3\pi - 8)\bar{\epsilon}_n^2 + \pi},\\
\bar{\theta}_n &= \widehat{m}(1 - e'\widehat{\phi}_n) + 4\bar{\sigma}_n\bar{\epsilon}_n/\sqrt{2\pi},
\end{aligned} \tag{15}$$

where

$$\widehat{c}_{k,Z} = \frac{1}{n}\sum_{t=p+1}^{n}\widehat{Z}_t^k$$

for $k = 2, 3$. If $\widehat{c}_{3,Z}/\widehat{c}_{2,Z}^{3/2} \notin (-c_0, c_0)$, the ME $(\bar{\theta}_n, \bar{\epsilon}_n, \bar{\sigma}_n^2)$ are not defined. Since $\widehat{Z}_t/\sigma$ does not depend on θ and σ, $\widehat{c}_{k,Z}/\sigma^k$ and $\widehat{c}_{3,Z}/\widehat{c}_{2,Z}^{3/2}$ do not depend on θ and σ.

The two ME $\widehat{\eta}_n$ and $\bar{\eta}_n$ are consistent. Indeed, the linear process (X_t) is strictly stationary and ergodic, and then so are the processes (X_tX_{t+k}) and $(X_tX_{t+k}^2)$ for any fixed k, see for instance [22, Theorem 3.5.8]. The pointwise ergodic theorem for stationary sequences [22, Theorem 3.5.7] asserts that $\widehat{m}$, $\widehat{m}_{2,k}$ and $\widehat{m}_{3,k}$ converge a.s. to m, $m_{2,k}$ and $m_{3,k}$, respectively. Since transformation (14) is continuous, we deduce that $\widehat{\eta}_n \xrightarrow{a.s.} \eta$. Since

$$\widehat{Z}_t = \widetilde{Z}_t + \sum_{i=1}^{p}(\phi_i - \widehat{\phi}_{n,i})\widetilde{X}_{t-i} + (m - \widehat{m})(1 - e'\widehat{\phi}_n)$$

and $\widehat{\phi}_n \xrightarrow{a.s.} \phi$, $\widehat{m} \xrightarrow{a.s.} m$, $\mathrm{E}\,|\widetilde{Z}_t^k\widetilde{X}_{t-1}^{k_1}\ldots\widetilde{X}_{t-p}^{k_p}| < \infty$ for all $k, k_1, \ldots, k_p \geq 0$, we have $\widehat{c}_{k,Z} \xrightarrow{a.s.} \mathrm{E}\,\widetilde{Z}_t^k$ for all $k \geq 1$. Hence, the continuity of transformation (15) implies that $\bar{\eta}_n \xrightarrow{a.s.} \eta$. The asymptotic normalities of $\widehat{\eta}_n$ and $\bar{\eta}_n$ are proved in the Appendix. The results are summarised in the following theorem.

Theorem 1. *Let (X_t) be defined by (4) where (Z_t) are iid random variables with an ESN $(\theta, \sigma, \epsilon)$ distribution. Let $\eta = (\phi', \theta, \epsilon, \sigma^2)' \in S$ and $\widehat{\eta}_n$, $\bar{\eta}_n$ be the ME of η defined by (14) and (15), respectively. Then, as $n \to \infty$,*

(i) $\widehat{\eta}_n \xrightarrow{a.s.} \eta$ *and* $\bar{\eta}_n \xrightarrow{a.s.} \eta$,

(ii) $n^{1/2}(\widehat{\eta}_n - \eta) \xrightarrow{d} \mathrm{N}(0, \Sigma_1)$ *and* $n^{1/2}(\bar{\eta}_n - \eta) \xrightarrow{d} \mathrm{N}(0, \Sigma_2)$.

Remark 1. Since $\widehat{\phi}_n$ is the Yule–Walker estimator of ϕ, we deduce from [23, Theorem 8.1.1] that $n^{1/2}(\widehat{\phi}_n - \phi) \xrightarrow{d} \mathrm{N}(0, c_{2,Z}M_2^{-1})$. Therefore,

$$\Sigma_{1,ij} = \Sigma_{2,ij} = c_{2,Z}M_{2,ij}^{-1} \quad \text{for } 1 \leq i, j \leq p. \tag{16}$$

Moreover, the covariance matrix $c_{2,Z}M_2^{-1}$ depends only on the parameters ϕ. The explicit expressions of the others components of matrices Σ_1 and Σ_2 seem quite cumbersome.

Remark 2. In the simple case of an iid sequence (Z_t) whose skewness ϵ is known a priori and is not estimated, the asymptotic covariance of $n^{1/2}(\check{\theta}_n - \theta, \check{\sigma}_n^2 - \sigma^2)$, where $\check{\theta}_n = \widehat{m} + 4\epsilon\check{\sigma}_n/\sqrt{2\pi}$ and $\check{\sigma}_n^2 = \pi\widehat{c}_{2,Z}/((3\pi - 8)\epsilon^2 + \pi)$ is given in [18, Theorem 3.1], showing in particular that $\check{\theta}_n$ and $\check{\sigma}_n^2$ are asymptotically correlated when $\epsilon \neq 0$.

2.2. Conditional maximum likelihood method

Here, we suppose that η_0 is the true value of η and we consider the likelihood estimator based on maximisation of the conditional likelihood of $(X_1, \ldots, X_n)$ conditionally to $(X_1, \ldots, X_p)$. According to (4), the logarithm of the conditional likelihood is

$$L_n(\eta) = \sum_{t=p+1}^{n} l(X_t, \ldots, X_{t-p}; \eta), \tag{17}$$

where, for all $x = (x_0, \ldots, x_p) \in \mathbb{R}^{p+1}$, $l(x; \eta) = \ln f(x_0 - \phi_1 x_1 - \cdots - \phi_p x_p)$ and f is the density defined by (1). Our main result is the following and is proved in the Appendix.

Theorem 2. *Let (X_t) be defined by (4) where ϕ is replaced by ϕ_0 and (Z_t) are iid random variables with an ESN $(\theta_0, \sigma_0, \epsilon_0)$ distribution, and let $\eta_0 = (\phi_0', \theta_0, \epsilon_0, \sigma_0^2)' \in S$. Then, there exists a sequence of estimators $(\widetilde{\eta}_n)$ such that, for any $\epsilon > 0$, there exists an event E with $\mathrm{P}(E) > 1 - \epsilon$ and an n_0 such that on E, for $n > n_0$, $\frac{\partial L_n}{\partial \eta}(\widetilde{\eta}_n) = 0$ and L_n attains a relative maximum at $\widetilde{\eta}_n$. Furthermore,*

(i) $\widetilde{\eta}_n \xrightarrow{a.s.} \eta_0$ *as* $n \to \infty$.

(ii) $n^{1/2}(\widetilde{\eta}_n - \eta_0) \xrightarrow{d} \mathrm{N}(0, \Sigma)$ *as* $n \to \infty$, *where*

$$\Sigma = \sigma_0^2(1 - \epsilon_0^2)\begin{pmatrix} M_2^{-1} & -mM_2^{-1}e & 0 & 0 \\ -me'M_2^{-1} & c_1 & c_2 & 0 \\ 0 & c_2 & c_3 & 0 \\ 0 & 0 & 0 & c_4 \end{pmatrix}, \tag{18}$$

$$c_1 = \frac{3\pi}{3\pi - 8} + m^2 e' M_2^{-1} e, \qquad c_2 = \frac{2\sqrt{2\pi}}{(3\pi - 8)\sigma_0}, \qquad c_3 = \frac{\pi}{(3\pi - 8)\sigma_0^2}, \qquad c_4 = \frac{2\sigma_0^2}{1 - \epsilon_0^2},$$

and M_2, m are calculated for $\eta = \eta_0$.

(iii) *The covariance matrix Σ can be estimated strongly consistently by replacing η_0 by $\widetilde{\eta}_n$ in its expression. One may also replace M_2 by the estimated covariance matrix $[\widehat{m}_{2,i-j}]_{i,j=1}^{p}$.*

Remark 3. It results from (18) that the MLE $\widetilde{\sigma}_n^2$ and $\widetilde{\phi}_n$ are asymptotically independent of $(\widetilde{\phi}_n', \widetilde{\theta}_n, \widetilde{\epsilon}_n)$ and $\widetilde{\epsilon}_n$, respectively. This property does not hold for the ME. Furthermore, we deduce from (16) and (18) that the asymptotic covariance of $\widetilde{\phi}_n$ is reduced compared to the asymptotic covariance of the Yule–Walker estimator $\widehat{\phi}_n$ by the factor

$$\frac{c_{2,Z}}{\sigma_0^2(1 - \epsilon_0^2)} = \frac{(3\pi - 8)\epsilon_0^2 + \pi}{\pi(1 - \epsilon_0^2)} \in [1, \infty).$$

This factor is a strictly increasing function of $|\epsilon_0|$ and is equal to 1 in the Gaussian case $\epsilon_0 = 0$. On the other hand, the asymptotic variances of $\widetilde{\phi}_n$, $\widetilde{\theta}_n$, $\widetilde{\epsilon}_n$ and $\widetilde{\sigma}_n^2$ depend on (ϕ_0, ϵ_0), $(\phi_0, \theta_0, \epsilon_0, \sigma_0^2)$, ϵ_0 and σ_0^2, respectively.

