ARMA MODELLING WITH NON-GAUSSIAN INNOVATIONS

BY W. K. LI

University of Hong Kong

AND

A. I. MCLEOD

The University of Western Ontario

First version received August 1985; final version accepted May 1987

Abstract. The problem of modelling time series driven by non-Gaussian innovations is considered. The asymptotic normality of the maximum likelihood estimator is established under some general conditions. The distribution of the residual autocorrelations is also obtained. This gives rise to a potentially useful goodness-of-fit statistic. Applications of the results to two important cases are discussed. Two real examples are considered.

Keywords. Autoregressive moving-average process; maximum likelihood estimation; non-Gaussian innovations; residual autocorrelations.

1. INTRODUCTION

Time series processes driven by non-Gaussian innovations are common in real situations. For example, in economics, Nelson and Granger (1979) considered a set of 21 time series; among these, only six were found to be Gaussian. Usually such time series exhibit 'saw-tooth' behaviour. Some well-known examples are daily riverflow time series and various other geophysical time series such as the Wölfer sunspot series (Weiss, 1977; Yakowitz, 1973). Simple monotonic transformations will not rectify the asymmetry (Weiss, 1975). However, as shown in Figure 1, an autoregressive model with positively skewed innovations will mimic such asymmetric behaviour. Jacobs and Lewis (1978,a,b) and Lawrance and Lewis (1980) represent steps towards modelling this type of time series. In these papers, the authors considered the construction of models which have a predesignated marginal distribution. Alternatively, Tong and Lim (1980) consider a nonlinear approach. A direct approach would be to consider time series driven by innovations with a prespecified common distribution.

In hydrology it is well known that daily precipitation can usually be considered as gamma or log-normal distributed. Thus, some hydrologists, notably Quimpo (1967) and O'Connel and Jones (1979) have considered fitting autoregressive models to riverflow series with log-normal disturbances. Their method, however, is not maximum-likelihood; rather, the Yule–Walker equation is used

0143-9782/88/02 0155-14 $02.50/0
JOURNAL OF TIME SERIES ANALYSIS Vol. 9, No. 2

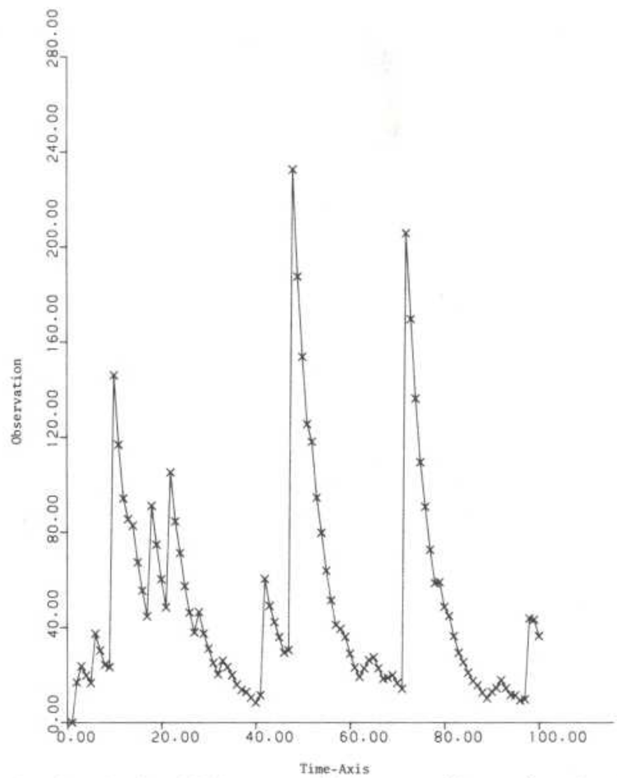

FIGURE 1. Sample path of an AR(1) process with log-normal innovations, $\phi = 0.8$, $n = 100$.

to obtain the autoregressive parameters. Since in many situations the maximum-likelihood estimators have many desirable properties, it is recommended that the maximum-likelihood procedure be used. It will be shown that under some very mild conditions on the distribution function of the innovations, maximum-likelihood estimators for autoregressive moving-average processes are always consistent and asymptotically normally distributed. Moreover, the asymptotic distribution of the residual autocorrelations will always be normal with mean $\mathbf{0}$ and the variance–covariance matrix will always assume a particular form. The above results will be applied to autoregressive models with log-normal and gamma innovations. Previously, Davies *et al.* (1980) discussed in detail the generation of symmetric and asymmetric time series from skewed innovations.

In practice, the distribution of the innovation series will not usually be known initially. However, since the standard least-squares methods are known to be Gaussian efficient (Whittle, 1962), initial model and parameter estimates may be obtained. Thus, the Gaussian assumption of the innovation series may be tested as in Granger and Newbold (1977, p. 314). Should the Gaussian assumption be rejected, then the new proposed techniques of this paper may be applied. Alternatively, one may consider the use of robust techniques. However, such procedures are designed mainly against symmetric departures from normality (Andrews *et* *al.*, 1972) and therefore differ from the situation here. The proposed procedure fits in well with the general framework for empirical modelling suggested by Box and Jenkins (1976). The results of this paper suggest that if, at the model-criticism stage, non-normal innovations are found, an improved model may be constructed by taking this into account in the model selection and calibration. Two real-life examples are given in section 6.

Previously, Basawa *et al.* (1976) considered the consistency and asymptotic normality of the maximum-likelihood estimator for stochastic processes under a different set of conditions. However, they considered neither problems in actual estimation nor problems in model diagnostic checking. Kabaila (1983) derived a lower bound on the asymptotic covariance matrix. Ledolter (1979) considered the sensitivity of ARIMA models to non-normal but symmetric error distributions. Klimko and Nelson (1978) obtained asymptotic results for a conditional least-squares approach.

