Statistica Sinica **6**(1996), 1-22

MAXIMUM LIKELIHOOD ESTIMATION FOR NONGAUSSIAN NONMINIMUM PHASE ARMA SEQUENCES

Keh-Shin Lii and Murray Rosenblatt

University of California, Riverside and University of California, San Diego

Abstract. We consider an approximate maximum likelihood procedure for estimating parameters of possibly noncausal and noninvertible autoregressive moving average processes driven by independent identically distributed nonGaussian noise. It is shown that the normalized approximate likelihood has a global maximum at true parameter values in the nonGaussian case. Under appropriate conditions, estimates of parameters that are solutions of likelihood equations exist, are consistent and asymptotically normal. An asymptotic covariance matrix is given. The procedure is illustrated with simulation examples of ARMA$(1,1)$ processes.

Key words and phrases: Maximum likelihood estimates, asymptotic normality, autoregressive moving average processes, noncausal, noninvertible, nonGaussian.

1. Introduction

We consider, in this paper, approximate maximum likelihood estimation for an autoregressive moving average (ARMA) process

$$x_t - \tilde{\phi}_1 x_{t-1} - \cdots - \tilde{\phi}_p x_{t-p} = z_t - \tilde{\theta}_1 z_{t-1} - \cdots - \tilde{\theta}_{p'} z_{t-p'}, \tag{1.1}$$

where the $\tilde{\phi}_j$'s and $\tilde{\theta}_k$'s are real weights and the z_τ's are real independent, identically distributed random variables with mean 0, variance σ^2 and finite fourth moment. Assume that the polynomials $\phi(z) = 1 - \sum_{j=1}^{p} \tilde{\phi}_j z^j$, $\theta(z) = 1 - \sum_{k=1}^{p'} \tilde{\theta}_k z^k$ have no zeros in common. There is then a stationary solution (which is uniquely determined) if and only if $\phi(z)$ has no zeros of absolute value one. The spectral density of the process x_t is then given by $S(\lambda) = (\sigma^2/2\pi)\ |\theta(e^{-i\lambda})/\phi(e^{-i\lambda})|^2$. Quadratic statistics can be used to estimate the second order spectral density but not usually to estimate $\phi(e^{-i\lambda})$, $\theta(e^{-i\lambda})$ since there are many distinct function pairs $\phi(e^{-i\lambda})$, $\theta(e^{-i\lambda})$ that are possible corresponding to the same spectral density $S(\lambda)$. If the process $\{x_t\}$ is Gaussian, the polynomials $\phi(z)$, $\theta(z)$ are not identifiable unless one makes, for example, the assumption that they are minimum phase, that is, their zeros all have modulus greater than one. In the case of a nonGaussian sequence $\{x_t\}$ the polynomials $\phi(z)$, $\theta(z)$ are identifiable. The process $\{x_t\}$ is causal if the polynomial $\phi(z)$ is minimum phase. The process is invertible if polynomial $\theta(z)$ is minimum phase. Interest in nonGaussian processes without minimum phase assumptions on the polynomials $\phi(z)$, $\theta(z)$ does arise in a number of applications. In one of these applications a deconvolution problem of the type that arises in seismic investigations is at issue. A discussion of this type of problem is given in Wiggins (1978, 1985) and Donoho (1981).

There is an extensive literature concerned with stationary Gaussian autoregressive moving average sequences that has been generated over an appreciable period of time. The interest in nonGaussian ARMA sequences is of fairly recent vintage, especially that concerned with an improvement in performance of parameter estimates beyond that provided by a "quasiGaussian" likelihood function, namely, that provided by estimates computed as if the sequence had a Gaussian likelihood (Rosenblatt (1985, Chapter 4)).

The stationary solution of (1.1) is causal, that is, has a representation of the form $x_t = \sum_{j=0}^{\infty} a_j z_{t-j}$ if and only if $\phi(z)$ has no zeros inside the unit disc in the complex plane. Thus, all moving average models $x_t = z_t - \tilde{\theta}_1 z_{t-1} - \cdots - \tilde{\theta}_{p'} z_{t-p'}$ are obviously causal whether they are minimum phase or not. The following four moving average models all have the same spectral density $x_t = z_t - (19/12)z_{t-1} + (15/24)z_{t-2}$ with $\sigma^2 = 1$; $x_t = z_t - (38/15)z_{t-1} + (8/5)z_{t-2}$ with $\sigma^2 = (5/8)^2$; $x_t = z_t - (39/20)z_{t-1} + (9/10)z_{t-2}$ with $\sigma^2 = (5/6)^2$; $x_t = z_t - (13/6)z_{t-1} + (20/18)z_{t-2}$ with $\sigma^2 = (3/4)^2$. They are different processes if z_t is nonGaussian. Only the first process is minimum phase. Yet it is clear that one would want to be able to distinguish between these viable causal models in the nonGaussian context.

It is clear that autoregressive models satisfying $x_t - \tilde{\phi}_1 x_{t-1} - \cdots - \tilde{\phi}_p x_{t-p} = z_t$ that are nonminimum phase are noncausal. Consider a one-dimensional random field or a transect of a random field as would be natural in the case of areal population surveys of trees or animal populations. If in one direction the model were causal, then in the other (opposite) direction it would be noncausal. We are dealing here with models in which t is not a time parameter but rather the scale along a transect. It would be equally natural in the case of a photographic film or the transect along a photographic plate. Thus, it is quite natural and of interest to deal with noncausal nonGaussian autoregressive schemes in such contexts.

Notice that since $\sigma|\theta(e^{-i\lambda})/\phi(e^{-i\lambda})| = (2\pi S(\lambda))^{1/2}$, additional information required in estimating $\theta(e^{-i\lambda})$ and $\phi(e^{-i\lambda})$ certainly involves the phase information $\arg\{\theta(e^{-i\lambda})\}$, $\arg\{\phi(e^{-i\lambda})\}$. This information is not available (or meaningful) in the Gaussian case but is available in the nonGaussian context. Though such information cannot be obtained by quadratic or second order spectral methods, it can be resolved by using third or higher order cumulant spectral estimates. An earlier paper using such methods is Lii and Rosenblatt (1982). There has been much interest in the engineering literature in such problems (Mendel

(1991)). Another area of application is to be found in what is called "speckle masking" in astronomy. Here related methods are used to overcome the degradation of telescopic images caused by atmospheric turbulence. Discussions of this application can be found in the papers of Lohmann, Weigelt, and Wirnitzer (1983) and Bartelt, Lohmann and Wirnitzer (1984). In Kay and Sengupta (1991) the improvement in spectral estimation for linear nonGaussian minimum phase processes is studied, when the distribution of the independent random variables generating the process is known. Of course, corresponding questions for second and higher order spectra are clearly of interest when the process studied is nonminimum phase. A recent survey of other applications of the models considered here is Mendel (1991) and references therein. Methods of estimation of parameters considered in these references are based upon higher order moments.

Methods based on higher (than second) order moments or cumulant spectra are not efficient in estimating parameters of finite parameter schemes. Kreiss (1987) has determined the behavior of asymptotically efficient parameter estimates when the functions $\phi(\,\cdot\,)$ and $\theta(\,\cdot\,)$ satisfy the minimum phase assumption. These estimates improve upon the standard estimates based on a Gaussian likelihood by the Fisher information of the density function. Breidt, Davis, Lii and Rosenblatt (1991, which will be abbreviated as BDLR) have studied the behavior of asymptotically efficient parameter estimates for possibly nonminimum phase autoregressive sequences. Corresponding results were obtained by a different procedure for the case of possibly nonminimum phase moving average sequences by Lii and Rosenblatt (1992, which will be abbreviated as LR). An appropriate modification of the procedure employed in this last paper is used to get the appropriate results for general stationary nonGaussian ARMA sequences.

The system of equations (1.1) can be rewritten as $\phi(B)x_t = \theta(B)z_t$ in terms of the backward shift operator B. Let

$$\begin{aligned}\phi(z) &= \phi^+(z)\phi^*(z) = (1-\phi_1 z-\cdots-\phi_r z^r)(1-\phi_{r+1}z-\cdots-\phi_p z^s),\\ \theta(z) &= \theta^+(z)\theta^*(z) = (1-\theta_1 z-\cdots-\theta_{r'} z^{r'})(1-\theta_{r'+1}z-\cdots-\theta_{p'} z^{s'}),\end{aligned} \tag{1.2}$$

where ϕ^+ and θ^+ have no roots on the closed unit disc and ϕ^* and θ^* have all their roots in the interior of the unit disc. Also

$$\begin{aligned}\phi^+(z)^{-1} &= \sum_{j=0}^{\infty}\alpha_j z^j\ , \quad \phi^*(z)^{-1} = \sum_{j=s}^{\infty}\beta_j z^{-j}\ ,\\ \theta^+(z)^{-1} &= \sum_{j=0}^{\infty}\alpha'_j z^j\ , \quad \theta^*(z)^{-1} = \sum_{j=s'}^{\infty}\beta'_j z^{-j}\ .\end{aligned} \tag{1.3}$$

The random variables z_t are assumed to have density function $f_\sigma(z) = \frac{1}{\sigma}f(\frac{z}{\sigma})$.