Remark 4. It is interesting to compare the asymptotic covariance matrix of the MLE of $(\theta_0, \epsilon_0, \sigma_0^2)$ in (18) with the asymptotic covariance matrix Π of the MLE of $(\theta_0, \epsilon_0, \sigma_0^2)$ obtained for an iid sequence ($\phi_0 = 0$) and given in [18, Theorem 4.7] by

$$\Pi = \sigma_0^2(1 - \epsilon_0^2)\begin{pmatrix} c_7 & c_2 & 0 \\ c_2 & c_3 & 0 \\ 0 & 0 & c_4 \end{pmatrix},$$

where $c_7 = 3\pi/(3\pi - 8)$. The asymptotic variances of the estimates of ϵ_0 and σ_0^2 are the same, as well as the asymptotic covariance between the estimates of θ_0 and ϵ_0, while the asymptotic variance of the estimate of θ_0 is inferior in the iid case unless (Z_t) is zero-mean. The variance stabilising transformations for the parameters ϵ_0 and σ_0^2 are therefore the ones given in [18, Theorem 4.11]. In particular, an approximate $(1 - \alpha)$ confidence interval for ϵ_0 is

$$\left(\sin\left\{\arcsin(\widetilde{\epsilon}_n) - z_{\alpha/2}\sqrt{\frac{\pi}{(3\pi - 8)n}}\right\}, \sin\left\{\arcsin(\widetilde{\epsilon}_n) + z_{\alpha/2}\sqrt{\frac{\pi}{(3\pi - 8)n}}\right\}\right), \tag{19}$$

where z_α denotes the $(1 - \alpha)$ quantile of the standard normal distribution.

Remark 5. When the skewness ϵ_0 is known a priori and is not estimated, the asymptotic covariance Ψ of the MLE of $(\phi_0', \theta_0, \sigma_0^2)'$ is obtained by inverting the matrix $V = \Sigma^{-1}$ whose $(p+2)$th row and $(p+2)$th column have been deleted. It results from (34) that

$$\Psi = \sigma_0^2(1-\epsilon_0^2)\begin{pmatrix} M_2^{-1} & -mM_2^{-1}e & 0 \\ -me'M_2^{-1} & c_8 & 0 \\ 0 & 0 & c_4 \end{pmatrix}, \tag{20}$$

where $c_8 = 1 + m^2 e' M_2^{-1} e$ and M_2, m are calculated for $\eta = \eta_0$. Therefore, the asymptotic variance of the MLE of the location θ_0 is reduced by the factor

$$\frac{c_1}{c_8} = \frac{\frac{3\pi}{3\pi-8} + m^2 e' M_2^{-1} e}{1 + m^2 e' M_2^{-1} e}$$

when the value of the skewness, if known a priori, is used. The asymptotic covariances of $\widetilde{\phi}_n$ and $\widetilde{\sigma}_n^2$ are unchanged.

Remark 6. When it is known that (X_t) is Gaussian and thus the skewness is not estimated, $(\widetilde{\phi}_n, \widetilde{\theta}_n)$ coincide with the usual least squares estimates (LSE) $(\phi_n^{\mathrm{LS}}, \theta_n^{\mathrm{LS}})$ obtained by minimising the sum of squares

$$S(\phi, \theta) = \sum_{t=p+1}^{n} (X_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p} - \theta)^2,$$

and $\widetilde{\sigma}_n^2 = \sigma_n^{2,\mathrm{LS}} = (n-p)^{-1} S(\phi_n^{\mathrm{LS}}, \theta_n^{\mathrm{LS}})$. The corresponding asymptotic covariance is given by (20) where $\epsilon_0 = 0$. To check this, we set $\upsilon_2 = 1$ in (26). Then, the partial derivatives $\partial l/\partial\phi_i$, $\partial l/\partial\theta$ and $\partial l/\partial\sigma^2$ are given by (28) where $\upsilon_2 = 1$ and $(\widetilde{\phi}_n, \widetilde{\theta}_n, \widetilde{\sigma}_n^2)$ satisfy the equations $\partial L_n/\partial\eta = 0$, i.e.,

$$\sum_{t=p+1}^{n} (X_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p} - \theta) X_{t-i} = 0,$$
$$\sum_{t=p+1}^{n} (X_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p} - \theta) = 0,$$
$$\sum_{t=p+1}^{n} (X_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p} - \theta)^2 = (n-p)\sigma^2,$$

whose solutions are indeed the LSE $(\phi_n^{\mathrm{LS}}, \theta_n^{\mathrm{LS}}, \sigma_n^{2,\mathrm{LS}})$.

Remark 7. It is instructive to study the properties of $(\phi_n^{\mathrm{LS}}, \theta_n^{\mathrm{LS}}, \sigma_n^{2,\mathrm{LS}})$ when $\epsilon_0 \neq 0$. According to the standard theory, $(\phi_n^{\mathrm{LS}}, \theta_n^{\mathrm{LS}}, \sigma_n^{2,\mathrm{LS}}) \xrightarrow{a.s.} (\phi_0, c_{1,Z}, c_{2,Z})$ and $n^{1/2}(\phi_n^{\mathrm{LS}} - \phi_0) \xrightarrow{d} \mathrm{N}(0, c_{2,Z} M_2^{-1})$. Therefore, when $\epsilon_0 \neq 0$, ϕ_n^{LS} is consistent but is not efficient, and we deduce from (3) that θ_n^{LS} tends to overestimate the location θ when $\epsilon_0 < 0$ and tends to underestimate θ when $\epsilon_0 > 0$, and $\sigma_n^{2,\mathrm{LS}}$ tends to overestimate σ^2.

3. Numerical results

3.1. Simulation study

The standard ESN$(0, 1, \epsilon)$ distribution is a mixture of two half-normal distributions and may be generated by $(1-U)(1-\epsilon)|N_1| - U(1+\epsilon)|N_2|$ where U, N_1, N_2 are mutually independent, $\mathrm{P}(U=1) = (1+\epsilon)/2 = 1 - \mathrm{P}(U=0)$, and N_1, N_2 are N(0, 1), see [18]. If Z has an ESN$(0, 1, \epsilon)$ distribution, then $\theta + \sigma Z$ has an ESN$(\theta, \sigma, \epsilon)$ distribution. Therefore, AR models with ESN innovations are easily generated.

In the following, we consider a causal AR(1) model defined by

$$X_t = \phi X_{t-1} + Z_t, \tag{21}$$

where $|\phi| < 1$ and (Z_t) is a sequence of iid random variables with an ESN$(\theta, \sigma, \epsilon)$ distribution. We compare the different ME of the parameters $\eta = (\phi, \theta, \epsilon, \sigma^2)'$ for finite samples, and we discuss the advantages of the MLE.

The centred series $\widetilde{X}_t$ satisfies the difference equation $\widetilde{X}_t = \phi\widetilde{X}_{t-1} + \widetilde{Z}_t$, where $\widetilde{Z}_t = Z_t - \mathrm{E}\,Z_t$. Therefore, we have

$$\mathrm{E}\,\widetilde{X}_t^3 = \mathrm{E}(\phi\widetilde{X}_{t-1} + \widetilde{Z}_t)^3 = \phi^3\,\mathrm{E}\,\widetilde{X}_{t-1}^3 + \mathrm{E}\,\widetilde{Z}_t^3,$$

which is equivalent to

$$m_{3,0}(1-\phi^3) = \frac{2\sqrt{2}\sigma^3\epsilon}{\pi^{3/2}}[(5\pi - 16)\epsilon^2 - \pi], \tag{22}$$

and we deduce from (9) and (22) that

$$g(\epsilon) = \frac{m_{3,0}(1-\phi^3)}{[m_{2,0}(1-\phi^2)]^{3/2}}, \tag{23}$$

where g is defined by (12). We can consider the ME $\mathring{\eta}_n = (\widehat{\phi}_n, \mathring{\theta}_n, \mathring{\epsilon}_n, \mathring{\sigma}_n^2)'$ of η obtained by replacing in (6), (8), (9) and (23) the moments m, $m_{2,0}$, $m_{2,1}$ and $m_{3,0}$ by the corresponding sample moments given in (13). Then $\mathring{\eta}_n$ is defined by

$$\begin{aligned} \widehat{\phi}_n &= \widehat{m}_{2,1}/\widehat{m}_{2,0}, \\ \mathring{\epsilon}_n &= g^{-1}\left(\frac{\widehat{m}_{3,0}(1-\widehat{\phi}_n^3)}{[\widehat{m}_{2,0}(1-\widehat{\phi}_n^2)]^{3/2}}\right), \\ \mathring{\sigma}_n^2 &= \frac{\widehat{m}_{2,0}(1-\widehat{\phi}_n^2)\pi}{(3\pi-8)\mathring{\epsilon}_n^2+\pi}, \\ \mathring{\theta}_n &= \widehat{m}(1-\widehat{\phi}_n) + 4\mathring{\sigma}_n\mathring{\epsilon}_n/\sqrt{2\pi}. \end{aligned} \tag{24}$$

If $\widehat{m}_{3,0}(1-\widehat{\phi}_n^3)[\widehat{m}_{2,0}(1-\widehat{\phi}_n^2)]^{-3/2} \notin (-c_0, c_0)$, the ME $(\mathring{\theta}_n, \mathring{\epsilon}_n, \mathring{\sigma}_n^2)$ are not defined. As in Section 2.1, we observe that $\widehat{m}_{3,0}(1-\widehat{\phi}_n^3)[\widehat{m}_{2,0}(1-\widehat{\phi}_n^2)]^{-3/2}$ does not depend on θ and σ. We deduce from the pointwise ergodic theorem for stationary sequences and the continuity of transformation (24) that $\mathring{\eta}_n \xrightarrow{a.s.} \eta$. The asymptotic normality of $(\mathring{\eta}_n - \eta)$ can be deduced from the asymptotic normality of $(\widehat{m}-m, \widehat{m}_{2,0}-m_{2,0}, \widehat{m}_{2,1}-m_{2,1}, \widehat{m}_{3,0}-m_{3,0})$ by the delta method.

For the three ME $\widehat{\eta}_n$, $\bar{\eta}_n$ and $\mathring{\eta}_n$ defined by (14), (15) and (24), respectively, ϕ is estimated by the Yule–Walker estimator $\widehat{\phi}_n$ and we have $n^{1/2}(\widehat{\phi}_n - \phi) \xrightarrow{d} \mathrm{N}(0, 1-\phi^2)$. Therefore, the asymptotic variance of $\widehat{\phi}_n$ depends only on $|\phi|$. We compare the percentages of undefined ME and the mean-squared error (MSE) of the ME of θ, ϵ and σ^2 for the three ME when $\theta = 10$, $\sigma^2 = 1$, and ϕ and ϵ vary in $(-1, 1)$.