2. PROPERTIES OF THE GENERAL MAXIMUM-LIKELIHOOD ESTIMATOR

Let X_t, $t = 1, \ldots, n$, be a stationary process satisfying

$$\phi(B)X_t = \theta(B)a_t, \tag{1}$$

where B is the backward-shift operator, $BX_t = X_{t-1}$, $\phi(B) = 1 - \phi_1 B - \cdots - \phi_p B^p$ and $\theta(B) = 1 - \theta_1 B - \cdots - \theta_q B^q$ and it is assumed that $\phi(B)$ and $\theta(B)$ satisfy the condition that all their roots lie outside the unit circle and have no common roots between them. The a_t are assumed to be independent and identically distributed with finite fourth moments and mean μ_a. The range of a_t is the open interval (a, b) where a and b may be infinite. Denote the probability density of a_t by $p_t = p(a_t \mid \alpha)$. Without loss of generality, the parameter α is assumed to be a scalar. Denote expected values by $\langle \cdot \rangle$. Define $\boldsymbol{\eta}^{\mathrm{T}} = (\phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q) = (\eta_1, \ldots, \eta_r)$, where $r = p + q$. Let $\boldsymbol{\lambda}^{\mathrm{T}} = (\boldsymbol{\eta}^{\mathrm{T}}, \alpha)$. It is assumed that p_t satisfies the following assumptions (Kendall and Stuart, 1961, p. 43).

ASSUMPTION 1. The derivatives

$$\frac{\partial \log p_t}{\partial \alpha}, \quad \frac{\partial \log p_t}{\partial a_t}, \quad \frac{\partial^2 \log p_t}{\partial \alpha^2}, \quad \frac{\partial^2 \log p_t}{\partial a_t\, \partial \alpha}, \quad \frac{\partial^2 \log p_t}{\partial a_t^2}$$

exist, and are continuous for almost all a_t and α in an open interval A of α including the true value.

ASSUMPTION 2. At the true values of $\boldsymbol{\lambda}$

$$\left\langle \frac{1}{p_t} \frac{\partial p_t}{\partial \alpha} \right\rangle = \left\langle \frac{1}{p_t} \frac{\partial^2 p_t}{\partial \alpha^2} \right\rangle = 0$$

and

$$\left\langle \frac{1}{p_t^2} \left(\frac{\partial p_t}{\partial \alpha} \right)^2 \right\rangle > 0,$$

and similarly

$$\left\langle \frac{1}{p_t} \frac{\partial^2 p_t}{\partial \alpha \, \partial a_t} \right\rangle = 0, \qquad 0 < \left\langle \left(\frac{1}{p_t} \frac{\partial p_t}{\partial a_t} \right)^2 \right\rangle < \infty.$$

ASSUMPTION 3.

$$\lim_{a_t \to a} p_t = \lim_{a_t \to b} p_t, \qquad \text{similarly for } \frac{\partial p_t}{\partial a_t}.$$

It is clear that assumption 1 implies the existence of derivatives of p_t with respect to η_i. Given n successive observations X_t, $t = 1, \ldots, n$, the log-likelihood conditional on the first p observations is defined by

$$L = \log \prod_{t=p+1}^{n} p(a_t \mid \alpha) = \sum \log p(a_t \mid \alpha).$$

It is assumed that $a_t = \mu_a$, for $t = 1, \ldots, p$. Denote the maximum-likelihood estimator of $\boldsymbol{\eta}^{\mathrm{T}}$ and α by $\hat{\boldsymbol{\eta}}$ and $\hat{\alpha}$ respectively. Define as in McLeod (1978) auxiliary processes u_t and v_t by

$$\phi(B)u_t = -a_t, \qquad \theta(B)v_t = a_t.$$

Note that from assumption 2, $\langle -\partial^2 L/\partial\alpha \cdot \partial\eta_i \rangle = \langle \partial L/\partial\alpha \cdot \partial L/\partial\eta_i \rangle = \langle \partial L/\partial\alpha \cdot \partial L/\partial a_t \cdot m_{t-i} \rangle$, where $m_{t-i} = v_{t-i}$ or u_{t-j} depending on $\eta_i = \theta_i$ or ϕ_j.

THEOREM. *Under assumptions 1–3 the conditional maximum-likelihood estimator $\hat{\boldsymbol{\lambda}}$ of $\boldsymbol{\lambda}$ exists and is consistent. Furthermore, $\sqrt{n}(\hat{\boldsymbol{\lambda}} - \boldsymbol{\lambda})$ is asymptotically normally distributed and the (i, j)th element of the Fisher information matrix per observation $\boldsymbol{I}$ of $\boldsymbol{\lambda}$ is given by*

$$I_{ij} = \left\langle \left(\frac{\partial p_t}{\partial a_t} \right)^2 \Big/ p_t^2 \right\rangle \langle \gamma_{\eta\eta}(j - i) + \mu_a^2/M \rangle, \qquad \text{if } 1 \leqslant i, j \leqslant r;$$

$$I_{ij} = \left\langle \frac{1}{p_t^2} \cdot \frac{\partial p_t}{\partial \alpha} \cdot \frac{\partial p_t}{\partial a_t} \cdot m_{t-h} \right\rangle, \qquad \begin{array}{l} m_{t-h} \text{ depends on } i \text{ if } 1 \leqslant i \leqslant r, j = r + 1 \\ m_{t-h} \text{ depends on } j \text{ if } 1 \leqslant j \leqslant r, i = r + 1; \end{array}$$

$$I_{ij} = \left\langle \frac{1}{p_t^2} \left(\frac{\partial p_t}{\partial \alpha} \right)^2 \right\rangle, \qquad \text{if } i = j = r + 1;$$

where $\gamma_{\eta\eta}(k)$ is the lag k cross-covariance of u_t, v_t or u_t and v_t depending on η_i and η_j. Similarly, M is either $\phi(1)^2$, $\theta(1)^2$ or $\phi(1)\theta(1)$ depending on η_i and η_j.

For convenience of presentation, the proof of the theorem is relegated to appendix 1.