Let

$$C1 = E\left(\frac{f'_\sigma}{f_\sigma}(z)\right)^2 \sigma^2 , \quad C2 = E\left(z\,\frac{f'_\sigma}{f_\sigma}(z)\right)^2 . \tag{1.4}$$

The object is to estimate the parameters $\boldsymbol{\eta} = (\eta_u,\, u = 1, \ldots, p+p'+1)$ where $\eta_u = \phi_u$, $u = 1, \ldots, p$; $\eta_u = \theta_{u-p}$, $u = p+1, \ldots, p+p'$; $\eta_{p+p'+1} = \sigma$.

The approximate likelihood function is derived in Section 2. The fact that the limit of the normalized likelihood has a global maximum at the true parameter value is established in Section 3. The limiting covariance properties of the logarithm of a quasi-likelihood (defined in terms of the z_t's rather than the actual observations) are derived in Section 4 and are shown to be given by covariance (1.5) and Table 1. The approximate likelihood equations in terms of $x_1, \ldots, x_n$ alone are given in Section 5. Given that r, s, r', s' have been determined, it is shown that there is a sequence of solutions $\boldsymbol{\eta}_n$ of the likelihood equations that are consistent and such that $n^{1/2}(\boldsymbol{\eta}_n - \boldsymbol{\eta})$ is asymptotically normal with limiting covariance matrix Σ^{-1} as given by (1.5). Here $\boldsymbol{\eta}$ is, of course, the true parameter vector. A number of simulations are described in Section 6. In the rest of this section we set down some further notation and conditions for the paper. Consider the positive definite matrix

$$\Sigma = (\sigma_{u,v}\ ;\ u, v = 1, \ldots, p+p'+1) \tag{1.5}$$

with $\sigma_{u,v}$ given in Table 1.

We shall show that under appropriate conditions approximate maximum likelihood estimates $\hat{\boldsymbol{\eta}}$ of the unknown parameters $\boldsymbol{\eta}$ are such that $\sqrt{n}\,(\hat{\boldsymbol{\eta}} - \boldsymbol{\eta})$ is asymptotically normally distributed with mean 0 and covariance matrix Σ^{-1}.

As in BDLR (1991) and LR (1992) we make the following assumptions on the density f of the i.i.d. random variable z_t which we call Assumption A.

A1. $f(x) > 0 \quad \text{for all} \quad x$

A2. $f \in C^2$

A3. $f' \in L_1 \quad \text{with} \quad \int f'(x)\,dx = f(x)\Big|_{-\infty}^{\infty} = 0$

A4. $\int x\,f'(x)\,dx = x\,f(x)\Big|_{-\infty}^{\infty} - \int f(x)\,dx = -1$

A5. $\int f''(x)\,dx = f'(x)\Big|_{-\infty}^{\infty} = 0$

A6. $\int x\,f''(x)\,dx = x\,f'(x)\Big|_{-\infty}^{\infty} - \int f'(x)\,dx = 0$

A7. $\int x^2 f''(x)\,dx = x^2 f'(x)\Big|_{-\infty}^{\infty} - 2\int x\,f'(x)\,dx = 2$

A8. $\int (1+x^2)(f'(x))^2/f(x)\,dx < \infty$.

We also assume

B. $|u(z+h)-u(z)| \le A((1+|z|^k)|h|+|h|^\ell)$ for all z, h with k, ℓ, A fixed positive constants and $u(\cdot) = f'/f\,,\ (f'/f)'$.

Table 1. $\sigma_{u,v}$ in (1.5)

$$\sigma_{u,v} = \begin{cases} C1 \sum_{j=0}^{\infty} \alpha_j \alpha_{j+|u-v|}, & \text{if } u,v = 1,\dots,r, \\ C1 \sum_{j=s}^{\infty} \beta_j \beta_{j+|u-v|}, & \text{if } u,v = r+1,\dots,p, \\ & \quad (u,v) \neq (p,p), \\ \beta_s^2(C2-1) + C1 \sum_{j=s}^{\infty} \beta_{j+1}^2, & \text{if } u = v = p, \\ \sum_j \alpha_{j-u}\beta_{j+v-r}, & \text{if } u=1,\dots,r,\ v=r+1,\dots,p, \\ C1 \sum_{j=0}^{\infty} \alpha'_j \alpha'_{j+|v-u|}, & \text{if } u,v = p+1,\dots,p+r', \\ C1 \sum_{j=s'}^{\infty} \beta'_j \beta'_{j+|v-u|}, & \text{if } u,v = p+r'+1,\dots,p+p', \\ & \quad (u,v) \neq (p+p',p+p'), \\ \beta_{s'}^{\prime 2}(C2-1) + C1 \sum_{j=s'}^{\infty} \beta_{j+1}^{\prime 2}, & \text{if } u = v = p+p', \\ \sum_j \alpha'_{j-u}\beta'_{j+v-r'}, & \text{if } u = p+1,\dots,p+r', \\ & \quad v = p+r'+1,\dots,p+p', \\ -C1 \sum_j \alpha_j \alpha'_{j+u-v+p}, & \text{if } u = 1,\dots,r, \\ & \quad v = p+1,\dots,p+r', \\ -C1 \sum_j \beta_j \beta'_{j+r-u-r'+v-p}, & \text{if } u = r+1,\dots,p, \\ & \quad v = p+r'+1,\dots,p+p', \\ & \quad (u,v) \neq (p,p+p'), \\ -\beta_s\beta'_{s'}(C2-1) - C1 \sum_{j=1}^{\infty} \beta_{s+j}\beta'_{s'+j}, & \text{if } (u,v) = (p,p+p'), \\ -\sum_j \alpha_{j-u}\beta'_{j+v-p-r'}, & \text{if } u = 1,\dots,r, \\ & \quad v = p+r'+1,\dots,p+p', \\ -\sum_j \alpha'_{j-v+p}\beta_{j+u-r}, & \text{if } u = r+1,\dots,p, \\ & \quad v = p+1,\dots,p+r', \\ \sigma^{-1}\beta_s(C2-1), & \text{if } u = p,\ v = p+p'+1, \\ -\sigma^{-1}\beta'_{s'}(C2-1), & \text{if } v = p+p',\ v = p+p'+1, \\ \sigma^{-2}(C2-1), & \text{if } u = v = p+p'+1, \\ 0, & \text{otherwise.} \end{cases}$$

2. The Likelihood Function

In this section we derive an approximation to the likelihood function of $\boldsymbol{x} = (x_1, \ldots, x_n)$. The basic idea is to augment the data vector $\boldsymbol{x}$ by a vector $\boldsymbol{w}$ of fixed size where $(\boldsymbol{w}, \boldsymbol{x})$ can be mapped linearly in a $1-1$ manner onto a vector $(\boldsymbol{y}, \boldsymbol{z})$ and $\boldsymbol{y}$ is a vector of fixed size independent of $\boldsymbol{z} = (z_k, \ldots, z_{n+j})$. Here j and k are fixed. The density of $(\boldsymbol{w}, \boldsymbol{x})$ can be written as $h(\boldsymbol{y}) \prod_{t=k}^{n+j} f_\sigma(z_t) \cdot D_n$ with h the density of $\boldsymbol{y}$ and D_n the Jacobian of the transformation from $(\boldsymbol{y}, \boldsymbol{z})$ to $(\boldsymbol{w}, \boldsymbol{x})$. The approximate likelihood is obtained by ignoring the term $h(\boldsymbol{y})$ and replacing z_t (as a function of the x's) by an approximation $z_t(q)$ that depends on the x_t's only for $q < t \leq n - q$ and not on $\boldsymbol{w}$ with $q = q(n) \to \infty$ as $n \to \infty$ and $q(n) = o(n)$. The contribution of z_t for $t < q$, $t > n - q$ is also ignored. Details are given in the following discussion. Let

$$\begin{aligned} u_t &= \phi^*(B) x_t, \quad \xi_t = \theta^*(B) z_t, \\ v_t &= \phi^+(B) x_t, \quad \zeta_t = \theta^+(B) z_t. \end{aligned} \tag{2.1}$$