To build Fig. 1, we generate 2000 independent realisations of size 1000 of model (21) and we plot the percentages of undefined ME when $n = 300$ and $n = 1000$. These percentages do not depend on θ and σ, and we have observed in the simulations that for each ME, they depend very few on the signs of ϕ and ϵ. For that reason, they are plotted in Fig. 1 as functions of $|\phi|$ and $|\epsilon|$. We observe that for all ME the percentages increase as $|\epsilon|$ increases (around 40% when $|\epsilon| = 0.95$) and decrease as n increases. Moreover, the percentages increase as $|\phi|$ increases for $\widehat{\eta}_n$ and $\mathring{\eta}_n$, these percentages being bigger for $\mathring{\eta}_n$, and the percentages depend very few on ϕ for $\bar{\eta}_n$ and are the smallest ones.

To build Figs. 2–4, we generate 1000 independent realisations of size 1000 of model (21) for which all ME are well defined. We have observed in the simulations that the MSE of the ME of ϵ and σ^2 depend very few on the signs of ϕ and ϵ, and therefore they are plotted as functions of $|\phi|$ and $|\epsilon|$. This is not the case for the MSE of the ME of θ. Fig. 2 shows that the MSE of the ME of θ are bigger for positive values of ϕ (compare the cases $\phi = -0.5$ and $\phi = 0.5$), and only the MSE of $\bar{\theta}_n$ seems to be an increasing function of ϕ for any fixed ϵ. When $\phi > 0$, the MSE decrease as ϵ increases, and when $\phi < 0$, the MSE increase as $|\epsilon|$ increases. The MSE of $\bar{\theta}_n$ is the smallest one for any (ϕ, ϵ) and is the less sensible to ϵ for any fixed ϕ. Figs. 3 and 4 show that the MSE of $\widehat{\epsilon}_n$, $\mathring{\epsilon}_n$, $\widehat{\sigma}_n^2$ and $\mathring{\sigma}_n^2$ increase as $|\phi|$ and $|\epsilon|$ increase, while the MSE of $\bar{\epsilon}_n$ and $\bar{\sigma}_n^2$ do not depend on ϕ and increase slightly as $|\epsilon|$ increase, and are the smallest ones.

The comparison of the three ME shows that $\bar{\eta}_n$ is the best one in terms of percentage of undefined ME and MSE of the estimation of θ, ϵ and σ^2, while $\mathring{\eta}_n$ is the worth one.

To compare $\bar{\eta}_n$ with the MLE $\widetilde{\eta}_n$, we generate 1000 independent realisations of size 1000 of model (21) for which the ME $\bar{\eta}_n$ are well defined, and for each realisation, $\bar{\eta}_n$ is used as initial value in a quasi-Newton method to find $\widetilde{\eta}_n$. We take $\phi = 0.8$, $\theta = 10$, $\sigma^2 = 1$ and ϵ varies in $(-1, 1)$. In Fig. 5, we plot the MSE of $\bar{\eta}_n$ and $\widetilde{\eta}_n$ and the asymptotic variances of $\widetilde{\eta}_n$. For the four parameters, the MSE of $\bar{\eta}_n$ are significantly greater than those of $\widetilde{\eta}_n$ and the differences between the MSE increase as $|\epsilon|$ increases. Furthermore, the MSE of $\widetilde{\eta}_n$ are close from the asymptotic variances given by (18).

In Fig. 6, we plot the LSE $(\phi_n^{\mathrm{LS}}, \theta_n^{\mathrm{LS}}, \sigma_n^{2,\mathrm{LS}})$ and the MLE $(\widetilde{\phi}_n, \widetilde{\theta}_n, \widetilde{\sigma}_n^2)$ obtained from 1000 independent realisations of size 1000 of model (21) where $\phi = 0.8$, $\theta = 10$, $\sigma^2 = 1$ and ϵ varies in $(-1, 1)$. We observe that $\widetilde{\phi}_n$ is closer from the true value than ϕ_n^{LS} when $|\epsilon|$ increases. Furthermore, as noticed in Remark 7, θ_n^{LS} and $\sigma_n^{2,\mathrm{LS}}$ are biased estimates of the location θ and the squared scale σ^2, respectively, and the bias increase as $|\epsilon|$ increases.

3.2. A real time series example

We consider the Dow–Jones Utilities Index from July 3, 1972 through December 20, 1972, and we ignore the unequal spacing of the data resulting from the five-day working week. The very slowly decaying positive sample autocorrelation function of this time series suggests differencing at lag one before attempting to fit a stationary model, see [24, Example 5.1.1]. Fig. 7 shows that the differenced series is asymmetric and may be modeled by an AR(1) process.

The MLE and the Gaussian LSE presented in Remark 6 are given in Table 1 where the variances are calculated using (18) where η_0 is replaced by the MLE, and (20) where $\epsilon_0 = 0$ and $(\phi_0, \theta_0, \sigma_0^2)$ are replaced by the LSE, respectively.

The approximate 95% confidence interval for the skewness parameter ϵ deduced from Table 1 is $(-0.56, -0.04)$ and the confidence interval deduced from (19) is $(-0.54, -0.04)$. On the basis of both intervals, we reject at the 5% significance level the hypothesis that ϵ is zero.

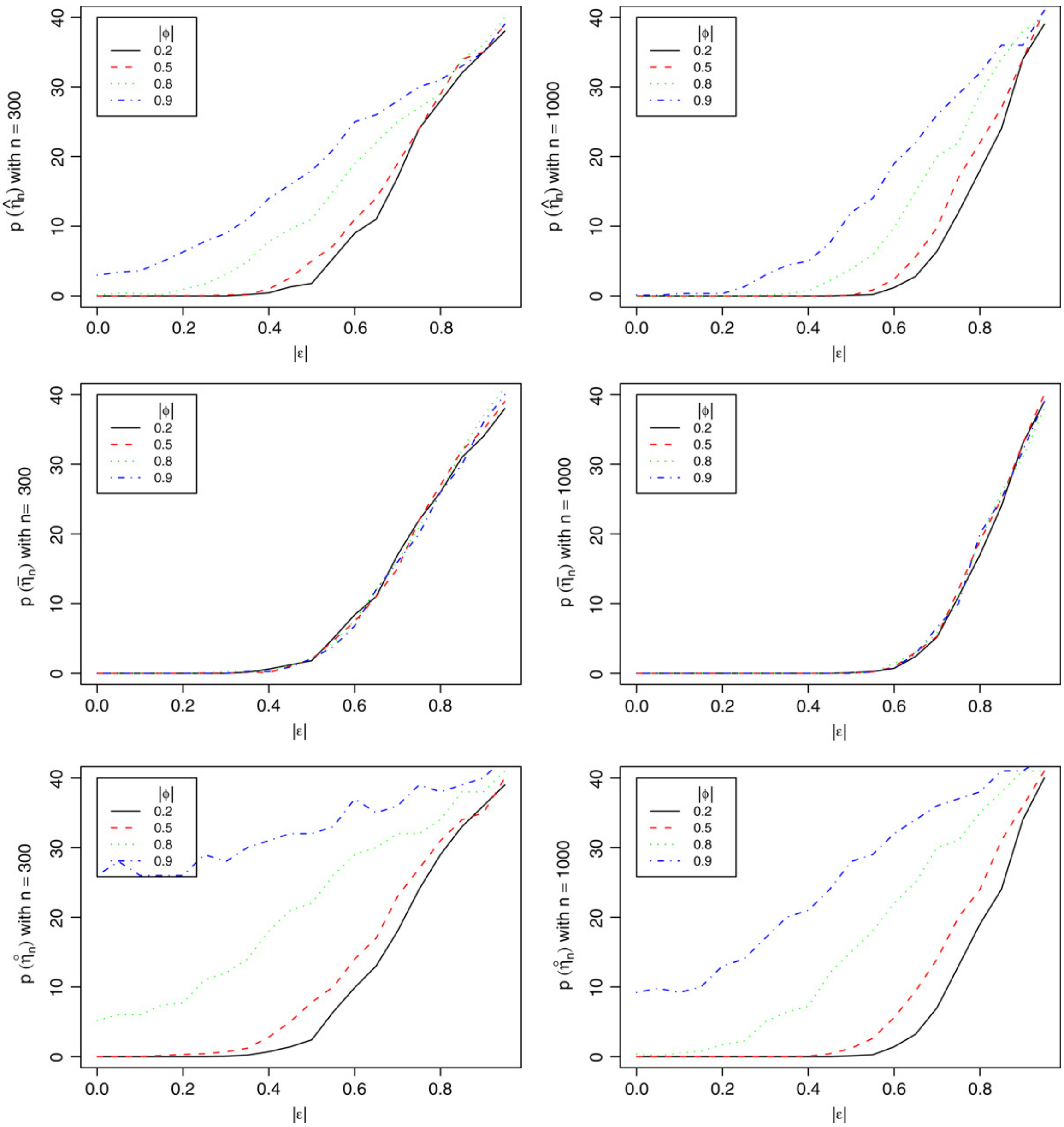

Fig. 1. Percentages of undefined ME, $p(\widehat{\eta}_n)$, $p(\bar{\eta}_n)$ and $p(\mathring{\eta}_n)$ when $\theta = 10$, $\sigma^2 = 1$, $n = 300$ and $n = 1000$.

Table 1
MLE and LSE of an AR(1) model fitted to the differenced series of the Dow–Jones Utilities Index (Jul. 3 – Dec. 20, 1972).