3. DISTRIBUTION OF THE RESIDUAL AUTOCORRELATIONS

We remark that the fourth moments of $\{a_t\}$ are assumed to be finite. Define the lth innovation autocorrelation by

$$r_a(l) = C_a(l)/C_a(0),$$

where $C_a(l)$ is the lth innovation autocovariance,

$$C_a(l) = \sum (a_t - \mu_a)(a_{t-l} - \mu_a)/n.$$

Then it can be shown that

$$r_a(l) = \bar{C}_a(l)/\bar{C}_a(0) + O_p(n^{-1}),$$

where

$$\bar{C}_a(l) = n^{-1} \sum (a_t - \bar{\mu}_a)(a_{t-l} - \bar{\mu}_a) + O_p(n^{-1}),$$

where $\bar{\mu}_a = \sum a_t/n$.

Similarly, it may be shown that

$$r_a(l) = C_a(l)/\langle C_a(0)\rangle + O_p(n^{-1}).$$

Thus, $r_a(l)$, $C_a(l)/\langle C_a(0)\rangle$ and $\bar{C}_a(l)/\bar{C}_a(0)$ will have the same asymptotic distribution. As in McLeod and Li (1983), from theorem 14 of Hannan (1970, p. 228) $\sqrt{n}\cdot \boldsymbol{r}^{\mathrm{T}} = \sqrt{n}(r_a(1), \ldots, r_a(m))$ is multivariate normal with covariance matrix $\mathbf{1}_m$.

The asymptotic cross-covariance of $\partial L/\partial\phi_i$ and $C_a(l)$ is obtained by noting

$$\int C_a(l) \prod_{t=p+1}^{n} p_t \, da_{p+1} \cdots a_n \cong 0. \tag{2}$$

Now $\prod_{t=p+1}^{n} p_t$ is really the likelihood of $X_{p+1}, \ldots, X_n$ conditional on $X_1, \ldots, X_p$ and is therefore dependent on α and $\boldsymbol{\eta}^{\mathrm{T}} = (\phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q)$. Differentiating equation (2) with respect to ϕ_i and interchanging the differential and integral sign gives

$$\begin{aligned}
&\frac{1}{n}\int \frac{\partial}{\partial\phi_i}\left(C_a(l) \prod p_t\right) da_{p+1} \cdots a_n \\
&\qquad = \frac{1}{n}\int \sum (a_t - \mu_a) u_{t-l-i} \prod p_t - \frac{1}{n}\int \sum u_{t-i}(a_{t-l} - \mu_a) \prod p_t \\
&\qquad\quad + \frac{1}{n}\int \sum a_t(a_{t-l} - \mu_a) \frac{\partial \prod p_t}{\partial\phi_i} \\
&= 0.
\end{aligned}$$

Note that the cumulant term is zero (Brillinger, 1975, p. 19). Now, the first term can be shown to be zero; the second term can be shown, as in McLeod (1978), to be equal to $\psi'_{l-i}\cdot\langle C_a(0)\rangle$ if $i \leqslant l$, where $\sum_{i=0}^{\infty} \psi'_i = 1/\phi(B)$; and the last term can be rewritten as

$$\frac{1}{n}\sum \int a_t(a_{t-l}-\mu_a)\frac{\partial L}{\partial \phi_i}\prod p_t\, da_{p+1}\cdots a_n.$$

Thus, the asymptotic cross-covariance of $\sqrt{n}\cdot r_a(l)$ and $n^{-1/2}\,\partial L/\partial\phi_i$ is just equal to ψ'_{l-i}. Similar results hold for $n^{-1/2}\,\partial L/\partial\theta_k$. Furthermore, it can be seen that $n\cdot\langle r\cdot\partial L/\partial\alpha\rangle = 0$, so that asymptotically the marginal distribution of $\sqrt{n}[(\hat{\boldsymbol{\eta}}-\boldsymbol{\eta})^{\mathrm{T}}, \boldsymbol{r}^{\mathrm{T}}]^{\mathrm{T}}$ is normal with mean $\mathbf{0}$ and covariance matrix

$$\begin{bmatrix} \boldsymbol{J} & \boldsymbol{J}\boldsymbol{X}^{\mathrm{T}} \\ \boldsymbol{X}\boldsymbol{J} & \mathbf{1}_m \end{bmatrix},$$

where $\boldsymbol{J}$ is the asymptotic covariance matrix of $\hat{\boldsymbol{\eta}}$ from section 2; $\boldsymbol{X}$ is given by $(\phi'_{i-j}\,|\,\theta'_{l-k})_{m\times(p+q)}$, $\sum_{j=0}^{\infty}\theta'_j = 1/\theta(B)$, and $\sum_{i=0}^{\infty}\psi'_i = 1/\phi(B)$.

The residual autocorrelations are defined by

$$\hat{r}(l) = \frac{\sum(\hat{a}_t-\hat{\bar{\mu}}_a)(\hat{a}_{t-l}-\hat{\bar{\mu}}_a)}{\sum(\hat{a}_t-\hat{\bar{\mu}}_a)^2},$$

where $\hat{\bar{\mu}}_a = \sum \hat{a}_t/n$.

By Taylor series expansion about $\hat{\boldsymbol{\lambda}}$ and evaluation at the true values of $\boldsymbol{\lambda}$, $\hat{\boldsymbol{r}}^{\mathrm{T}} = [\hat{r}(1), \ldots, \hat{r}(m)]$, can be shown (McLeod, 1978) to be

$$\hat{\boldsymbol{r}} = \boldsymbol{r} - \boldsymbol{X}(\hat{\boldsymbol{\eta}}-\boldsymbol{\eta}) + \mathrm{O}_p(n^{-1}). \tag{3}$$

By the martingale central limit theorem (Billingsley, 1961), any linear combination of $\sqrt{n}\cdot\hat{\boldsymbol{r}}$ is asymptotically normal, and it follows from equation (3) that $\sqrt{n}\cdot\hat{\boldsymbol{r}}$ is asymptotically multivariate normally distributed with mean $\mathbf{0}$ and covariance matrix

$$\mathbf{1}_m - \boldsymbol{X}\boldsymbol{J}\boldsymbol{X}^{\mathrm{T}}. \tag{4}$$