We consider transformations

$$(x_{1-p}, \ldots, x_{n+p}) \to (u_{1-r}, \ldots, u_{n+p}, v_{n+p+1-s}, \ldots, v_{n+p}), \tag{2.2}$$

$$(\xi_{1-r'}, \ldots, \xi_0, u_{1-r}, \ldots, u_{n+p}) \to (u_{1-r}, \ldots, u_0, \xi_{1-r'}, \ldots, \xi_{n+p}), \tag{2.3}$$

$$(\xi_{1-r'}, \ldots, \xi_{n+p}, \zeta_{n+p+1-s'}, \ldots, \zeta_{n+p}) \to (z_{1-p'}, \ldots, z_{n+p}). \tag{2.4}$$

The transformation (2.2) is the linear mapping

$$\begin{bmatrix} -\phi_p & -\phi_{p-1} & \cdots & -\phi_{r+1} & 1 & 0 & 0 & \cdots & 0 \\ 0 & -\phi_p & & \cdots & -\phi_{r+1} & 1 & 0 & \cdots & 0 \\ & & \ddots & & & \ddots & & & \\ & & & \ddots & & & & & 1 \\ \hline & -\phi_r & & -\phi_{r-1} & \cdots & 1 & 0 & \cdots & 0 \\ & & \ddots & \ddots & \ddots & \ddots & & & \\ & & & & -\phi_r & -\phi_{r-1} & \cdots & & 1 \end{bmatrix} \begin{bmatrix} x_{1-p} \\ \vdots \\ x_{n+r} \\ x_{n+r+1} \\ \vdots \\ x_{n+p} \end{bmatrix} = \begin{bmatrix} u_{1-r} \\ \vdots \\ u_{n+p} \\ v_{n+p+1-s} \\ \vdots \\ v_{n+p} \end{bmatrix},$$

where the lower right rectangular submatrix of the transformation matrix indicated is $s \times (r+s)$. The Jacobian of the transformation is seen to be $|\phi_p|^{n+p} h_1(\phi_1, \ldots, \phi_p)$ for a function h_1. The linear transformation (2.3) is based on the minimum phase system $\phi^+(B) u_t = \theta^+(B) \xi_t$ and so has a unit Jacobian. The transformation (2.4) is the transformation of (2.2) in the inverse direction using (2.1). The transformation (2.4) has Jacobian $|\theta_{p'}|^{-n-p} h_2^{-1}(\theta_1, \ldots, \theta_{p'})$. Let us consider the following sequence of transformations obtained by augmenting the variables in transformations (2.2) to (2.4)

$$\begin{aligned}
&(\xi_{1-r'},\ldots,\xi_0,\, x_{1-p},\ldots,x_{n+p},\, \zeta_{n+p+1-s'},\ldots,\zeta_{n+p})\\
&\to (\xi_{1-r'},\ldots,\xi_0,\, u_{1-r},\ldots,\, u_{n+p},\, v_{n+p+1-s},\ldots,\, v_{n+p},\, \zeta_{n+p+1-s'},\ldots,\zeta_{n+p})\\
&\to (u_{1-r},\ldots,\, u_0,\, \xi_{1-r'},\ldots,\, \xi_{n+p},\, v_{n+p+1-s},\ldots,\, v_{n+p},\, \zeta_{n+p+1-s'},\ldots,\, \zeta_{n+p})\\
&\to (u_{1-r},\ldots,\, u_0,\, z_{1-p'},\ldots,\, z_{n+p},\, v_{n+p+1-s},\ldots,v_{n+p}). \qquad (2.5)
\end{aligned}$$

Note that $u_t = \phi^*(B)x_t = \phi^+(B)^{-1}\theta(B)z_t = \sum_{j=0}^{\infty}\alpha_j\{z_{t-j} - \tilde{\theta}_1 z_{t-j-1}\cdots - \tilde{\theta}_{p'}z_{t-j-p'}\}$ is independent of $\{z_{t+1},\, z_{t+2},\ldots\}$ and that $v_t = \phi^+(B)x_t = \phi^*(B)^{-1}$ $\theta(B)z_t$ is independent of $\{z_{t+s-p'-1},\, z_{t+s-p'-2},\ldots\}$. The joint probability density of the variables in the parentheses on the extreme right of (2.5) can be written

$$g_1(u_{1-r},\ldots,u_0,\, z_{1-p'},\ldots,z_0)\prod_{j=1}^{n+p-p'} f_\sigma(z_j)g_2(z_{n+p-p'+1},\ldots,z_{n+p},v_{n+p+1-s},\ldots,v_{n+p})$$

with g_1, g_2 joint probability densities of the arguments listed. Thus the joint probability density of $(\xi_{1-r'},\ldots,\, \xi_0,\, x_{1-p},\ldots,x_{n+p},\, \zeta_{n+p+1-s'},\ldots,\zeta_{n+p})$ is

$$\ell(\boldsymbol{\eta}) \equiv g_1 \prod_{j=1}^{n+p-p'} f_\sigma(z_j)g_2\cdot|\phi_p/\theta_{p'}|^n c(\phi_1,\ldots,\, \phi_p,\theta_1,\ldots,\, \theta_{p'}), \tag{2.6}$$

where g_1, g_2, and $c(\phi_1,\ldots,\phi_p,\, \theta_1,\ldots,\, \theta_{p'})$ do not depend on n and so in later arguments can be ignored for large n.

The expressions obtained suggest that one can approximate $1/n$ times the log of the likelihood (2.6) by

$$\frac{1}{n-2q}\sum_{t=q}^{n-q}\log f_\sigma(z_t) + \log|\phi_p| - \log|\theta_{p'}|, \tag{2.7}$$

where z_t is unobserved (but expressible in terms of $\boldsymbol{\xi}$, $\boldsymbol{x}$, $\boldsymbol{\zeta}$) and $q = q(n) \to \infty$ as $n \to \infty$ with $q(n) = o(n)$. A further approximation to (2.7) is given by

$$\frac{1}{n-2q}\sum_{t=q+1}^{n-q}\log f_\sigma(z_t(q)) + \log|\phi_p| - \log|\theta_{p'}| \tag{2.8}$$

with

$$z_t(q) = [\theta(B)^{-1}\phi(B)]_q x_t, \tag{2.9}$$

where $[\theta(z)^{-1}\phi(z)]_q$ is the truncated Laurent expansion of $\theta(z)^{-1}\phi(z)$ extended over powers z^k with $|k| \leq q$. This is plausible since the coefficients of z^k in the expansion decay exponentially fast to zero as $|k| \to \infty$. Note that $z_t(q)$ is completely expressible in terms of the observations $\{x_1, \ldots, x_n\}$ for $t = q + 1, \ldots, n - q$ and is computable given a set of parameters $\boldsymbol{\eta}$.

In Section 4 we proceed initially with our derivations using (2.7) as if the true z_t's can be computed. This would be the case if the infinite sequence $\{x_t\}$ were available. Likelihood equations will be derived in terms of (2.7) for quasi-estimates $\hat{\boldsymbol{\eta}}$ of $\boldsymbol{\eta}$. The asymptotic behavior of a sequence of such quasi-estimates will be examined as $n \to \infty$. Then, in Section 5, we show that the solutions of corresponding set of likelihood equations derived from the approximation (2.8) have the same statistical properties asymptotically.

3. Extremum of the Limiting Normalized Approximate Likelihood

The actual approximate loglikelihood that we consider has the form

$$L_q(\boldsymbol{\eta}) = \sum_{t=q}^{n-q} \log f_\sigma(z_t(q)) + (n-2q)[\log|\phi_p| - \log|\theta_{p'}|],$$

where $q = q(n) \to \infty$ as $n \to \infty$ but $q(n) = o(n)$. Here $z_t(q)$ is given in (2.9). With distinct real zeros of $\phi(z)$, $\theta(z)$ and no zeros reciprocal of other zeros, there are $2^{p+p'}$ different likelihoods corresponding to root location inside or outside the unit circle in the complex plane.

First, note that Assumptions A and B imply that

$$E_{\boldsymbol{\eta}_0}|\log f_\sigma(z_t(q)) - \log f_\sigma(M(\boldsymbol{\phi}, \boldsymbol{\theta})x_t)| \to 0 \tag{3.1}$$

as $n \to \infty$, where $M(\boldsymbol{\phi}, \boldsymbol{\theta}) = \theta(B)^{-1}\phi(B)$ with $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_p)$, $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_{p'})$ and $\boldsymbol{\eta}_0$ is the true parameter vector. This follows directly on using Taylor's formula with error term, Assumption B and the bounds on moments available.