	MLE		LSE	
	Estimate	Variance	Estimate	Variance
ϕ	0.47	5.8×10^{-3}	0.49	6.4×10^{-3}
θ	−0.11	6.3×10^{-3}	0.05	1.2×10^{-3}
σ^2	0.12	2.6×10^{-4}	0.13	2.8×10^{-4}
ϵ	−0.30	1.7×10^{-2}		

The p-value of the Shapiro–Wilk test applied to the residuals of the AR(1) model obtained with the LSE is 1.9×10^{-2} and the p-value of the Jarque–Bera test applied to these residuals is 4.6×10^{-4}. Therefore, both tests reject the null hypothesis of normality at the 95% confidence level.

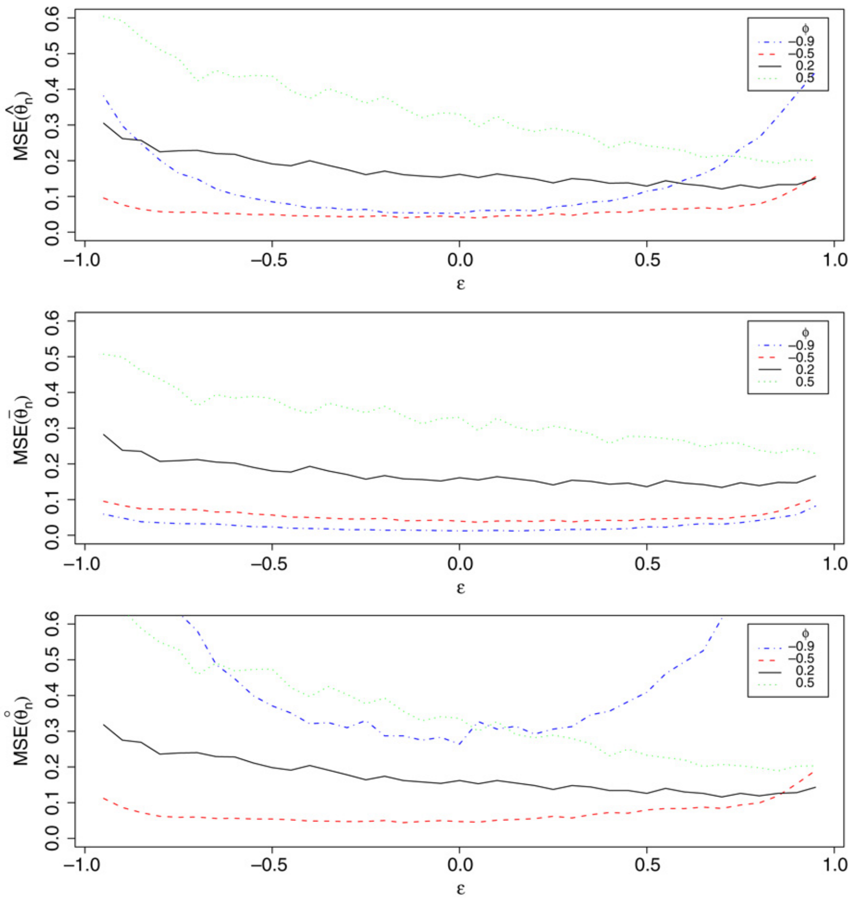

Fig. 2. MSE of $\widehat{\theta}_n$, $\bar{\theta}_n$ and $\mathring{\theta}_n$ when $\theta = 10$, $\sigma^2 = 1$ and $n = 1000$.

Acknowledgments

The author is grateful to two anonymous referees for their helpful comments and suggestions. Part of this work was developed while the author was visiting the Department of Statistics, Pontificia Universidad Católica de Chile, whose kind hospitality is gratefully acknowledged. The author would also like to thank Reinaldo Arellano, Wilfredo Palma and Mario Tagle for insightful remarks.

Appendix

Asymptotic normality of $\widehat{\eta}_n$. Let

$$\widetilde{m}_{2,k} = \frac{1}{n}\sum_{t=1}^{n} \widetilde{X}_t \widetilde{X}_{t+k},$$

$$\widetilde{m}_{3,k} = \frac{1}{n}\sum_{t=1}^{n} \widetilde{X}_t \widetilde{X}_{t+k}^2,$$

$$\widehat{m}_n = (\widehat{m} - m, \widehat{m}_{2,0} - m_{2,0}, \ldots, \widehat{m}_{2,p} - m_{2,p}, \widehat{m}_{3,0} - m_{3,0}, \ldots, \widehat{m}_{3,p} - m_{3,p})',$$
$$\widetilde{m}_n = (\widehat{m} - m, \widetilde{m}_{2,0} - m_{2,0}, \ldots, \widetilde{m}_{2,p} - m_{2,p}, \widetilde{m}_{3,0} - m_{3,0}, \ldots, \widetilde{m}_{3,p} - m_{3,p})'.$$

According to [23, Propositions 7.3.4], $n^{1/2}(\widetilde{m}_{2,k} - \widehat{m}_{2,k}) \xrightarrow{p} 0$. Simple algebra gives

$$n^{1/2}(\widetilde{m}_{3,k} - \widehat{m}_{3,k}) = n^{1/2}(\widehat{m} - m)\left[\frac{1}{n}\sum_{t=1}^{n-k}\widetilde{X}_{t+k}^2 + \frac{2}{n}\sum_{t=1}^{n-k}\widetilde{X}_t\widetilde{X}_{t+k} - (\widehat{m} - m)\frac{1}{n}\sum_{t=1}^{n-k}(2\widetilde{X}_{t+k} + \widetilde{X}_t) + \frac{n-k}{n}(\widehat{m} - m)^2\right] + n^{-1/2}\sum_{t=n-k+1}^{n}\widetilde{X}_t\widetilde{X}_{t+k}^2.$$

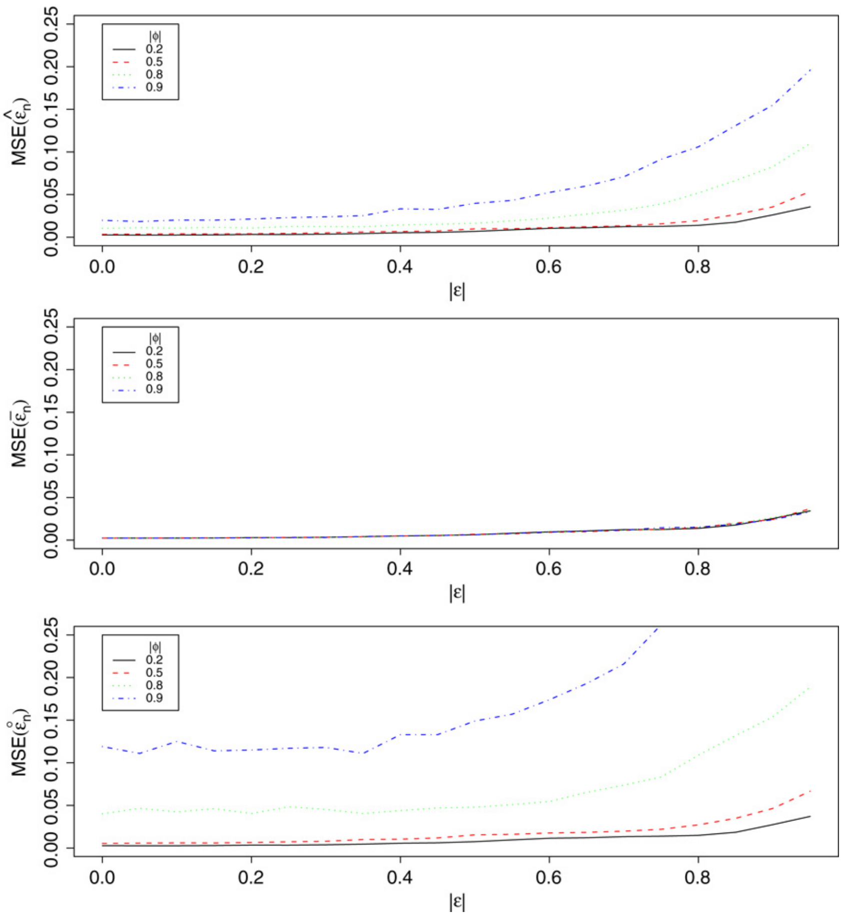

Fig. 3. MSE of $\widehat{\epsilon}_n$, $\bar{\epsilon}_n$ and $\mathring{\epsilon}_n$ when $\theta = 10, \sigma^2 = 1$ and $n = 1000$.

Since $n^{-1/2}\,\mathrm{E}\,|\sum_{t=n-k+1}^{n}\widetilde{X}_t\widetilde{X}_{t+k}^2| \le n^{-1/2}k(\mathrm{E}\,\widetilde{X}_t^2\,\mathrm{E}\,\widetilde{X}_t^4)^{1/2}$, the last term converges to 0 in probability. The term in square brackets converges a.s. to $m_{2,0} + 2m_{2,k}$ and $n^{1/2}(\widehat{m} - m)$ converges in distribution. Therefore, it follows from Slutsky's lemma that $n^{1/2}\widehat{m}_n$ and $n^{1/2}(\widetilde{m}_n - (\widehat{m} - m)d_1)$ have the same asymptotic distribution, where

$$d_1 = (0, 0, \ldots, 0, 3m_{2,0}, m_{2,0} + 2m_{2,1}, \ldots, m_{2,0} + 2m_{2,p})'.$$

This asymptotic distribution is easily obtained from the one of $n^{1/2}\widetilde{m}_n$. Now, using the same technique as in [23, Propositions 7.3.1, 7.3.2, 7.3.3], we can show that $n^{1/2}\widetilde{m}_n \xrightarrow{d} \mathrm{N}(0, \Upsilon_1)$, where $\Upsilon_1 = \lim_{n\to\infty}\Upsilon_{1,n}$ and $\Upsilon_{1,n}$ is the covariance matrix of $n^{1/2}\widetilde{m}_n$. Then, $n^{1/2}\widehat{m}_n \xrightarrow{d} \mathrm{N}(0, \Omega_1)$, where

$$\Omega_{1,ij} = \Omega_{1,ji} = \Upsilon_{1,ij} - d_{1,j}\Upsilon_{1,1i} - d_{1,i}\Upsilon_{1,1j} + d_{1,i}d_{1,j}\Upsilon_{1,11}, \tag{25}$$

and $n^{1/2}(\widehat{\eta}_n - \eta) \xrightarrow{d} \mathrm{N}(0, \Sigma_1)$, where $\Sigma_1 = D_1\Omega_1 D_1'$ and D_1 is the gradient of the transformation (14) that associates $\widehat{\eta}_n$ to $(\widehat{m}, \widehat{m}_{2,0}, \ldots, \widehat{m}_{2,p}, \widehat{m}_{3,0}, \ldots, \widehat{m}_{3,p})$.