Thus, if an autoregressive moving-average process is driven by innovations having finite fourth-order moments and probability densities satisfying assumptions 1 to 3, and the model parameters are estimated by the maximum-likelihood method, the resulting residual autocorrelations will be asymptotically normally distributed with covariance matrix given by (4). The result (4) reduces the problem of the asymptotic distribution of the residual autocorrelations (in the general situation) to the simpler problem of evaluating the quantities $\boldsymbol{X}$ and $\boldsymbol{J}$. Note that equation (4) is not necessarily idempotent as in the Gaussian situation. Nevertheless, a portmanteau statistic can still be defined by

$$Q_m = \hat{\boldsymbol{r}}^{\mathrm{T}}(\mathbf{1}_m - \hat{\boldsymbol{X}}\hat{\boldsymbol{J}}\hat{\boldsymbol{X}}^{\mathrm{T}})^{-1}\hat{\boldsymbol{r}}$$

where $\hat{\boldsymbol{X}}$ and $\hat{\boldsymbol{J}}$ are the estimates of $\boldsymbol{X}$ and $\boldsymbol{J}$ when $\boldsymbol{\lambda}$ is replaced by $\hat{\boldsymbol{\lambda}}$. Q_m would be approximately asymptotically chi-square distributed with m degrees of freedom.

4. APPLICATIONS TO THE LOG-NORMAL AND GAMMA INNOVATION SITUATIONS

For simplicity, consider the autoregressive process

$$\phi(B)X_t = a_t,$$

where the a_t are log-normal distributed with probability densities

$$p(a_t) = [a_t\sqrt{2}\pi\sigma]^{-1} \exp[-\tfrac{1}{2}(\log a_t)^2/\sigma^2].$$

The moments of a_t about 0 are given by

$$\mu_r' = \exp(\tfrac{1}{2}\sigma^2 r^2).$$

The log-likelihood L conditional on the first p observations is

$$L = \text{constant} - \sum_{t=p+1}^{n} \log a_t - \frac{(n-p)}{2} \log \sigma^2 - \frac{1}{2} \sum_{t=p+1}^{n} (\log a_t)^2/\sigma^2. \qquad (5)$$

When a_t is log-normal distributed with $\langle \log a_t \rangle = 0$ and $\langle (\log a_t)^2 \rangle = \sigma^2$, assumptions 1 to 3 of section 2 are satisfied and the information matrix of $\boldsymbol{\eta}$ is given by

$$(I_{kj}) = \exp 2\sigma^2(1 + \sigma^{-2})[\gamma(k-j) + \mu_a^2/\phi(1)^2],$$

where $\mu_a^2 = \exp \sigma^2$. See Li (1981) for a more detailed discussion.

Note that the maximum-likelihood estimator for σ^2 is simply $\sum (\log a_t)^2/n$; thus, after maximizing over σ^2, the log-likelihood (5) can be written

$$L_{(\max)} = \text{constant} - \sum_{p+1}^{n} \log a_t - \frac{(n-p)}{2} \log(\textstyle\sum (\log a_t)^2/n).$$

A non-linear optimization algorithm can then be used to find the maximum-likelihood estimate $\hat{\boldsymbol{\eta}}$. The two-parameter situation presents no additional problem, but the three-parameter log-normal situation is much more difficult. However, Hill (1963) has suggested maximum-likelihood estimates which may be useful in this situation.

As an example, consider the ARMA(1, 0) process

$$(1 - \phi B)Z_t = a_t,$$

where $\log a_t$ is $N(0, 1)$. Then straightforward calculation yields

$$I = \begin{bmatrix} a & -e/(1-\phi) \\ -e/(1-\phi) & \frac{1}{2} \end{bmatrix}, \qquad (6)$$

where

$$a = \left(\frac{e(e-1)}{1-\phi^2} + \frac{e}{(1-\phi)^2}\right)2e^2.$$

TABLE I

EMPIRICAL RELATIVE EFFICIENCY OF GAUSSIAN ESTIMATOR TO MAXIMUM-LIKELIHOOD FOR ϕ_1

n	ϕ_1	Empirical mean Gaussian estimator	Empirical variance Gaussian estimator	Empirical mean MLE	Empirical variance MLE	Theoretical variance MLE	Empirical relative efficiency	Theoretical relative efficiency
50	0.8	0.72892	0.01190	0.80077	0.00007	0.000025	0.00582	0.00348
	0.4	0.36235	0.01408	0.40308	0.00049	0.000131	0.03486	0.00779
	0	−0.02277	0.01728	0.00603	0.00095	0.000212	0.05482	0.01060
200	0.8	0.78339	0.00214	0.80043	0.00001	0.000006	0.00425	0·00348
	0.4	0.39364	0.00584	0.40199	0.00006	0.000033	0.00987	0.00779
	0	−0.00076	0.00535	−0.00176	0.00005	0.000053	0.00947	0.01060

This implies that the asymptotic variance of $\hat{r}(k)$, $k > 0$, is

$$\frac{1}{n}[1 - \tfrac{1}{2}\Delta^{-1}\phi^{2(k-1)}],$$

where Δ is the determinant of (6). Hence the asymptotic variance for $\hat{r}(k)$ should be much closer to $1/n$ than in the corresponding Gaussian situation.

Simulation experiments have been performed to compare the performances of the conditional Gaussian estimator and the approximate maximum-likelihood estimator, in the first-order autoregressive case, for sample sizes $n = 50$, and 200, when the innovations are log-normal distributed with $\sigma = 1$. The values of ϕ_1 used are 0, 0.4, 0.8. There are 1000 replications for each combination of values of ϕ_1 and n. The random number generator Super-Duper (Marsaglia, 1976) is used with the method of Box and Muller (1958), to generate normal variates which are then exponentiated to give the log-normal innovations. The simulation results are summarized in Table I. It can be seen that the Gaussian estimator may be heavily downward-biased for small sample sizes and large ϕ_1 values.