Lemma 1. *Assumptions* A *and* B *imply that*

$$\frac{1}{n}L_q(\boldsymbol{\eta}) \to E_{\boldsymbol{\eta}_0} \log f_\sigma(M(\boldsymbol{\phi}, \boldsymbol{\theta})x_t) + \log|\phi_p| - \log|\theta_{p'}| \tag{3.2}$$

in mean as $n \to \infty$ where $\boldsymbol{\eta}_0$ is the true parameter vector. Further, the right hand side of (3.2) *takes on the maximal value at $\boldsymbol{\eta} = \boldsymbol{\eta}_0$ in the nonGaussian case when* $\log(g_1 g_2)$ *is integrable for all admissible parameter values.*

The limiting relation (3.2) follows directly from (3.1) and the ergodic theorem. The expectation on the right side of (3.2) can be rewritten

$$\begin{aligned} &E_{\boldsymbol{\eta}_0} \log f_1(M(\boldsymbol{\phi}, \boldsymbol{\theta})\sigma^{-1}x_t) + \log|\phi_p| - \log|\theta_{p'}| - \log\sigma \\ &= E_{\boldsymbol{\eta}_0} \log f_1(M(\boldsymbol{\phi}, \boldsymbol{\theta})\sigma^{-1}M(\boldsymbol{\phi}_0, \boldsymbol{\theta}_0)^{-1}\sigma_0\xi_t) + \log|\phi_p| - \log|\theta_{p'}| - \log\sigma, \end{aligned} \tag{3.3}$$

where the ξ_t's are independent with common density f_1. The integrability of $\log(g_1 g_2)$ implies that

$$\frac{1}{n} E_{\boldsymbol{\eta}_0} | \log \ell(\boldsymbol{\eta}) - \log L_q(\boldsymbol{\eta}) | \to 0$$

as $n \to \infty$. Jensen's inequality implies that $E_{\boldsymbol{\eta}_0}[\log \ell(\boldsymbol{\eta})]$ is uniquely maximized by $\boldsymbol{\eta} = \boldsymbol{\eta}_0$ for

$$E_{\boldsymbol{\eta}_0}(\log\{\frac{\ell(\boldsymbol{\eta})}{\ell(\boldsymbol{\eta}_0)}\}) < \log E_{\boldsymbol{\eta}_0}(\frac{\ell(\boldsymbol{\eta})}{\ell(\boldsymbol{\eta}_0)}) = 0 \qquad \text{if } \boldsymbol{\eta} \neq \boldsymbol{\eta}_0,$$

and so $E_{\boldsymbol{\eta}_0}\{\log \ell(\boldsymbol{\eta})\} < E_{\boldsymbol{\eta}_0}\{\log \ell(\boldsymbol{\eta}_0)\}$ if $\boldsymbol{\eta} \neq \boldsymbol{\eta}_0$. In the limit as $n \to \infty$ it is clear that (3.2) is maximized at $\boldsymbol{\eta} = \boldsymbol{\eta}_0$.

In computational practice the isolation of the proper surface of the $2^{p+p'}$ surfaces can proceed by using a grid of points on the surfaces at which the approximate likelihood would be evaluated and the maximum value computed. Given the choice of the proper surface, one would then carry on the search there. Of course, a number of alternate procedures could be used. For example, a preliminary search for the proper surface could be made by using the cruder nonparametric estimate given in Lii and Rosenblatt (1982). With such an initial search one can determine r, s and r', s'. This is illustrated in the examples of Section 6. We note that $z_t(q)$ is expressed in terms of observed $\{x_1, \ldots, x_n\}$ for $t = q+1, \ldots, n-q$ and is computable given a set of parameters $\boldsymbol{\eta}$.

In the next section we proceed with our derivations using (2.7) as if true z_t can actually be computed. They could be if the infinite sequence of $\{x_t\}$ were available. Likelihood equations will be derived in terms of (2.7) for quasi-estimates $\hat{\boldsymbol{\eta}}$ of $\boldsymbol{\eta}$. The asymptotic behavior of a sequence of such quasi-estimates will be examined as $n \to \infty$. Then we show, in Section 5, that the solutions of a corresponding set of likelihood equations derived from the approximation by (2.8) have the same statistical properties asymptotically.

4. Covariance Properties

Let

$$L(\boldsymbol{\eta}) = L(\boldsymbol{\phi}, \boldsymbol{\theta}, \sigma) = \sum_{t=a+1}^{n-a} g_t(\boldsymbol{\phi}, \boldsymbol{\theta}, \sigma) = \sum_{t=a+1}^{n-a} g_t(\boldsymbol{\eta}) \tag{4.1}$$

with

$$g_t(\boldsymbol{\eta}) = \log\, f_\sigma(\phi^+(B)\phi^*(B)\theta^+(B)^{-1}\theta^*(B)^{-1}x_t) + \log|\eta_p| - \log|\eta_{p+p'}|\,. \tag{4.2}$$

Then

$$\begin{aligned} \frac{\partial}{\partial \phi_j}\, g_t(\boldsymbol{\eta}) &= -\frac{f'_\sigma}{f_\sigma}\,(z_t)\,\cdot\, B^j \phi^*(B)\theta^+(B)^{-1}\theta^*(B)^{-1}x_t \\ &= -\frac{f'_\sigma}{f_\sigma}\,(z_t)\,\cdot\, B^j \phi^+(B)^{-1} z_t\,, \qquad j = 1, \ldots, r, \end{aligned}$$

$$\begin{aligned}
\frac{\partial}{\partial\theta_k}\, g_t(\boldsymbol{\eta}) &= \frac{f'_\sigma}{f_\sigma}\,(z_t)\,\cdot\, B^k\theta^+(B)^{-1}z_t, \quad k=1,\ldots,r'\,, \\
\frac{\partial}{\partial\phi_\ell}\, g_t(\boldsymbol{\eta}) &= -\frac{f'_\sigma}{f_\sigma}\,(z_t)B^{\ell-r}\phi^*(B)^{-1}z_t + \delta_{p,\ell}\frac{1}{\phi_p}\,, \quad \ell=r+1,\ldots,p\,, \\
\frac{\partial}{\partial\theta_m}\, g_t(\boldsymbol{\eta}) &= \frac{f'_\sigma}{f_\sigma}\,(z_t)B^{m-r'}\theta^*(B)^{-1}z_t - \delta_{p',m}\frac{1}{\theta_{p'}}\,, \quad m=r'+1,\ldots,p'\,, \\
\frac{\partial}{\partial\sigma}\, g_t(\boldsymbol{\eta}) &= \frac{\partial}{\partial\eta_{p+q+1}}\, g(\boldsymbol{\eta}) = -\frac{1}{\sigma}\left(z_t\frac{f'_\sigma(z_t)}{f_\sigma(z_t)}+1\right),
\end{aligned}$$

where $f_\sigma(x)=(1/\sigma)f(x/\sigma)$ and $f'_\sigma(x)=(1/\sigma^2)f'(x/\sigma)$. Using (1.3) and the fact that $\beta_s=-(1/\phi_p)$ and $\beta_{s'}=-(1/\theta_{p'})$, together with Assumption A we can readily verify that $E((\partial/\partial\eta_i)\; g_t)=0$ for $i=1,\ldots,p+p'+1$. From (1.3) we can verify that

$$\mathrm{Cov}\;\left(\frac{\partial g_t}{\partial\phi_j},\frac{\partial g_\tau}{\partial\theta_k}\right)=\begin{cases}0, & \text{if } t\neq\tau,\\ -E\left(\frac{f'_\sigma}{f_\sigma}\,(z_t)\right)^2\,\cdot\,\sum_i \alpha_i\alpha'_{i+j-k}\sigma^2, & \text{if } t=\tau,\end{cases}$$

for $j=1,\ldots,r$; $k=1,\ldots,r'$. It is understood that $\alpha_\ell=\alpha'_\ell=0$ for $\ell<0$.