Asymptotic normality of $\bar{\eta}_n$. Let

$$\widetilde{c}_{k,Z} = \frac{1}{n}\sum_{t=p+1}^{n}\widetilde{Z}_t^k,$$

$$\widehat{c}_n = (\widehat{m} - m, \widehat{m}_{2,0} - m_{2,0}, \ldots, \widehat{m}_{2,p} - m_{2,p}, \widehat{c}_{2,Z} - c_{2,Z}, \widehat{c}_{3,Z} - c_{3,Z})',$$
$$\widetilde{c}_n = (\widehat{m} - m, \widetilde{m}_{2,0} - m_{2,0}, \ldots, \widetilde{m}_{2,p} - m_{2,p}, \widetilde{c}_{2,Z} - c_{2,Z}, \widetilde{c}_{3,Z} - c_{3,Z})',$$

where $c_{2,Z}$ and $c_{3,Z}$ are given by (3). Using the causality of model (7) and Slutsky's lemma, it follows from easy calculations that $n^{1/2}(\widetilde{c}_{2,Z} - \widehat{c}_{2,Z}) \xrightarrow{p} 0$ and $n^{1/2}(\widetilde{c}_{3,Z} - \widehat{c}_{3,Z} - d(\widehat{m} - m)) \xrightarrow{p} 0$ where $d = 3c_{2,Z}(1 - e'\phi)$. Then $n^{1/2}\widehat{c}_n$ and $n^{1/2}(\widetilde{c}_n - (\widehat{m} - m)d_2)$ have the same asymptotic distribution, where

$$d_2 = (0, \ldots, 0, d)'.$$

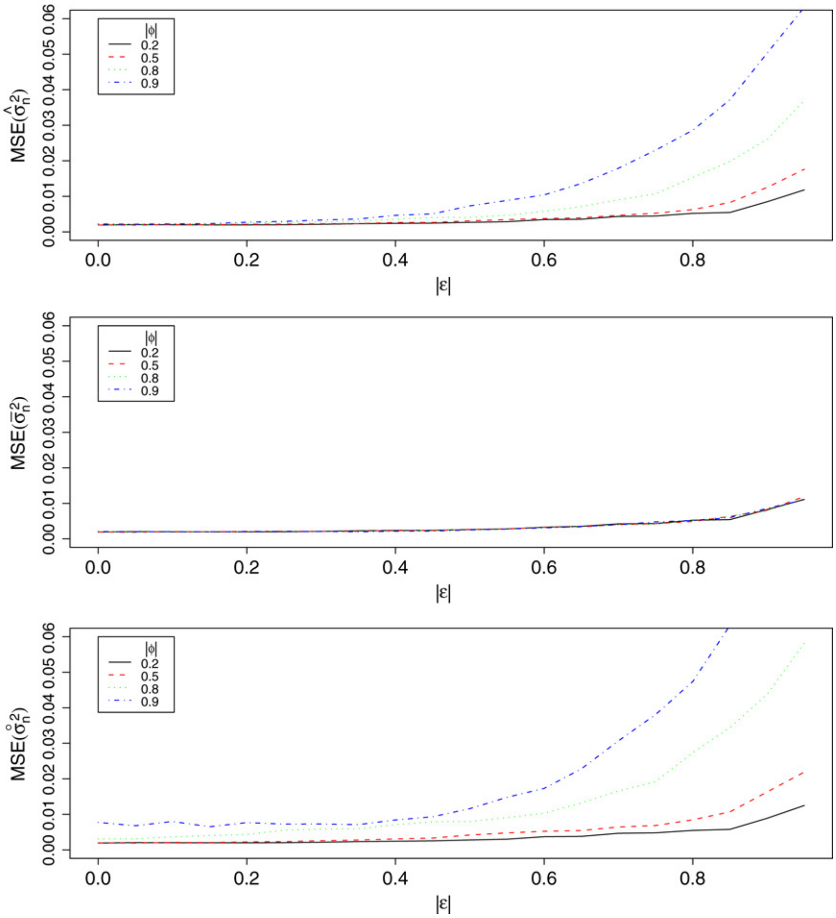

Fig. 4. MSE of $\widehat{\sigma}_n^2$, $\bar{\sigma}_n^2$ and $\mathring{\sigma}_n^2$ when $\theta = 10$, $\sigma^2 = 1$ and $n = 1000$.

As in [23, Propositions 7.3.1, 7.3.2, 7.3.3], we can establish that $n^{1/2}\widetilde{c}_n \xrightarrow{d} \mathrm{N}(0, \varUpsilon_2)$, where $\varUpsilon_2 = \lim_{n\to\infty} \varUpsilon_{2,n}$ and $\varUpsilon_{2,n}$ is the covariance matrix of $n^{1/2}\widetilde{c}_n$. Then $n^{1/2}\widehat{c}_n \xrightarrow{d} \mathrm{N}(0, \varOmega_2)$, where $\varOmega_2$ is given by (25) in which $\varOmega_1, \varUpsilon_1, d_1$ are replaced by $\varOmega_2, \varUpsilon_2, d_2$, respectively. Hence, $n^{1/2}(\bar{\eta}_n - \eta) \xrightarrow{d} \mathrm{N}(0, \varSigma_2)$, where $\varSigma_2 = D_2\varOmega_2 D_2'$ and D_2 is the gradient of the transformation (15) that associates $\bar{\eta}_n$ to $(\widehat{m}, \widehat{m}_{2,0}, \dots, \widehat{m}_{2,p}, \widehat{c}_{2,Z}, \widehat{c}_{3,Z})$.

Proof of Theorem 2. The basic technique of proof is to control the behavior of the first and second order terms in a Taylor expansion of $L_n(\eta)$ about η_0. The log-likelihood $L_n(\eta)$ is defined by (17) where for all $x \in \mathbb{R}^{p+1}$,

$$l(x; \eta) = -\frac{1}{2}\ln(2\pi\sigma^2) - \frac{u^2 v_2}{2\sigma^2} \tag{26}$$

and

$$\begin{aligned} u &= x_0 - \phi_1 x_1 - \cdots - \phi_p x_p - \theta, \\ v_n &= \frac{\mathbb{1}_{(-\infty,0)}(u)}{(1+\epsilon)^n} + (-1)^n \frac{\mathbb{1}_{(0,+\infty)}(u)}{(1-\epsilon)^n}, \quad \forall n \in \mathbb{N}. \end{aligned} \tag{27}$$

Function $l(x; \eta)$ is twice continuously differentiable with respect to η in some neighbourhood $\mathcal{N}_0$ of η_0 for almost all $x \in \mathbb{R}^{p+1}$. Without further notice, all neighbourhoods defined below are taken to be contained in $\mathcal{N}_0$. Then, for $\delta > 0$, $\|\eta - \eta_0\| < \delta$,

$$L_n(\eta) = L_n(\eta_0) + (\eta-\eta_0)'\frac{\partial L_n}{\partial \eta}(\eta_0) + \frac{1}{2}(\eta-\eta_0)'V_n(\eta-\eta_0) + \frac{1}{2}(\eta-\eta_0)'T_n(\eta^\star)(\eta-\eta_0),$$

where

$$V_n = \frac{\partial^2 L_n}{\partial\eta\partial\eta'}(\eta_0), \qquad T_n(\eta^\star) = \frac{\partial^2 L_n}{\partial\eta\partial\eta'}(\eta^\star) - \frac{\partial^2 L_n}{\partial\eta\partial\eta'}(\eta_0),$$

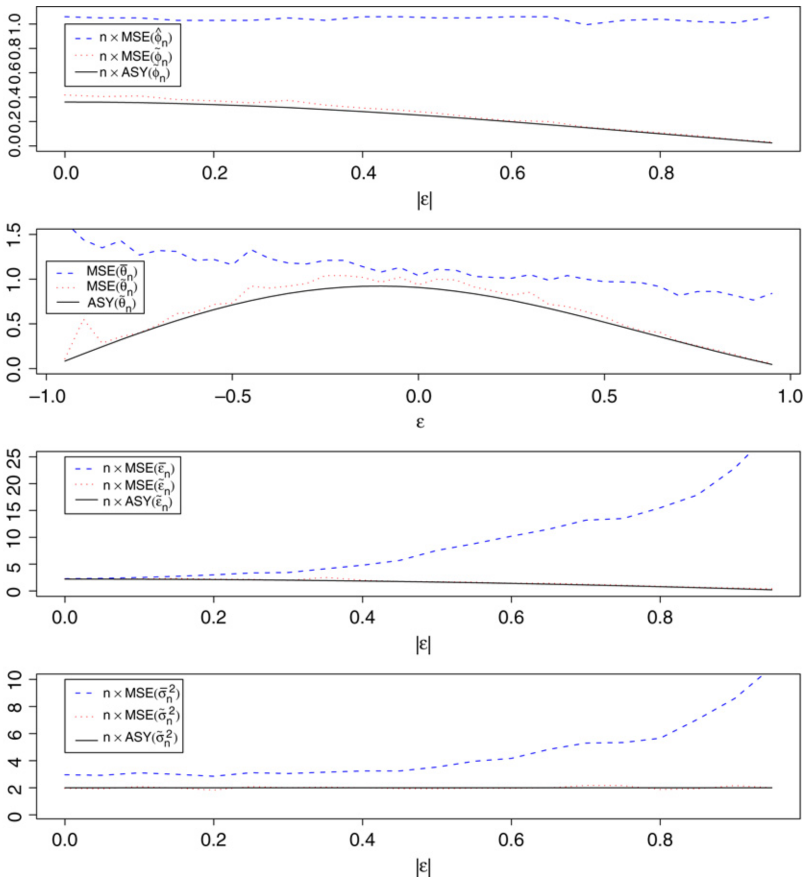

Fig. 5. MSE and asymptotic variances (ASY) of $\bar{\eta}_n$ and $\widetilde{\eta}_n$ when $\phi = 0.8$, $\theta = 10$, $\sigma^2 = 1$ and $n = 1000$.