It may also be seen that the maximum-likelihood estimator performs much better for this range of values of n, although values of the empirical variance of the maximum likelihood estimator are somewhat larger than those obtained from theoretical calculations. However, it appears that these discrepancies are getting smaller as n increases. In fact, when $n = 200$ and $\phi_1 = 0$, the theoretical and empirical variances of the maximum-likelihood estimator are virtually identical. The general pattern of the simulation results confirms that the maximum-likelihood estimator is greatly superior to the Gaussian estimator. On the other hand, even in the Gaussian situation, McLeod (1974, p. 79) has demonstrated that there may be great discrepancies between the theoretical and empirical variances of the estimated parameters. Simulation experiments have also been performed to compare the asymptotic variance and the sampling variance of $\hat{r}(1)$ about zero for ARMA(1, 0) models, when $\phi_1 = 0$, 0.2, 0.4, 0.6, and 0.8 and σ equals to 1. The length of each series is 200 and the number of replications for each value of ϕ_1 is 500. The IMSL subroutine GGNLG was used to generate the log-normal variates. The results are summarized in Table II. Values inside the

TABLE II

EMPIRICAL VARIANCE OF $\hat{r}(1)$ FOR AUTOREGRESSIVE PROCESS OF ORDER 1. LOG-NORMAL INNOVATIONS

ϕ_1	Theoretical variance of $\hat{r}(1)$	Empirical variance of $\hat{r}(1)$
0	0.0049	0.0046 ($\pm$0.0007)
0.2	0.0049	0.0045 ($\pm$0.0007)
0.4	0.0050	0.0043 ($\pm$0.0007)
0.6	0.0050	0.0043 ($\pm$0.0006)
0.8	0.0050	0.0045 ($\pm$0.0007)

Series length = 200
Number of replications = 500
$\sigma^2 = 1$

bracket are two times the standard error of the empirical variance of $\hat{r}(1)$. It can be seen that the empirical variances are closer to the predicted values if ϕ is small. As a whole the empirical variances are somewhat downward-biased. On the other hand, most predicted values are within the 95% confidence limits.

Now suppose that the innovations a_t in

$$\phi(B)X_t = a_t$$

are gamma-distributed with probability densities given by

$$p(a_t) = \frac{a_t^{\alpha-1}\exp(-a_t)}{\Gamma(\alpha)}, \qquad a_t > 0.$$

Then, if $\alpha > 2$, it can be shown that the conditions in section 2 are all satisfied. In this case the log-likelihood L conditional on the first p observations is given by

$$L = \sum_{t=p+1}^{n} (\alpha - 1)\log a_t - \sum_{t=p+1}^{n} a_t - (n-p)\log\Gamma(\alpha).$$

As in the log-normal case the (j, k)th element in the information matrix for $\boldsymbol{\eta}$ is given by

$$\begin{aligned}\frac{1}{n}\left\langle\frac{\partial^2 L}{\partial\phi_j\,\partial\phi_k}\right\rangle &= -\left(\gamma(k-j) + \frac{\alpha^2}{\phi(1)^2}\right)\cdot\int_0^\infty \frac{\alpha-1}{a_t^2}\,\frac{a_t^{\alpha-1}\exp(-a_t)}{\Gamma(\alpha)}\,da_t \\ &= -\left(\gamma(k-j) + \frac{\alpha^2}{\phi(1)^2}\right)\Big/(\alpha-2) \qquad (7)\end{aligned}$$

where

$$\gamma(k-j) = \left\langle\left(X_{t-j} - \frac{\mu_a}{\phi(1)}\right)\left(X_{t-k} - \frac{\mu_a}{\phi(1)}\right)\right\rangle.$$

Similar results can be obtained for other derivatives. It may be noted that if $\alpha \leqslant 1$, then assumption 3 is not satisfied and if $\alpha \leqslant 2$, then the integral in (7) does not exist.

Evaluation of the likelihood for the ARMA parameters $\dot{\boldsymbol{\eta}}$ can then be obtained using $\dot{\alpha} = \sum \dot{a}_t/n$. Estimation becomes more difficult if the distribution of a_t is given by the three-parameter gamma,

$$p(a_t) = \frac{(a_t - \gamma)^{\alpha-1} \exp[-(a_t - \gamma)/\beta]}{\beta^\alpha \Gamma(\alpha)},$$

where $\alpha > 0$, $\beta > 0$, $a_t > \gamma$. When γ is known, however, the solution to the maximum-likelihood equation is always possible and $\boldsymbol{\eta} = \dot{\boldsymbol{\eta}}$ is given by substituting $\beta = A/\alpha$ into

$$y = \log \alpha - \psi(\alpha),$$

where $\psi(\alpha) = d \log \Gamma(\alpha)/d\alpha$, $y = \log A - \log G$, where A is the arithmetic mean and G the geometric mean of $a_t - \gamma$, respectively. Note that if γ is unknown it may be estimated by the smallest observation.

Greenwood and Durand (1960) have provided tables of $y\alpha$ as a function of y and from these tables values of α can be obtained from interpolation. Choi and Wette (1968) suggested using a series expansion of $\psi(\alpha)$ and $\psi'(\alpha)$. However, we have obtained satisfactory estimates of α using the series expansion of $1/\Gamma(\alpha)$ as given by Abramowitz and Stegun (1965, p. 256).

5. TWO APPLICATIONS

As a demonstration of the potential of the non-Gaussian innovation approach, we consider fitting such models to the Wölfer yearly sunspot numbers and the Canadian lynx data. It has been pretty well known that both data sets exhibit non-Gaussian or perhaps non-linear behaviour (see, for example, Tong, 1983), and various authors have proposed different models for both. We consider first the sunspot series from 1770 to 1955. The best linear Gaussian model for the sunspot series is that of an AR(9) with an estimated residual variance of 199.27. Tong and Lim (1980) considered a threshold model that gives a residual variance of 153.71. Gabr and Subba Rao (1981) considered a subset bilinear model with an estimated residual variance of 124.33. Here we consider the simple AR(2) model which has been considered by Box and Jenkins (1976). A two-parameter log-normal distribution was considered for the innovations. Powell's (1964) conjugate direction algorithm was used to obtain the conditional maximum-likelihood estimates. Approximate Gaussian estimates of ϕ_1 and ϕ_2 were used as initial values. The resulting model is

$$X_t = 1.6759X_{t-1} - 0.7840X_{t-2} + a_t,$$

where a_t is log-normal distributed with estimated mean 13.88 and variance 153.39. The residual variance is obtained using the method of Finney (Johnson and Kotz, 1970) as the usual product moment estimate can be very inefficient.