Similarly

$$\mathrm{Cov}\;\left(\frac{\partial g_t}{\partial\phi_\ell},\;\frac{\partial g_\tau}{\partial\theta_m}\right)=\begin{cases}0, & \text{if } t\neq\tau,\\ -E\left(\frac{f'_\sigma}{f_\sigma}(z_t)\right)^2\,\cdot\,\sum_j \beta_j\beta'_{j+r-\ell-r'+m}\sigma^2\,, & \text{if } t=\tau,\end{cases}$$

for $\ell=r+1,\ldots,p$, $m=r'+1,\ldots,p'$, $(\ell,m)\neq(p,p')$ where it is understood that $\beta_\ell=0$ if $\ell<s$ and $\beta'_\ell=0$ if $\ell<s'$. In a similar way we can verify the following relations

$$\mathrm{Cov}\;\left(\frac{\partial g_t}{\partial\phi_j},\frac{\partial g_\tau}{\partial\theta_m}\right)=\begin{cases}0, & \text{if } t\leq\tau,\\ -\alpha_{t-\tau-j}\beta'_{t-\tau+m-r'}, & \text{if } t>\tau,\end{cases}$$

for $j=1,\ldots,r$; $m=r'+1,\ldots,p'$,

$$\mathrm{Cov}\;\left(\frac{\partial g_t}{\partial\phi_\ell},\frac{\partial g_\tau}{\partial\theta_k}\right)=\begin{cases}0, & \text{if } t\geq\tau,\\ -\alpha'_{\tau-t-k}\beta_{\tau-t+\ell-r}, & \text{if } t<\tau,\end{cases}$$

for $\ell=r+1,\ldots,p$; $k=1,\ldots,r'$. Also

$$\begin{aligned}
&\mathrm{Cov}\;\left(\frac{\partial g_t}{\partial\phi_p},\frac{\partial g_\tau}{\partial\theta_{p'}}\right)\\
&=\mathrm{Cov}\;\left(-\frac{f'_\sigma}{f_\sigma}(z_t)B^s\sum_{j=s}^{\infty}\beta_jB^{-j}z_t+\frac{1}{\phi_p}\,,\frac{f'_\sigma}{f_\sigma}(z_\tau)B^{s'}\sum_{k=s'}^{\infty}\beta'_kB^{-k}z_\tau-\frac{1}{\theta_{p'}}\right)\\
&=\begin{cases}0, & \text{if } t\neq\tau,\\ -\beta_s\beta'_{s'}\left[E\left(z_t\frac{f'_\sigma}{f_\sigma}(z_t)\right)^2-1\right]-E\left(\frac{f'_\sigma}{f_\sigma}(z_t)\right)^2\,\cdot\,\sum_{j=1}^{\infty}\beta_{s+j}\beta'_{s'+j}\sigma^2, & \text{if } t=\tau\,,\end{cases}
\end{aligned}$$

$$\text{Cov}\left(\frac{\partial g_t}{\partial \phi_j}, \frac{\partial g_\tau}{\partial \sigma}\right) = \begin{cases} 0\,, & j=1,\ldots,p-1, \text{ for all } t,\tau, \\ 0\,, & j=p \text{ and } t \neq \tau, \\ \sigma^{-1}\beta_s E\left(z_t \frac{f'_\sigma}{f_\sigma}(z_t)\right)^2 - \sigma^{-1}\beta_s\,, & j=p \text{ and } t=\tau\,, \end{cases}$$

$$\text{Cov}\left(\frac{\partial g_t}{\partial \theta_j}, \frac{\partial g_\tau}{\partial \sigma}\right) = \begin{cases} 0, & j \neq p'\,, \text{ for all } t,\tau, \\ 0, & j = p'\,,\ t \neq \tau, \\ -\sigma^{-1}\beta'_{s'} E\left(z_t \frac{f'_\sigma}{f_\sigma}(z_t)\right)^2 + \sigma^{-1}\beta'_{s'}, & j = p',\ t=\tau\,, \end{cases}$$

and

$$\text{Cov}\ \left(\frac{\partial g_t}{\partial \sigma}, \frac{\partial g_\tau}{\partial \sigma}\right) = \begin{cases} 0, & t \neq \tau, \\ \sigma^{-2} E\left(z_t \frac{f'_\sigma}{f_\sigma}(z_t)\right)^2 - \sigma^{-2}, & t = \tau\,. \end{cases}$$

It is clear that Cov $(\partial g_t/\partial \phi_u, \partial g_\tau/\partial \phi_v)$ for $u,v = 1,\ldots,p$ will be the same as that given in BDLR (1991) and Cov $(\partial g_t/\partial \theta_u, \partial g_\tau/\partial \theta_v)$ for $u,v = 1,\ldots,p'$ will be the same as that given in LR (1992). We summarize these results in

Theorem 1. *Under the Assumptions* A, *for $a > 0$ we have*

$$\text{Cov}\ \left(\sum_{t=a+1}^{n-a} \frac{\partial g_t}{\partial \eta_u}, \sum_{\tau=a+1}^{n-a} \frac{\partial g_\tau}{\partial \eta_v}\right) \cong (n-2a)\sigma_{u,v}\,, \quad u,v = 1,\ldots,p+p'+1, \tag{4.3}$$

where $\sigma_{u,v}$ is given in Table 1. *Later, we shall take a as a function of n with $a(n) = o(n)$.*

Lemma 2. *Given Assumptions* A

$$(n-2a)^{1/2} \sum_{t=a}^{n-a} \frac{\partial g_t}{\partial \boldsymbol{\eta}} \xrightarrow{d} N(0,\Sigma) \tag{4.4}$$

as $n \to \infty$ with Σ given by (1.5).

The partial derivatives $\partial g_t/\partial \eta_u$, $u = 1,\ldots,p+p'+1$ are just as in BDLR (1991) and LR (1992), except possibly for centering of the form $\sum_j (f'_\sigma/f_\sigma)(z_t) \cdot\gamma_j z_{t-j}$ with weights γ_j that tend to zero exponentially fast as $|j| \to \infty$. If the expansions are truncated by neglecting terms indexed beyond a large $|j|$, the corresponding approximations for the partial derivatives are finite step dependent and thus asymptotically jointly normal. The asymptotic covariance behavior of the expressions on the left of (4.4) is given by Σ. A standard approximation argument gives the conclusion of the lemma.

Lemma 3. *Under Assumptions* A,

$$E\left(\frac{\partial^2 g_t}{\partial \eta_u \partial \eta_v}\right) = -\sigma_{u,v}\,, \quad u,v = 1,\ldots,p+p'+1\,.$$

It is clear that this identity is valid for $u, v = 1, \dots, p$ (see BDLR (1991)) and for $u, v = p+1, \dots, p+p'$ (see LR (1992)) as well as $u = p$, $v = p+p'+1$ and $u = p+p'$, $v = p+p'+1$. We shall just carry out the explicit computation for $u = 1, \dots, r$, $v = p+1, \dots, p+r'$ since the other cases with u an autoregressive variable and v a moving average variable are quite similar. If $u = 1, \dots, r$, as already noted, $\partial g_t / \partial \phi_u = -(f'_\sigma / f_\sigma)\ (z_t) B^u \phi^+(B)^{-1} z_t$ and so

$$\frac{\partial^2 g_t}{\partial \phi_u \partial \theta_v} = -\,\sigma^{-2}\Big\{ h'(\sigma^{-1} z_t) \Big[\sum_{j=0}^{\infty} \alpha_j B^{j+u} z_t\Big]\Big[\sum_{k=0}^{\infty} \alpha'_k B^{k+v-p} z_t\Big] + \sigma^2 \frac{f'_\sigma}{f_\sigma}\,(\cdots) \sum_{j=0}^{\infty} \alpha_j B^{j+u} \sum_{k=0}^{\infty} \alpha'_k B^{k+v-p} z_t \Big\}\ ,$$

$u = 1, \dots, r$, $v = p+1, \dots, p+r'$ with $h(x) = f'(x)/f(x)$. We can then easily see that

$$E\left(\frac{\partial^2 g_t}{\partial \phi_u \partial \theta_v}\right) = -C1 \sum_j \alpha_j \alpha'_{j+u-v+p} = -\sigma_{u,v}\ ,\ u = 1, \dots, r,\ v = p+1, \dots, p+r'\ .$$

5. Asymptotics

Our purpose is to show that there is a sequence of solutions $\hat{\boldsymbol{\eta}}_n$ to the approximate likelihood equations

$$\frac{\partial L_q(\boldsymbol{\eta})}{\partial \eta_u} = 0\ , \qquad u = 1, \dots, p+p'+1\ , \tag{5.1}$$

where

$$L_q(\boldsymbol{\eta}) = \sum_{t=q}^{n-q} \log\ f_\sigma(z_t(q)) + (n-2q)\,[\log|\phi_p| - \log|\theta_{p'}|]\ , \tag{5.2}$$

$q = q(n) \to \infty$ as $n \to \infty$ but $q(n) = o(n)$, that is consistent and asymptotically efficient in the sense that $n^{1/2}(\hat{\boldsymbol{\eta}}_n - \boldsymbol{\eta}) \xrightarrow{d} N(0, \Sigma^{-1})$ with Σ the Fisher information matrix given by (1.5). The approximate likelihood function $L_q(\boldsymbol{\eta})$ is actually fully given in terms of the observations x_t, $t = 1, \dots, n$. As an intermediate step we consider a likelihood like ${}_0L(\boldsymbol{\eta}) = \sum_{t=q}^{n-q} g_t(\boldsymbol{\eta})$ that requires more knowledge than that given by just the observations x_t, $t = 1, \dots, n$ (since in (4.2) we need to know the infinite past and future of the x_t's).