and $\eta^\star = \eta^\star(X_1, \ldots, X_n; \eta)$ is an intermediate point between η and η_0. According to [25, Corollary 2.1 and Theorem 2.2], see also [26, Theorem 3.2.23], Theorem 2(i) holds if the three following conditions are satisfied as $n \to \infty$,

(A1) $n^{-1} \frac{\partial L_n}{\partial \eta}(\eta_0) \xrightarrow{a.s.} 0$,

(A2) $n^{-1} V_n \xrightarrow{a.s.} -V$, where V is a positive definite matrix,

(A3) for $1 \le i, j \le p+3$, $\lim_{n\to\infty} \sup_{\delta\to 0} (n\delta)^{-1} |T_n(\eta^\star)_{i,j}| < \infty$ a.s., where $T_n(\eta^\star)_{i,j}$ is the (i,j)th component of $T_n(\eta^\star)$.

If, in addition,

(A4) $n^{-1/2} \frac{\partial L_n}{\partial \eta}(\eta_0) \xrightarrow{d} \mathrm{N}(0, V)$,

then Theorem 2(ii) holds with $\Sigma = V^{-1}$.

Let us check condition (A1). Using that $\mathrm{d}v_n/\mathrm{d}\epsilon = -n v_{n+1}$, we deduce from (26) and (27), that

$$\frac{\partial l}{\partial \phi_i}(x; \eta) = \frac{u x_i v_2}{\sigma^2}, \qquad \frac{\partial l}{\partial \theta}(x; \eta) = \frac{u v_2}{\sigma^2}, \qquad \frac{\partial l}{\partial \epsilon}(x; \eta) = \frac{u^2 v_3}{\sigma^2}, \qquad \frac{\partial l}{\partial \sigma^2}(x; \eta) = -\frac{1}{2\sigma^2} + \frac{u^2 v_2}{2\sigma^4}. \tag{28}$$

Observe that

$$|v_n| \le \frac{2}{(1-|\epsilon|)^n}. \tag{29}$$

Therefore, finiteness of $\mathrm{E}\, X_t^2$ and (28) imply that

$$\mathrm{E}\left| \frac{\partial l}{\partial \eta_i}(X_t, \ldots, X_{t-p}; \eta_0) \right| < \infty, \quad \text{for } 1 \le i \le p+3. \tag{30}$$

Since (X_t) is strictly stationary and ergodic, so is the process $(U_t)_{t>p}$ where

$$U_t = \frac{\partial l}{\partial \eta}(X_t, \ldots, X_{t-p}; \eta_0), \tag{31}$$

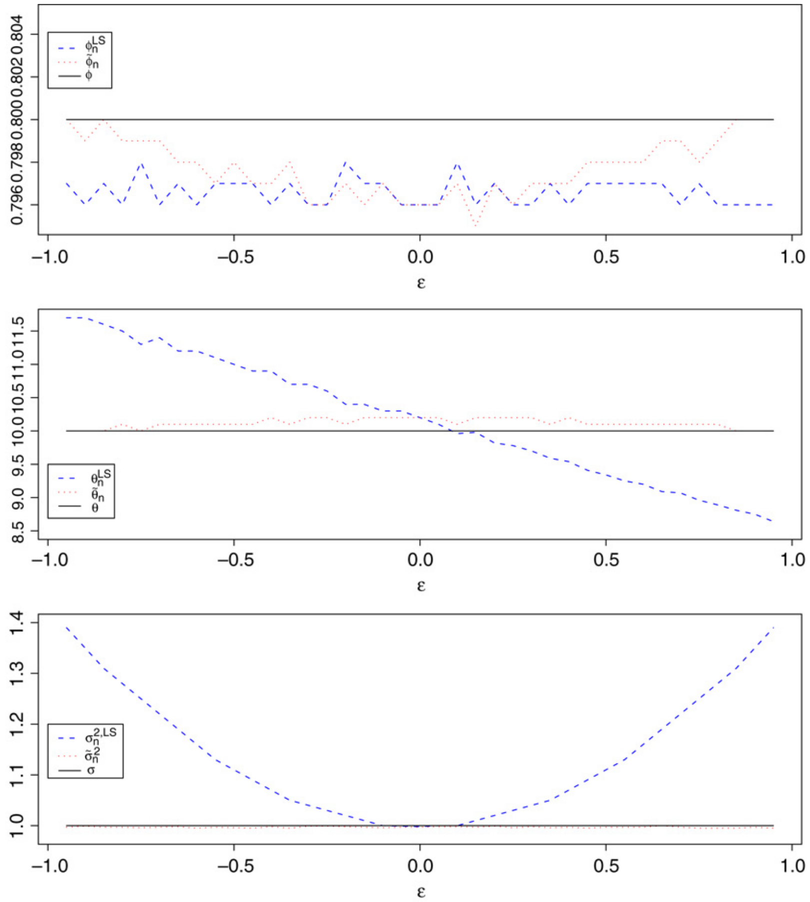

Fig. 6. LSE and MLE of (ϕ, θ, σ^2) when $\phi = 0.8, \theta = 10, \sigma^2 = 1$ and $n = 1000$.

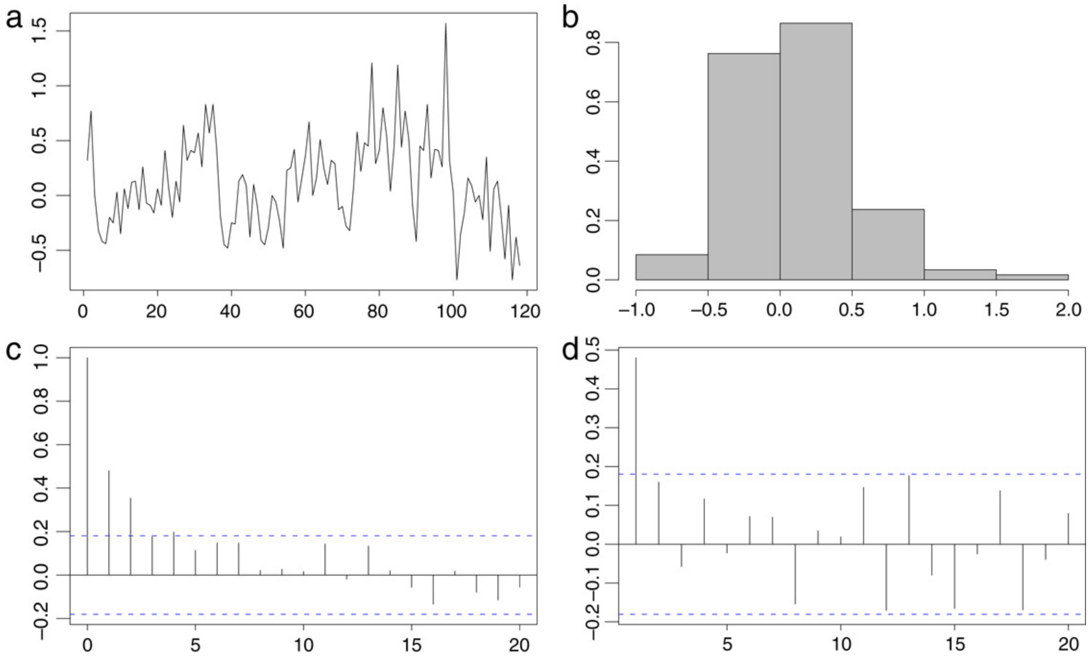

Fig. 7. Differenced series of the Dow–Jones Utilities Index (Jul. 3 – Dec. 20, 1972); (a) Series, (b) Histogram, (c) Autocorrelation function, (d) Partial autocorrelation function.

and it follows from the pointwise ergodic theorem for stationary sequences [22, Theorem 3.5.7] and (30) that $n^{-1}\frac{\partial L_n}{\partial \eta}(\eta_0) \xrightarrow{a.s.} \mathrm{E}\, U_t$. Since (X_t) is causal, X_{t-i} and Z_t are independent when $i > 0$, and then for all non negative integers k, l, m, n and positive integers i, j,

$$\mathrm{E}\left[(Z_t-\theta_0)^k X_{t-i}^l X_{t-j}^m \left(\frac{\mathbb{1}_{(-\infty,\theta_0)}(Z_t)}{(1+\epsilon_0)^n} + (-1)^n \frac{\mathbb{1}_{(\theta_0,+\infty)}(Z_t)}{(1-\epsilon_0)^n}\right)\right]$$
$$= \frac{\sigma_0^k}{\sqrt{2\pi}}\left[\frac{(-1)^k I_k(-\epsilon_0)}{(1+\epsilon_0)^n} + \frac{(-1)^n I_k(\epsilon_0)}{(1-\epsilon_0)^n}\right] \mathrm{E}\, X_{t-i}^l X_{t-j}^m, \tag{32}$$

where $I_k(\epsilon)$ is defined by (2). Using that $X_t - \phi_{0,1}X_{t-1} - \phi_{0,p}X_{t-p} - \theta_0 = Z_t - \theta_0$, (28) and (32), we get $\mathrm{E}\, U_t = 0$, so that (A1) holds.