The residual variance of 153.39 is comparable to that of the threshold model, although it is still greater than that of the subset bilinear model. The AR(2) model is certainly more parsimonious.

Next, we consider the Canadian lynx data from 1821 to 1934. Moran (1953) considered an AR(2) model for the $\log_{10}$-transformed data set with a residual variance of 0.0459. The Yule–Walker equations were used to estimate the autoregressive parameters. Nicholls and Quinn (1982) considered a random coefficient autoregressive model of order 2 that gives a residual variance of 0.0391. Here an AR(2) model with two-parameter gamma innovations is considered. As in other works, the $\log_{10}$-transformation is applied first to the original data. Denote the transformed observations by X_t and let $X'_t = X_t - 1.591$ where 1.591 is just slightly smaller than the smallest value $\log_{10}(39)$ of X_t. The resulting model is

$$X'_t = 1.4771X'_{t-1} - 0.6210X'_{t-2} + a_t,$$

where a_t is distributed as a two-parameter gamma with estimated shape and scale parameters equal to 2.658 and 0.1183 respectively. These values together give an estimated residual variance of 0.0372. This is a 5% reduction of residual variance over the random coefficient model.

Although our experience with real data is still quite limited, the above two examples demonstrate that linear time series model with non-Gaussian innovation can be a useful tool in time series modelling.

6. CONCLUSIONS

Non-Gaussian time series are important in many applications. It has been shown that under mild restrictions on the probability density of the innovation series, conditional maximum-likelihood estimators for autoregressive moving-average parameters are consistent and asymptotically normal. The asymptotic distribution of the residual autocorrelations under this general situation is also obtained. This is found to assume a particularly simple form. Simulation experiments also indicate the superiority of the maximum-likelihood estimator to the Gaussian estimator. As real examples, autoregressive models with log-normal and gamma innovations are fitted to the sunspot and the Canadian lynx data respectively. The results of this paper answer, to a certain extent, the question raised by Granger (1979) on non-Gaussian time series modelling. It is believed that these results will contribute to this important area of time series analysis.

ACKNOWLEDGEMENTS

The authors are grateful to Drs K. W. Hipel, E. McKenzie, I. B. MacNeill and G. Tunnicliffe Wilson, and to a referee for helpful comments. This research was supported by a NSERC Grant.

APPENDIX I

We now present a proof of the theorem stated in section 2.

Denote by $\dot{\eta}_i$ and $\dot{\alpha}$ parameter values of η_i and α satisfying assumptions 1 to 3, of the model (1) in a closed interval R consisting the true parameters in its interior. Denote the corresponding a_t by $\dot{a}_t$ and similarly denote $\dot{u}_t$ and $\dot{v}_t$.

The first-order derivatives of L evaluated at $\dot{\boldsymbol{\eta}}$ and $\dot{\alpha}$ are given by

$$\frac{\partial \dot{L}}{\partial \phi_i} = \sum \frac{1}{\dot{p}_t} \frac{\partial \dot{p}_t}{\partial a_t} \dot{u}_{t-i}, \qquad \frac{\partial \dot{L}}{\partial \theta_i} = -\sum \frac{1}{\dot{p}_t} \frac{\partial \dot{p}_t}{\partial a_t} \dot{v}_{t-i}, \qquad \frac{\partial \dot{L}}{\partial \alpha} = \sum \frac{1}{\dot{p}_t} \frac{\partial \dot{p}_t}{\partial \alpha}.$$

The second-order derivatives of L are given by

$$\frac{\partial^2 \dot{L}}{\partial \alpha^2} = \sum \frac{1}{\dot{p}_t} \frac{\partial^2 \dot{p}_t}{\partial \alpha^2} + \sum \frac{-1}{\dot{p}_t^2} \left(\frac{\partial \dot{p}_t}{\partial \alpha}\right)^2,$$

$$\frac{\partial^2 \dot{L}}{\partial \alpha\, \partial \phi_i} = -\sum \frac{1}{\dot{p}_t^2} \frac{\partial \dot{p}_t}{\partial \alpha} \frac{\partial \dot{p}_t}{\partial a_t} \dot{u}_{t-i} + \sum \frac{1}{\dot{p}_t} \frac{\partial^2 \dot{p}_t}{\partial \alpha\, \partial a_t} \dot{u}_{t-i},$$

$$\frac{\partial^2 \dot{L}}{\partial \alpha\, \partial \theta_k} = \sum \frac{1}{\dot{p}_t^2} \frac{\partial \dot{p}_t}{\partial \alpha} \frac{\partial \dot{p}_t}{\partial a_t} \dot{v}_{t-k} - \sum \frac{1}{\dot{p}_t} \frac{\partial^2 \dot{p}_t}{\partial \alpha\, \partial a_t} \dot{v}_{t-k},$$

$$\frac{\partial^2 \dot{L}}{\partial \phi_j\, \partial \phi_i} = \sum \left[\frac{1}{\dot{p}_t} \frac{\partial^2 \dot{p}_t}{\partial a_t^2} - \frac{1}{\dot{p}_t^2} \left(\frac{\partial \dot{p}_t}{\partial a_t}\right)^2\right] \dot{u}_{t-i} \dot{u}_{t-j},$$

$$\frac{\partial^2 \dot{L}}{\partial \phi_i\, \partial \theta_k} = \sum \left[\frac{1}{\dot{p}_t^2} \frac{\partial \dot{p}_t}{\partial a_t} \frac{\partial \dot{p}_t}{\partial a_t} - \frac{1}{\dot{p}_t} \frac{\partial^2 \dot{p}_t}{\partial a_t^2}\right] \dot{v}_{t-k} \dot{u}_{t-i} - \sum \frac{1}{\dot{p}_t} \frac{\partial \dot{p}_t}{\partial a_t} \frac{\partial^2 a_t}{\partial \phi_i\, \partial \theta_k},$$

$$\frac{\partial^2 \dot{L}}{\partial \theta_k\, \partial \theta_l} = \sum \left[\frac{1}{\dot{p}_t} \frac{\partial^2 \dot{p}_t}{\partial a_t^2} - \frac{1}{\dot{p}_t^2} \left(\frac{\partial \dot{p}_t}{\partial a_t}\right)^2\right] \dot{v}_{t-k} \dot{v}_{t-l}.$$