Theorem 2. *Let x_t be a zero mean* ARMA *process of order p, p' having the factorization* (1.2) *with $\{z_t\}$ an independent, identically distributed sequence of nonnormal random variables with mean zero, positive variance and probability* *density* $(1/\sigma)f(z/\sigma)$. *The nonnormal density is assumed to satisfy Assumptions* A *and* B. *Then there is a sequence of solutions* $\hat{\boldsymbol{\eta}}_n$ *of the approximate likelihood equations* (5.1) *that is consistent.*

The discussion of the existence of a sequence of consistent estimators that satisfy $\partial_0 L(\boldsymbol{\eta})/\partial\eta_j = 0$, $j = 1, \ldots, p+p'+1$, will follow that given on page 430 of Lehmann (1983). Initially, we assume s, s' in (1.2) are fixed. Then the parameter space corresponding to the model is

$$\begin{aligned}\theta_{s,s'} = \{\boldsymbol{\eta} \in R^{p+p'+1} : {} & 1 - \phi_1 z - \cdots - \phi_r z^r \neq 0 \text{ and} \\ & 1 - \theta_1 z - \cdots - \theta_{r'} z^{r'} \neq 0 \quad \text{for} \quad |z| \leq 1 ; \\ & 1 - \phi_{r+1} z - \cdots - \phi_p z^s \neq 0 \text{ and} \\ & 1 - \theta_{r'+1} z - \cdots - \theta_{p'} z^{s'} \neq 0 \text{ for } |z| \geq 1 ; \\ & \phi_r, \theta_{r'}, \phi_p, \theta_{p'} \neq 0 , \ \eta_{p+p'+1} = \sigma > 0\} .\end{aligned}$$

Let $\boldsymbol{\eta}_0 = (\eta_{0,1}, \ldots, \eta_{0,p+p'+1}) \in \theta_{s,s'}$ be the true parameter value. Set $Q_\epsilon = \{\boldsymbol{\eta} \in R^{p+p'+1} : |\boldsymbol{\eta} - \boldsymbol{\eta}_0| \leq \epsilon\}$, where $|\cdot|$ is the maximum norm on $R^{p+p'+1}$. Now $\theta_{s,s'}$ is open, and given small $\epsilon > 0$ there is a $d < 1$ such that for all $\boldsymbol{\eta} \in Q_\epsilon$,

$$\begin{aligned}\phi^+(z) &= 1 - \phi_1 z - \cdots - \phi_r z^r \neq 0 , \\ \theta^+(z) &= 1 - \theta_1 z - \cdots - \theta_{r'} z^{r'} \neq 0 , \qquad \text{for} \quad |z| < d^{-1} , \\ \phi^*(z) &= 1 - \phi_{r+1} z - \cdots - \phi_p z^s \neq 0 , \\ \theta^*(z) &= 1 - \theta_{r'+1} z - \cdots - \theta_{p'} z^{s'} \neq 0 , \quad \text{for} \quad |z| > d,\end{aligned}$$

and for $d < |z| < d^{-1}$

$$\begin{aligned}\phi(z) &= \phi^+(z)\phi^*(z) = 1 - \tilde{\phi}_1 z - \cdots - \tilde{\phi}_p z^p \neq 0 , \\ \theta(z) &= \theta^+(z)\theta^*(z) = 1 - \tilde{\theta}_1 z - \cdots - \tilde{\theta}_{p'} z^{p'} \neq 0 .\end{aligned}$$

There is then a $C > 0$ such that

$$\begin{aligned}&\sup_{\boldsymbol{\eta}\in Q_\epsilon} |\phi_j - \phi_{0,j}| < C\epsilon , \quad j = 1, \ldots, p, \\ &\sup_{\boldsymbol{\eta}\in Q_\epsilon} |\theta_j - \theta_{0,j}| < C\epsilon , \quad j = 1, \ldots, p', \\ &\sup_{\boldsymbol{\eta}\in Q_\epsilon} |\alpha'_j| \leq C d^{|j|} , \quad \sup_{\boldsymbol{\eta}\in Q_\epsilon} |\beta'_j| \leq C d^{|j|} , \quad j = 0, \pm 1, \ldots, \\ &\sup_{\boldsymbol{\eta}\in Q_\epsilon} |\alpha'_j - \alpha'_{0,j}| \leq C\epsilon d^{|j|} , \quad \sup_{\boldsymbol{\eta}\in Q_\epsilon} |\beta'_j - \beta'_{0,j}| \leq C\epsilon d^{|j|} , \quad j = 0, \pm 1, \ldots ,\end{aligned} \tag{5.3}$$

where, for example, $\{\alpha'_j\}$, $\{\beta'_j\}$ are the power series coefficients in (1.3) with parameter $\boldsymbol{\eta}$, and $\{\alpha'_{0,j}\}$, and $\{\beta'_{0,j}\}$ those with parameter $\boldsymbol{\eta}_0$ assuming that the roots of the polynomials $\theta^+(z)$, $\theta^*(z)$ are distinct. If the roots are of multiplicity greater than one, the best one can do is

$$\sup_{\boldsymbol{\eta}\in Q_\epsilon} |\alpha'_j - \alpha'_{0,j}| \,, \quad \sup_{\boldsymbol{\eta}\in Q_\epsilon} |\beta'_j - \beta'_{0,j}| \le C\epsilon^{\frac{1}{m}} d^{|j|} \,, \quad j = 0, \pm 1, \dots,$$

with $m = \max(r', s')$.

Note that

$$\begin{aligned}
&\frac{1}{n-2q}\left({}_0L(\boldsymbol{\eta}) - {}_0L(\boldsymbol{\eta}_0)\right)\\
&= \frac{1}{n-2q}\sum_{j=1}^{p+p'+1} A_j(\boldsymbol{\eta}_0)(\eta_j - \eta_{0,j})\\
&\quad + \frac{1}{2(n-2q)}\sum_{j,k=1}^{p+p'+1} B_{jk}(\boldsymbol{\eta}_0)(\eta_j - \eta_{0,j})(\eta_k - \eta_{0,k})\\
&\quad + \frac{1}{2(n-2q)}\sum_{j,k=1}^{p+p'+1} (B_{jk}(\boldsymbol{\eta}^*) - B_{jk}(\boldsymbol{\eta}_0))(\eta_j - \eta_{0,j})(\eta_k - \eta_{0,k})\\
&= S_1 + S_2 + S_3,
\end{aligned}$$

where $A_j(\boldsymbol{\eta}) = \sum_{t=q}^{n-q}(\partial g_t/\partial\eta_j)(\boldsymbol{\eta})$, $B_{jk}(\boldsymbol{\eta}) = \sum_{t=q}^{n-q}(\partial^2 g_t/\partial\eta_j\partial\eta_k)(\boldsymbol{\eta})$ and $\boldsymbol{\eta}^*$ is on the line segment joining $\boldsymbol{\eta}_0$ and $\boldsymbol{\eta}$. The ergodic theorem implies that

$$S_1 = \sum_{j=1}^{p+p'+1} \frac{1}{n-2q}\sum_{t=q}^{n-q} \frac{\partial g_t(\boldsymbol{\eta}_0)}{\partial \eta_j}(\eta_j - \eta_{0,j}) \to \sum_{j=1}^{p+p'+1} E\,\frac{\partial g_t(\boldsymbol{\eta}_0)}{\partial \eta_j}(\eta_j - \eta_{0,j}) = 0$$

and

$$S_2 = \frac{1}{2}\sum_{j,k=1}^{p+p'+1} \frac{1}{n-2q} B_{jk}(\boldsymbol{\eta}_0)(\eta_j - \eta_{0,j})(\eta_k - \eta_{0,k}) \to -\frac{1}{2}(\boldsymbol{\eta} - \boldsymbol{\eta}_0)'\Sigma\,(\boldsymbol{\eta} - \boldsymbol{\eta}_0)$$

almost surely as $n \to \infty$ where $\Sigma > 0$ is given by (1.5). By a standard elaboration of the corresponding argument in LR (1992) one can show that

$$\overline{\lim_{n\to\infty}} \sup_{\boldsymbol{\eta}\in Q_\epsilon} \frac{1}{n-2q}|B_{jk}(\boldsymbol{\eta}) - B_{jk}(\boldsymbol{\eta}_0)| \to 0$$

almost surely as $\epsilon \downarrow 0$ for $j, k = 1, \dots, p+p'+1$. For this one requires Assumptions A and B. It is now clear that, for ϵ sufficiently small, $\sup\,(S_1 + S_2 + S_3) < 0$ almost surely as $n \to \infty$ with the sup taken over all $\boldsymbol{\eta}$ on the boundary of Q_ϵ. Consequently there exist a $\delta = \delta(\epsilon) > 0$ such that for n large enough

$$\sup_{\boldsymbol{\eta}\in B(Q_\epsilon)} \frac{1}{n}{}_0L(\boldsymbol{\eta}) < \frac{1}{n}{}_0L(\boldsymbol{\eta}_0) - \delta(\epsilon)$$

almost surely with $B(Q_\epsilon)$ the boundary of Q_ϵ. It follows that ${}_0L(\boldsymbol{\eta})$ has a local maximum in the interior of Q_ϵ.