We now check condition (A2). According to (28), we have

$$\begin{aligned}
&\frac{\partial^2 l}{\partial\phi_i\partial\phi_j}(x;\eta) = -\frac{x_i x_j v_2}{\sigma^2}, \quad \frac{\partial^2 l}{\partial\phi_i\partial\theta}(x;\eta) = -\frac{x_i v_2}{\sigma^2}, \quad \frac{\partial^2 l}{\partial\phi_i\partial\epsilon}(x;\eta) = -\frac{2u x_i v_3}{\sigma^2},\\
&\frac{\partial l}{\partial\phi_i\partial\sigma^2}(x;\eta) = -\frac{u x_i v_2}{\sigma^4}, \quad \frac{\partial^2 l}{\partial\theta^2}(x;\eta) = -\frac{v_2}{\sigma^2}, \quad \frac{\partial^2 l}{\partial\theta\partial\epsilon}(x;\eta) = -\frac{2u v_3}{\sigma^2},\\
&\frac{\partial^2 l}{\partial\theta\partial\sigma^2}(x;\eta) = -\frac{u v_2}{\sigma^4}, \quad \frac{\partial^2 l}{\partial\epsilon^2}(x;\eta) = -\frac{3u^2 v_4}{\sigma^2}, \quad \frac{\partial^2 l}{\partial\epsilon\partial\sigma^2}(x;\eta) = -\frac{u^2 v_3}{\sigma^4},\\
&\frac{\partial^2 l}{\partial\sigma^4}(x;\eta) = \frac{1}{2\sigma^4} - \frac{u^2 v_2}{\sigma^6}.
\end{aligned} \tag{33}$$

Using (29) and finiteness of $\mathrm{E}\, X_t^2$, we easily deduce from (33) that

$$\mathrm{E}\left|\frac{\partial^2 l}{\partial\eta_i\partial\eta_j}(X_t,\ldots,X_{t-p};\eta_0)\right| < \infty, \quad \text{for } 1 \le i,j \le p+3.$$

The strict stationarity and ergodicity of the process $\left(\frac{\partial^2 l}{\partial\eta\partial\eta'}(X_t,\ldots,X_{t-p};\eta_0)\right)$ follows from those of (X_t), and the pointwise ergodic theorem asserts that $n^{-1}V_n \xrightarrow{a.s.} -V$ where $V = -\mathrm{E}\,\frac{\partial^2 l}{\partial\eta\partial\eta'}(X_t,\ldots,X_{t-p};\eta_0)$. Using (32) and (33), we get

$$V = \frac{1}{\sigma_0^2(1-\epsilon_0^2)}\begin{pmatrix} M_2 + m^2 ee' & me & mc_5 e & 0\\ me' & 1 & c_5 & 0\\ mc_5 e' & c_5 & 3\sigma_0^2 & 0\\ 0 & 0 & 0 & c_6\end{pmatrix} \tag{34}$$

where $c_5 = -4\sigma_0/\sqrt{2\pi}$, $c_6 = (1-\epsilon_0^2)/(2\sigma_0^2)$, and M_2, m are calculated for $\eta = \eta_0$. To show that V is positive definite, take $a = (a_1', a_2')'$ where $a_1 \in \mathbb{R}^p$ and $a_2 = (a_{21}, a_{22}, a_{23})' \in \mathbb{R}^3$. Simple algebra shows that

$$\sigma_0^2(1-\epsilon_0^2)a'Va = a_1'M_2a_1 + (me'a_1 + a_{21} + c_5a_{22})^2 + \frac{\sigma_0^2}{\pi}(3\pi-8)a_{22}^2 + c_6a_{23}^2.$$

Then $a'Va > 0$ for all $a \neq 0$ since M_2 is positive definite, $3\pi - 8 > 0$ and $c_6 > 0$. This shows (A2).

To establish condition (A3), we shall prove that there exist measurable functions $g_{i,j,k} : \mathbb{R}^{p+1} \to \mathbb{R}$, $1 \le i,j,k \le p+3$, such that, for all η in some neighbourhood $\mathcal{N}$ of η_0 and for almost all $x \in \mathbb{R}^{p+1}$,

$$\left|\frac{\partial^3 l}{\partial\eta_i\partial\eta_j\partial\eta_k}(x;\eta)\right| \le g_{i,j,k}(x) \quad \text{and} \quad \mathrm{E}\, g_{i,j,k}(X_t,\ldots,X_{t-p}) < \infty. \tag{35}$$

Indeed, (A3) follows from (35), the mean value theorem and the ergodic theorem. According to (33), all the third order partial derivatives of $l(x;\eta)$ with respect to η take the form

$$h(x;\eta) = \frac{r u^k x_i^l x_j^m v_n}{\sigma^{2q}}, \tag{36}$$

where r is an integer and k, l, m, n, q are non negative integers, except $\frac{\partial^3 l}{\partial\sigma^6}(x;\eta)$ which is the sum of two functions of type (36). To prove (35), it is therefore sufficient to show that there exists a measurable function $g : \mathbb{R}^{p+1} \to \mathbb{R}$ such that, for all $\eta \in \mathcal{N}$ and for almost all $x \in \mathbb{R}^{p+1}$, $|h(x;\eta)| \le g(x)$ and $\mathrm{E}\, g(X_t,\ldots,X_{t-p}) < \infty$. Firstly, we observe that $\phi(z) \neq 0$ for $|z| \le 1$ implies that the coefficients ϕ_k satisfy $|\phi_k| < \binom{p}{k}$. Indeed, let $z_1,\ldots,z_p$ be the unnecessarily distinct roots of $\phi(z)$, we have $\phi(z) = \prod_{i=1}^{p}(1 - z_i^{-1}z)$ and then $\phi_k = (-1)^{k+1} \sum_{1\le i_1<\cdots<i_k\le p} z_{i_1}^{-1}\cdots z_{i_k}^{-1}$, from which the result follows. Now, we choose $\mathcal{N}$ such that, for all $\eta \in \mathcal{N}$,

$$|\phi_i| < \binom{p}{i}, \quad |\epsilon| < 1-\delta \text{ where } \delta \in (0,1) \text{ is small enough}, \quad \sigma > \sigma_0/2,\ |\theta| \le 2|\theta_0|. \tag{37}$$

Then,

$$\forall(\eta, x) \in \mathcal{N} \times \mathbb{R}^{p+1}, \quad |h(x;\eta)| \le \frac{2^{2q+1}|r|}{\sigma_0^{2q}\delta^n}\left(\sum_{i=0}^{p}\binom{p}{i}|x_i| + 2|\theta_0|\right)^k |x_i|^l|x_j|^m \text{ a.e.}, \tag{38}$$

and finiteness of $\mathrm{E}\,|X_t|^k$ for all $k \ge 1$ implies that

$$\mathrm{E}\left(\sum_{i=0}^{p}\binom{p}{i}|X_{t-i}| + 2|\theta_0|\right)^k |X_{t-i}|^l|X_{t-j}|^m < \infty.$$

Thus (A3) holds.

We check now condition (A4). Firstly, we prove that the sequence $(U_t)_{t>p}$ defined by (31) is a martingale difference relative to $(\mathcal{F}_t)$ where $\mathcal{F}_t$ is the σ-algebra generated by $\{X_1,\ldots,X_t\}$. We have

$$\mathrm{E}[U_t|X_1 = x_1,\ldots,X_{t-1} = x_{t-1}] = \int_{\mathbb{R}} \frac{\partial l}{\partial\eta}(x_t, x_{t-1},\ldots,x_{t-p};\eta_0) f(x_t|x_1,\ldots,x_{t-1})\mathrm{d}x_t,$$

where $f(\cdot|x_1,\ldots,x_{t-1})$ is the true conditional density of X_t given $X_1 = x_1,\ldots,X_{t-1} = x_{t-1}$. Since (X_t) is the Markov process (4),

$$f(x_t|x_1,\ldots,x_{t-1}) = f_0(x_t - \phi_{0,1}x_{t-1} - \cdots - \phi_{0,p}x_{t-p}),$$

where f_0 is the density (1) with $(\theta,\sigma,\epsilon) = (\theta_0,\sigma_0,\epsilon_0)$. Therefore,

$$\mathrm{E}[U_t|X_1 = x_1,\ldots,X_{t-1} = x_{t-1}] = \int_{\mathbb{R}} \left.\frac{\partial f(x_t - \phi_1x_{t-1} - \cdots - \phi_px_{t-p})}{\partial\eta}\right|_{\eta_0} \mathrm{d}x_t. \tag{39}$$

The right-hand side of (39) is zero if we can interchange integration and differentiation, and this is the case if there exist functions $g_i : \mathbb{R}^{p+1} \to \mathbb{R}$, $1 \le i \le p+3$, such that, for all η in some neighbourhood $\mathcal{N}$ of η_0, for all $(x_1,\ldots,x_p) \in \mathbb{R}^p$, and for almost all $x_0 \in \mathbb{R}$,

$$\left|\frac{\partial f(x_0 - \phi_1x_1 - \cdots - \phi_px_p)}{\partial\eta_i}\right| \le g_i(x) \quad \text{and} \quad \int_{\mathbb{R}} g_i(x)\mathrm{d}x_0 < \infty, \tag{40}$$

where $x = (x_0,\ldots,x_p)$, see e.g. [27, Théorème 8.40]. According to (28), all the partial derivatives of $l(x;\eta)$ with respect to η either take the form (36) or are sums of functions of type (36). To prove (40), it is therefore sufficient to show that there exists a function $g : \mathbb{R}^{p+1} \to \mathbb{R}$ such that, for all $(\eta, x_1,\ldots,x_p) \in \mathcal{N} \times \mathbb{R}^p$,

$$|h(x;\eta)| f(x_0 - \phi_1x_1 - \cdots - \phi_px_p) \le g(x) \text{ a.e.} \quad \text{and} \quad \int_{\mathbb{R}} g(x)\mathrm{d}x_0 < \infty. \tag{41}$$