The above equations follow from the fact that

$$\frac{\partial^2 a_t}{\partial \phi_i\, \partial \phi_j} = \frac{\partial^2 a_t}{\partial \theta_k\, \partial \theta_l} = 0.$$

It can be seen that by assumption 3 that

$$\left\langle \frac{1}{p_t} \frac{\partial p_t}{\partial a_t} \right\rangle = \int \frac{1}{p_t} \frac{\partial p_t}{\partial a_t} p_t\, da_t = \lim_{a_t \to b} p(a_t \mid \alpha) - \lim_{a_t \to b} p(a_t \mid \alpha) = 0.$$

Similarly, it can be shown that $\langle (\partial^2 p_t / \partial a_t^2)/p_t \rangle = 0$. Thus the quantities $W_t = (\partial p_t / \partial a_t \cdot u_{t-j})/p_t$ are uncorrelated over t and this implies that $n^{-1}\, \partial L/\partial \phi_i$ and $n^{-1}\, \partial L/\partial \theta_i$ converge to zero in probability. Since $\{v_t\}$, $\{u_t\}$ are stationary and $\{p_t\}$ is independent,

$$-\lim_{n\to\infty} \frac{\partial^2 L}{\partial \eta_i\, \partial \eta_j} \bigg/ n \;\xrightarrow{p}\; \left\langle \frac{1}{p_t^2} \left(\frac{\partial p_t}{\partial a_t}\right)^2 m_{t-i} m_{t-j} \right\rangle,$$

$$-\lim_{n\to\infty} \frac{\partial^2 L}{\partial \alpha\, \partial \eta_i} \bigg/ n \;\xrightarrow{p}\; \left\langle \frac{1}{p_t^2} \frac{\partial p_t}{\partial \alpha} \frac{\partial p_t}{\partial a_t} m_{t-i} \right\rangle,$$

$$-\lim_{n\to\infty} \frac{\partial^2 L}{\partial \alpha^2} \bigg/ n \;\xrightarrow{p}\; \left\langle \frac{1}{p_t^2} \left(\frac{\partial p_t}{\partial \alpha}\right)^2 \right\rangle,$$

where $m_{t-i} = v_{t-i}$ or u_{t-i} depending on $\eta_i = \phi_i$ or θ_i. A similar result holds for derivatives of L with respect to α.

Now, expanding $n^{-1} \cdot \partial L/\partial \boldsymbol{\lambda}$ about the true parameter $\boldsymbol{\lambda}$ and evaluating at $\dot{\boldsymbol{\lambda}}$ gives

$$n^{-1} \cdot \partial L/\partial \boldsymbol{\lambda} \mid \dot{\boldsymbol{\lambda}} = n^{-1} \cdot \partial L/\partial \boldsymbol{\lambda} \mid \boldsymbol{\lambda} + n^{-1} \cdot \partial^2 L/\partial \boldsymbol{\lambda}\, \partial \boldsymbol{\lambda}^{\mathrm{T}} \mid \boldsymbol{\lambda}^* (\dot{\boldsymbol{\lambda}} - \boldsymbol{\lambda}),$$

where $\boldsymbol{\lambda}^*$ lies between $\boldsymbol{\lambda}$ and $\dot{\boldsymbol{\lambda}}$. By assumption 2, the first term converges stochastically to zero with variance equal to $-n^{-1}$ (expected value of $n^{-1} \cdot \partial L/\partial \lambda\, \partial \lambda^{\mathrm{T}}$). Thus, as in Crowder (1976), the conditional maximum-likelihood estimator $\hat{\boldsymbol{\lambda}}$ of $\boldsymbol{\lambda}$ exists and can be shown to be consistent. Furthermore, by the martingale central limit theorem (Billingsley, 1961) $\sqrt{n}(\hat{\boldsymbol{\lambda}} - \boldsymbol{\lambda})$ is asymptotically normally distributed. The Fisher's information per observation of $\boldsymbol{\lambda}$ is given by $\boldsymbol{I} = (I_{ij})$ where

$$I_{ij} = -\lim_{n\to\infty} n^{-1}\frac{\partial^2 L}{\partial\eta_i\,\partial\eta_j} = \left\langle\left(\frac{\partial p_t}{\partial a_t}\right)^2\Big/p_t^2\right\rangle\cdot\left\langle\gamma_{\eta\eta}(j-i)+\frac{\mu_a^2}{M}\right\rangle \qquad \text{if } 1\leqslant i, j\leqslant r;$$

$$I_{ij} = \left\langle\frac{1}{p_t^2}\cdot\frac{\partial p_t}{\partial\alpha}\cdot\frac{\partial p_t}{\partial a_t}\cdot m_{t-h}\right\rangle, \qquad \begin{array}{l} m_{t-h} \text{ depends on } i \text{ if } 1\leqslant i\leqslant r, j=r+1\\ m_{t-h} \text{ depends on } j \text{ if } 1\leqslant j\leqslant r, i=r+1;\end{array}$$

$$I_{ij} = \left\langle\frac{1}{p_t^2}\left(\frac{\partial p_t}{\partial\alpha}\right)^2\right\rangle, \qquad \text{if } i=j=r+1;$$

where $\gamma_{\eta\eta}(k)$ is the lag k cross-covariance of u_t, v_t or u_t and v_t depending on η_i and η_j. Similarly, M is either $\phi(1)^2$, $\theta(1)^2$ or $\phi(1)\theta(1)$ depending on η_i and η_j.

REFERENCES

ABRAMOWITZ, M. and STEGUN, I. A. (1965) *Handbook of Mathematical Functions*. New York: Dover.