What remains to be shown is that the replacement of z_t by $z_t(q)$ (leading to the replacement of g_t by $g_{t,q}$ where $g_{t,q}(\boldsymbol{\eta}) = \log f_\sigma(z_t(q)) + \log|\eta_p| - \log|\eta_{p+p'}|$ is valid asymptotically.

Consider now the approximate likelihood

$$\begin{aligned}\frac{1}{n-2q}\,L_q(\boldsymbol{\eta}) &= \frac{1}{(n-2q)}\sum_{t=q}^{n-q}\{\log\, f_\sigma(z_t(q)) + \log\,|\phi_p| - \log\,|\theta_{p'}|\}\\ &= \frac{1}{n-2q}\sum_{t=q}^{n-q} g_{t,q}(\boldsymbol{\eta})\ .\end{aligned}$$

From this point on a tilde over a random variable indicates the dependence of the random variable on the parameter $\boldsymbol{\eta}$ while a tildeless random variable will depend on the true model parameter $\boldsymbol{\eta} = \boldsymbol{\eta}_0$. But

$$\begin{aligned}\frac{1}{n-2q}\,\{{}_0L(\boldsymbol{\eta}) - L_q(\boldsymbol{\eta})\} &= \frac{1}{n-2q}\sum_{t=q}^{n-q}\{\log\, f_\sigma(\tilde z_t) - \log\, f_\sigma(\tilde z_t(q))\}\\ &= \frac{1}{n-2q}\sum_{t=q}^{n-q}(\tilde z_t - \tilde z_t(q))\,\frac{f'_\sigma}{f_\sigma}\,(\tilde z_t + \mu_t\{\tilde z_t(q) - \tilde z_t\})\end{aligned}$$

with $0 \le \mu_t \le 1$. Notice also that $|\tilde z_t - \tilde z_t(q)| \le \sum_{|j|\ge q} d^{|j|}|x_{t-j}|$ and, by Assumption B,

$$\begin{aligned}\left|\frac{f'_\sigma}{f_\sigma}\,(\tilde z_t + \mu_t\{\tilde z_t(q) - \tilde z_t\}) - \frac{f'_\sigma}{f_\sigma}\,(z_t)\right| &\le A[1 + |\tilde z_t - z_t + \mu_t\{\tilde z_t(q) - z_t\}|\;|z_t|^k\\ &\qquad + |\tilde z_t - z_t + \mu_t\{\tilde z_t(q) - \tilde z_t\}|^\ell]\ .\end{aligned}$$

The inequality (5.3) implies that

$$\sup_{\boldsymbol{\eta}\in Q_\epsilon} |\tilde z_t - z_t| \le C\,\epsilon^{1/m}\sum_{|j|\ge q} d^{|j|}\,|x_{t-j}|\ .$$

Thus

$$\sup_{\boldsymbol{\eta}\in Q_\epsilon}\frac{1}{n}\,|{}_0L(\boldsymbol{\eta}) - L_q(\boldsymbol{\eta})| \to 0$$

as $n \to \infty$ with probability one and consequently $(1/n)\;L_q(\boldsymbol{\eta})$ for n large enough will almost surely have a local maximum in the interior of Q_ϵ. There is then with probability one a consistent sequence of estimators $\hat{\boldsymbol{\eta}}_n$ satisfying the approximate likelihood equations $\partial L_q(\boldsymbol{\eta})/\partial\eta_j = 0$, $j = 1,\ldots,p+p'+1$.

Theorem 3. *Under the assumptions of Theorem 2 the sequence of solutions* $\hat{\eta}_n$ *of the approximate likelihood equations referred to there is asymptotically normal with mean* $\boldsymbol{\eta}_0$ *and covariance matrix* $n^{-1}\Sigma^{-1}$ *where* Σ *is given by* (1.5).

Consider the equation

$$0 = n^{-1/2}\frac{\partial L_q(\hat{\boldsymbol{\eta}}_n)}{\partial \boldsymbol{\eta}} = n^{-1/2}\sum_{t=q}^{n-q}\frac{\partial g_{t,q}(\boldsymbol{\eta}_0)}{\partial \boldsymbol{\eta}} + n^{-1/2}B_q(\boldsymbol{\eta}^*)n^{1/2}(\hat{\boldsymbol{\eta}}_n - \boldsymbol{\eta}_0)$$

with $B_q(\boldsymbol{\eta})$ the $(p+p'+1)\times(p+p'+1)$ matrix having entries

$$\sum_{t=q}^{n-q}\frac{\partial g_{t,q}(\boldsymbol{\eta})}{\partial \eta_i \partial \eta_j}\ , \quad i,j = 1,\ldots,p+p'+1,$$

and $\boldsymbol{\eta}^*$ on the line segment joining $\hat{\boldsymbol{\eta}}_n$ and $\boldsymbol{\eta}_0$. One can show

$$n^{-1/2}E\left|\sum_{t=q}^{n-q}\left\{\frac{\partial g_{t,q}(\boldsymbol{\eta}_0)}{\partial \boldsymbol{\eta}} - \frac{\partial g_t(\boldsymbol{\eta}_0)}{\partial \boldsymbol{\eta}}\right\}\right| \longrightarrow 0$$

as $n\to\infty$ and $n^{-1/2}\sum_{t=q}^{n-q}\partial g_{t,q}(\boldsymbol{\eta}_0)/\partial\boldsymbol{\eta}$ is asymptotically $N(0,\Sigma)$ as $n\to\infty$. Set $B(\boldsymbol{\eta}) = \{B_{jk}(\boldsymbol{\eta})\ ;\ \ j,k=1,\ldots,p+p'+1\}$. Then $n^{-1}B_q(\boldsymbol{\eta}^*) = n^{-1}B(\boldsymbol{\eta}_0) + n^{-1}\{B(\boldsymbol{\eta}^*) - B(\boldsymbol{\eta}_0)\} + n^{-1}\{B_q(\boldsymbol{\eta}^*) - B(\boldsymbol{\eta}^*)\}$. The desired conclusion follows from the fact that $n^{-1}\{B(\boldsymbol{\eta}^*) - B(\boldsymbol{\eta}_0)\} \longrightarrow 0$, $n^{-1}\{B_q(\boldsymbol{\eta}^*) - B(\boldsymbol{\eta}^*)\} \longrightarrow 0$, $n^{-1}B(\boldsymbol{\eta}_0) \longrightarrow -\Sigma$ in probability as $n\to\infty$.

Theorems 2 and 3 demonstrate the existence of a sequence of consistent and asymptotically normal solutions of the approximate likelihood equations (5.1). Theorem 2 is still valid if the approximate likelihood $L_q(\boldsymbol{\eta})$ depends on s, s', the numbers of zeros lying inside the unit disc in the complex plane. If more than one solution to these equations exists with s, s' known or unknown, the theorem does not tell us which solution to take as the estimator. The plausible candidate to take is clearly the $\tilde{s}, \tilde{s}', \tilde{\boldsymbol{\eta}}_n$ obtained by maximizing $L_q(\boldsymbol{\eta})$ with respect to s, s' and $\boldsymbol{\eta}$. Given appropriate conditions, $\tilde{s}, \tilde{s}^1, \tilde{\boldsymbol{\eta}}_n$ will be consistent and this will imply the asymptotic normality of $\tilde{\boldsymbol{\eta}}_n$. An argument for this follows.