We choose $\mathcal{N}$ such that (37) holds for all $\eta \in \mathcal{N}$. Then,

$$|u| \ge \frac{|x_0|}{2} \quad \text{if } |x_0| \ge M \quad \text{where } M = 2\left[\sum_{i=1}^{p}\binom{p}{i}|x_i| + 2|\theta_0|\right]. \tag{42}$$

Indeed, since u takes the form $u = x_0 - y$ where $|y| \le M/2$, we have

$$|u| \ge |x_0| - |y| \ge |x_0| - M/2 \ge |x_0|/2$$

if $|x_0| \ge M$. We assume further that $\sigma < 2\sigma_0$ for all $\eta \in \mathcal{N}$. Since $v_2 > 1/4$, it results from (26) and (42) that

$$f(x_0 - \phi_1x_1 - \cdots - \phi_px_p) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{u^2v_2}{2\sigma^2}\right) \le \begin{cases} \dfrac{2}{\sqrt{2\pi}\sigma_0}\exp\left(-\dfrac{x_0^2}{128\sigma_0^2}\right) & \text{if } |x_0| \ge M, \\ \dfrac{2}{\sqrt{2\pi}\sigma_0} & \text{if } |x_0| < M. \end{cases}$$

Since (38) is also satisfied, there exists a function g satisfying (41) and it follows from (39) that $\mathrm{E}[U_t|\mathcal{F}_{t-1}] = 0$. According to the central limit theorem of [28,29] for stationary ergodic martingale differences, if the components of $W = \mathrm{E}\,U_tU_t'$ are finite, then $n^{-1/2}\frac{\partial L_n}{\partial\eta}(\eta_0) \xrightarrow{d} \mathrm{N}(0, W)$. Therefore, (A4) will be proved if we show that $W = V$. This may be verified directly using (28) and (32) and $v_m v_n = v_{m+n}$ for all non negative integers m, n. Alternatively, one may show that we can interchange integration and one more differentiation in (39) since this implies that $\mathrm{E}[U_t U_t' | \mathcal{F}_{t-1}] = -\mathrm{E}[\frac{\partial^2 l}{\partial \eta \partial \eta'}(X_t, \ldots, X_{t-p}; \eta_0) | \mathcal{F}_{t-1}]$, and this is the case when there exist functions $g_{i,j} : \mathbb{R}^{p+1} \to \mathbb{R}$, $1 \le i, j \le p+3$, such that, for all η in some neighbourhood $\mathcal{N}$ of η_0, for all $(x_1, \ldots, x_p) \in \mathbb{R}^p$, and for almost all $x_0 \in \mathbb{R}$,

$$\left| \left[\frac{\partial^2 l}{\partial \eta_i \partial \eta_j}(x; \eta) + \frac{\partial l}{\partial \eta_i}(x; \eta) \frac{\partial l}{\partial \eta_j}(x; \eta) \right] f(x_0 - \phi_1 x_1 - \cdots - \phi_p x_p) \right| \le g_{i,j}(x), \quad \int_{\mathbb{R}} g_{i,j}(x) \mathrm{d}x_0 < \infty, \tag{43}$$

where $x = (x_0, \ldots, x_p)$. According to (33), all the second order partial derivatives of $l(x; \eta)$ with respect to η either take the form (36) or are sums of functions of type (36). According to (28), the same property is true for double products of first order partial derivatives of $l(x; \eta)$ with respect to η. Therefore, (41) implies (43), and (A4) is established. Simple algebra shows that Σ given by (18) is the inverse of V given by (34).

To prove Theorem 2(iii), we view Σ as a function g of η calculated at η_0. Since $\widetilde{\eta}_n \xrightarrow{a.s.} \eta_0$, $g(\widetilde{\eta}_n) \xrightarrow{a.s.} g(\eta_0)$ if g is continuous at η_0. This is indeed the case since m given by (6) is a continuous function of η in S, $c_{2,Z}$ given by (3) is a continuous function of (ϵ, σ^2) in $(-1, 1) \times (0, \infty)$, and for any given lag k, $m_{2,k}$ is a continuous function of $(c_{2,Z}, \phi)$ in $(0, \infty) \times C$. The last point follows from $m_{2,k} = \frac{c_{2,Z}}{\pi} \int_0^\pi \cos(k\nu) |\phi(\mathrm{e}^{-i\nu})|^{-2} \mathrm{d}\nu$ where $(\nu, \phi) \mapsto \cos(k\nu)|\phi(\mathrm{e}^{-i\nu})|^{-2}$ is continuous in $[0, \pi] \times C$ and every point in C admits a compact neighbourhood. Lastly, since for any fixed k, $\widehat{m}_{2,k} \xrightarrow{a.s.} m_{2,k}$, one may also replace M_2 by $[\widehat{m}_{2,i-j}]_{i,j=1}^p$ in (18) to estimate Σ strongly consistently.

References

[1] H. Tong, Nonlinear Time Series, in: Oxford Statistical Science Series, vol. 6, The Clarendon Press Oxford University Press, New York, 1990.
[2] J. Fan, Q. Yao, Nonlinear Time Series, in: Springer Series in Statistics, Springer-Verlag, New York, 2003, Nonparametric and parametric methods.
[3] G.K. Grunwald, R.J. Hyndman, L. Tedesco, R.L. Tweedie, Non-Gaussian conditional linear AR(1) models, Aust. N. Z. J. Stat. 42 (4) (2000) 479–495.
[4] P.A. Jacobs, P.A.W. Lewis, Stationary discrete autoregressive-moving average time series generated by mixtures, J. Time Ser. Anal. 4 (1) (1983) 19–36.
[5] J. Anděl, On AR(1) processes with exponential white noise, Comm. Statist. Theory Methods 17 (5) (1988) 1481–1495.
[6] W.K. Li, A.I. McLeod, ARMA modelling with non-Gaussian innovations, J. Time Ser. Anal. 9 (2) (1988) 155–168.
[7] M.L. Tiku, W.-K. Wong, G. Bian, Estimating parameters in autoregressive models in non-normal situations: Symmetric innovations, Comm. Statist. Theory Methods 28 (2) (1999) 315–341.
[8] M.L. Tiku, W.-K. Wong, D.C. Vaughan, G. Bian, Time series models in non-normal situations: Symmetric innovations, J. Time Ser. Anal. 21 (5) (2000) 571–596.
[9] M.L. Tiku, W.-K. Wong, G. Bian, Time series models with asymmetric innovations, Comm. Statist. Theory Methods 28 (6) (1999) 1331–1360.
[10] A. Akkaya, M.L. Tiku, Estimating parameters in autoregressive models in non-normal situations: Asymmetric innovations, Comm. Statist. Theory Methods 30 (3) (2001) 517–536.
[11] W.-K. Wong, G. Bian, Estimating parameters in autoregressive models with asymmetric innovations, Statist. Probab. Lett. 71 (1) (2005) 61–70.
[12] G.J. Janacek, A.L. Swift, A class of models for nonnormal time series, J. Time Ser. Anal. 11 (1) (1990) 19–31.
[13] M. Pourahmadi, Skew-normal ARMA models with nonlinear heteroscedastic predictors, Comm. Statist. Theory Methods 36 (2007) 1803–1819.
[14] A. Azzalini, A. Dalla Valle, The multivariate skew-normal distribution, Biometrika 83 (4) (1996) 715–726.
[15] A. Azzalini, A. Capitanio, Statistical applications of the multivariate skew normal distribution, J. R. Stat. Soc. Ser. B Stat. Methodol. 61 (3) (1999) 579–602.
[16] G. González-Farías, A. Domínguez-Molina, A.K. Gupta, Additive properties of skew normal random vectors, J. Statist. Plann. Inference 126 (2) (2004) 521–534.
[17] R.B. Arellano-Valle, A. Azzalini, On the unification of families of skew-normal distributions, Scand. J. Statist. 33 (3) (2006) 561–574.
[18] G.S. Mudholkar, A.D. Hutson, The epsilon-skew-normal distribution for analyzing near-normal data, J. Statist. Plann. Inference 83 (2) (2000) 291–309.
[19] A.D. Hutson, Utilizing the flexibility of the epsilon-skew-normal distribution for common regression problems, J. Appl. Stat. 31 (6) (2004) 673–683.
[20] A. Stuart, J.K. Ord, Kendall's Advanced Theory of Statistics. vol. 1 : Distribution Theory, sixth edition, Edward Arnold, London, 1994.
[21] D.R. Brillinger, Time Series. Data Analysis and Theory, Holt Rinehart and Winston Inc., New York, 1975.
[22] W.F. Stout, Almost Sure Convergence, Academic Press, New York, 1974.
[23] P.J. Brockwell, R.A. Davis, Time Series: Theory and Methods, second edition, Springer Verlag, New York, 1991.
[24] P.J. Brockwell, R.A. Davis, Introduction to Time Series and Forecasting, Springer Verlag, New York, 1996.
[25] L.A. Klimko, P.I. Nelson, On conditional least squares estimation for stochastic processes, Ann. Statist. 6 (1978).
[26] M. Taniguchi, Y. Kakizawa, Asymptotic Theory of Statistical Inference for Time Series, in: Springer Series in Statistics, Springer-Verlag, New York, 2000.
[27] P. Doukhan, J.C. Sifre, Agrégation de Mathématiques - Cours d'Analyse : Analyse Réelle et Intégration, Dunod, Paris, 2001.
[28] P. Billingsley, The Lindeberg–Lévy theorem for martingales, Proc. Amer. Math. Soc. 12 (1961) 788–792.
[29] I.A. Ibragimov, A central limit theorem for a class of dependent random variables, Theory Probab. Appl. 8 (1963) 83–89.