ANDREWS, D. F., BICKEL, P. J., HAMPEL, F. R., HUBER, P. J., ROGERS, W. H. and TUKEY, J. W. (1972) *Robust Estimates of Location: Survey and Advances*. New Jersey: Princeton University Press.

BASAWA, I. V., FEIGN, P. D. and HEYDE, C. C. (1976) Asymptotic properties of maximum likelihood estimators for stochastic processes. *Sankyha A* 38, 259–70.

BILLINGSLEY, P. (1961) The Lindeberg–Levy Theorem for martingales. *Proc. Amer. Math. Soc.* 12, 788–92.

BOX, G. E. P. and JENKINS, G. M. (1976) *Time Series Analysis, Forecasting and Control* (2nd edn). San Francisco: Holden-Day.

BOX, G. E. P. and MULLER, M. E. (1958) A note on generation of normal variates. *Ann. Math. Statist.* 28, 610–11.

BRILLINGER, D. R. (1975) *Time Series Data Analysis and Theory*. New York: Holt, Rinehart and Winston.

CHOI, S. C. and WETTE, R. (1968) Maximum likelihood estimation of the parameters of the gamma distribution and their bias. *Technometrics* 11, 683–90.

CROWDER, M. J. (1976) Maximum likelihood estimation for dependent observations. *J. Roy. Statist. Soc. Ser. B* 38, 45–53.

DAVIES, N., SPEDDING, T. and WATSON, W. (1980) Autoregressive moving average processes with non-normal residuals. *J. Time Series Anal.* 1, 103–9.

GABR, M. M. and SUBBA RAO, T. (1981) The estimation and prediction of subset bilinear time series models with application. *J. Time Series Anal.* 2, 155–71.

GRANGER, C. W. J. (1979) Near-normality and some econometric models. *Econometrica* 47, 781–4.

GRANGER, C. W. J. and NEWBOLD, P. (1977) *Forecasting Economic Time Series*. New York: Academic Press.

GREENWOOD, J. A. and DURAND, D. (1960) Aids for fitting the gamma distribution by maximum likelihood. *Technometrics* 2, 55–65.

HANNAN, E. J. (1970) *Multiple Time Series*. New York: Wiley.

HILL, B. M. (1963) The three-parameter lognormal distribution and Bayesian analysis of a point source epidemic. *J. Amer. Statist. Ass.* 58, 72–84.

JACOBS, P. A. and LEWIS, P. A. W. (1978a) Discrete time series generated by mixture I: correlational and runs properties. *J. Roy. Statist. Soc. Ser. B* 40, 94–105.

—— (1978b) Discrete time series generated by mixtures II: asymptotic properties. *J. Roy. Statist. Soc. Ser. B* 40, 222–8.

JOHNSON, N. L. and KOTZ, S. (1970) *Continuous Univariate Distributions*, Vol. 2. New York: Wiley.

KABAILA, P. (1983) On the asymptotic efficiency of estimators of the parameters of an ARMA process. *J. Time Series Anal.* 4, 37–47.

KENDALL, M. G. and STUART, ALAN (1961) *The Advanced Theory of Statistics*, Vol. II. New York: Macmillan.

KLIMKO, L. A. and NELSON, P. I. (1978) On conditional least squares estimation for stochastic processes. *Ann. Statist.* 6, 629–42.

LAWRANCE, A. J. and LEWIS, P. A. W. (1980) The exponential autoregressive moving-average EARMA (*p*, *q*) process. *J. Roy. Statist. Soc. Ser. B* 42, 150–61.

LEDOLTER, J. (1979) Inference robustness of ARIMA models under non-normality—Special application to stock price data. *Metrika* 26, 43–56.

LI, W. K. (1981) Topics in time series modelling. Unpublished Ph.D. Thesis, University of Western Ontario, Canada.

MARSAGLIA, G. (1976) Random number generation. In *Encyclopedia of Computer Science* (ed. A. Ralston). New York: Petrocelli and Charter pp. 1192–7.

MCLEOD, A. I. (1974) Contributions to applied time series. Master Thesis, University of Waterloo.

—— (1978) On the distribution of residual autocorrelations in Box–Jenkins Models. *J. Roy. Statist. Soc. Ser. B* 40, 296–302.

MCLEOD, A. I. and LI, W. K. (1983) Diagnostic checking ARMA time series models using squared-residual autocorrelations. *J. Time Series Anal.* 4, 269–73.

MORAN, P. A. P. (1953) The statistical analysis of the Canadian lynx cycle. *Aust. J. Zool.* 1, 163–73; 291–8.

NELSON, H. L. and GRANGER, C. W. J. (1979) Experiences with using the Box–Cox transformation when forecasting economic time series. *J. Econometrics* 10, 57–69.

NICHOLLS, D. F. and QUINN, B. G. (1982) *Random Coefficient Autoregressive Models: An Introduction.* New York: Springer-Verlag.

O'CONNELL, P. and JONES, D. A. (1979) Some experience with the development of models for the stochastic simulation of daily flows, inputs for risk analysis in water systems. *Water Resources Publications*, Colorado, 281–314.

POWELL, M. J. D. (1964) An efficient method for finding the minimum of a function of several variables without calculating derivatives. *Computer J.* 7, 155–62.

QUIMPO, R. G. (1967) Stochastic model for daily river flow sequences. Hydrology papers, Colorado State University, Colorado.

TONG, H. (1983). *Threshold Models in Non-linear Time Series Analysis.* New York: Springer-Verlag.

TONG, H. and LIM, K. S. (1980) Threshold autoregression, limit cycles and cyclical data (with discussion). *J. Roy. Statist. Soc. Ser. B* 42, 245–92.

WEISS, G. (1975) Time-reversibility of linear stochastic processes. *J. Appl. Prob.* 12, 831–6.

—— (1977) Shot noise models for the generation of synthetic streamflow data. *Water Resources Research* 13, 101–8.

WHITTLE, P. (1962) Gaussian estimation in stationary time series. *Bull Int. Statist. Inst.* 39, 105–29.

YAKOWITZ, S. J. (1973) A stochastic model for daily riverflow time series. *Water Resources Research* 9, 1271–85.