Restrict the parameter space

$$\Omega = \{\boldsymbol{\eta}\in R^{p+p'+1} : \phi_p\theta_{p'} \neq 0,\ \phi(z)\theta(z)\neq 0, \quad \text{for} \quad |z| = 1,\ \sigma > 0\}$$

to any compact subset Ω_c that contains the true parameter point $\boldsymbol{\eta}_0$. Using the same type of argument as that given in Theorem 2 for consistency, one can show that, with probability one,

$$\frac{1}{n-2q}L_q(\boldsymbol{\eta}) \to E_{\boldsymbol{\eta}_0}\log[f_\sigma(M(\boldsymbol{\phi},\boldsymbol{\theta})x_t)|\phi_p/\theta_{p'}|]$$

uniformly on Ω_c. If the limit has a unique point where the global maximum is assumed, then by a usual compactness argument, any global maximum of $(1/(n-2q))L_q(\boldsymbol{\eta})$ must converge to the global maximum of the limit and so the mle is consistent. Once the consistency is established the asymptotic normality is shown following the usual argument.

We remark that the asymptotics for the approximate maximum likelihood estimates of $\{\tilde{\phi}_i\}_{i=1}^{p}$ and $\{\tilde{\theta}_i\}_{i=1}^{p'}$ and σ can be obtained from the theorems by standard techniques. In particular, let $\tilde{\boldsymbol{\eta}} = (\tilde{\eta}_1, \ldots, \tilde{\eta}_p, \tilde{\eta}_{p+1}, \ldots, \tilde{\eta}_{p+p'}, \tilde{\eta}_{p+p'+1}) = (\tilde{\phi}_1, \ldots, \tilde{\phi}_p, \tilde{\theta}_1, \ldots, \tilde{\theta}_p, \sigma)$ in (1.1), then $n^{1/2}(\tilde{\boldsymbol{\eta}}_n - \tilde{\boldsymbol{\eta}}) \xrightarrow{d} N(0, R\Sigma^{-1}R')$ where Σ is given in (1.5) and $R = [\partial\tilde{\eta}_i/\partial\eta_j]$ for $i, j = 1\ldots, p+p'+1$ which is computed from

$$\tilde{\eta}_j = \begin{cases} \phi_j - \sum_{i=1}^{j} \phi_{j-i}\phi_{r+i}, & j = 1, \ldots, r, \\ -\sum_{i=j-r}^{j} \phi_{j-i}\phi_{r+i}, & j = r+1, \ldots, p, \end{cases}$$

$$\tilde{\eta}_{p+j} = \begin{cases} \theta_j - \sum_{i=1}^{j} \theta_{j-i}\theta_{r'+i}, & j = 1, \ldots, r', \\ -\sum_{i=j-r'}^{j} \theta_{j-i}\theta_{r'+i}, & j = r'+1, \ldots, p', \end{cases}$$

$$\tilde{\eta}_{p+p'+1} = \eta_{p+p'+1} = \sigma.$$

We make a few remarks on a possible reparametrization of the model $\phi(B)x_t = \theta(B)z_t$. Introduce $\phi^{**}(z) = \phi^*(z)/\phi_p$, $\theta^{**}(z) = \theta^*(z)/\theta_{p'}$, $\phi_p\theta_{p'} \neq 0$ with the random variables $z_t^* = z_t\theta_{p'}/\phi_p$. Model (1.1) can then be written $\phi^+(B)\phi^{**}(B)x_t = \theta^+(B)\theta^{**}(B)z_t^*$ with parameters

$$\begin{aligned} \boldsymbol{\eta}^* &= (\eta_1^*, \ldots, \eta_{p+p'}^*, \eta_{p+p'+1}^*) \\ &= (\phi_1, \ldots, \phi_r, \phi_{r+1}^*, \ldots, \phi_p^*, \theta_1, \ldots, \theta_{r'}, \theta_{r'+1}^*, \ldots, \theta_{p'}^*, \sigma^*), \end{aligned}$$

where

$$\eta_i^* = \begin{cases} \phi_i, & 1 \le i \le r, \\ \phi_{p+r-i}/\phi_p = \phi_i^*, & r+1 \le i < p, \\ \phi_p^{-1} = \phi_p^*, & i = p, \\ \theta_{i-p}, & p+1 \le i \le p+r', \\ \theta_{p+p'+r'-i}/\theta_{p'} = \theta_{i-p}^*, & p+r'+1 \le i < p+p', \\ \theta_{p'}^{-1} = \theta_{p'}^*, & i = p+p', \\ \sigma|\theta_{p'}/\phi_p|, & i = p+p'+1. \end{cases}$$

If ϕ^* and θ^* are replaced by ϕ^{**} and θ^{**} respectively in (2.1), the transformations that correspond to (2.2) and (2.4) have unit Jacobian. The approximation to the loglikelihood that corresponds to (2.7) is $1/(n-2q)\sum_{t=q+1}^{n-q}\log f_{\sigma^*}(z_t^*)$ with $\phi^+(B)\phi^{**}(B)x_t = \theta^+(B)\theta^{**}(B)z_t^*$. Derivations similar to those of Sections 3, 4 and 5 can be carried out. The relationship between $\boldsymbol{\eta}^*$ and the $\tilde{\phi}$'s, $\tilde{\theta}$'s and σ is more complicated for this than for our initial parameterization.

6. Simulation Methods and Results

For a general possibly noncausal ARMA process $\phi(B)x_t = \theta(B)z_t$ as given by (1.1) we have $x_t = (\theta(B)/(\phi^+(B)\phi^*(B)))\ z_t = \theta(B)[(\phi^{+'}(B)/\phi^+(B)) + (\phi^{*'}(B)/\phi^*(B))]z_t = \theta(B)\left[\sum_{j=0}^{\infty}\overline{\alpha}_j B^j + \sum_{j=1}^{\infty}\overline{\beta}_j B^{-j}\right] z_t = \sum_{j=-\infty}^{\infty}\gamma_j z_{t-j}$ for some γ_j's where $\phi^{+'}(B)$ and $\phi^{*'}(B)$ are determined by a partial fraction expansion and

$$\frac{\phi^{+'}(B)}{\phi^+(B)} = \phi^{+'}(B)\sum_{j=0}^{\infty}\alpha_j B^j = \sum_{j=0}^{\infty}\overline{\alpha}_j B^j,$$

$$\frac{\phi^{*'}(B)}{\phi^*(B)} = \phi^{*'}(B)\sum_{j=s}^{\infty}\beta_j B^{-j} = \sum_{j=1}^{\infty}\overline{\beta}_j B^{-j}\ .$$

Alternatively, one can consider $\sum_{j=-\infty}^{\infty}\gamma_j B_{t-j}$ as the Laurent series expansion of $\theta(B)/\phi(B)$. We note that $\gamma_j \to 0$ exponentially as $|j|\to\infty$. We approximate x_t by a truncation $\sum_{j=-M}^{M}\gamma_j z_{t-j}$ for large M. Given $\phi(B)$ and $\theta(B)$ to obtain $x_1,\ldots,x_n$ satisfying (1.1) we generate independent identically distributed z_t for $t=-M, -M+1,\ldots,n+M$ and then use $x_t = \sum_{j=-M}^{M}\gamma_j z_{t-j}$ for $t=1,\ldots,n$. In our examples we choose $M=50$ and the model ARMA $(1,1)$

$$x_t - \phi x_{t-1} = z_t - \theta z_{t-1}\ . \tag{6.1}$$

The distributions used for the input z_t are the Laplace (double exponential) distribution and the Student's t distribution with four degrees of freedom. The input scale factor σ is set equal to 1.0. The computation of the approximate log likelihood is given by (5.2) where $L_q(\phi,\theta,\sigma)$ depends on s and s', the number of zeros of $\phi(B)$ and $\theta(B)$ inside the unit circle respectively. The maximization of $L_q(\phi,\theta,\sigma)$ is carried out by a searching procedure. We used two different sample sizes $n=50$ and 100 in the simulation for $\{x_t\}_{t=1}^n$. When the function of z_t is the Laplace density

$$f_\sigma(x) = \frac{1}{\sigma}\, f\left(\frac{x}{\sigma}\right) = \frac{1}{2\sigma}\, e^{-|x|/\sigma}, \tag{6.2}$$

the maximum likelihood estimate of σ can be expressed in the closed form as a function of $\{x_t\}$ and the parameters in $\phi(z)$ and $\theta(z)$. This procedure is used in the computations when we assume that the input density function is known to be Laplacian. This is also used in the surface perspective and contour plot of Figures 1 and 2 which show the function $L_q(\phi,\theta,\sigma)$ with $n=100$ and $q=10$ in

:/a6n1/a920743.ps

θ ϕ

Figure 1. Approximate log-likelihood surface plot. The grid mesh size is 0.05 with $\phi = 0.9$, $\theta = -1.1$. The range is rescaled to 1.0 for plotting.

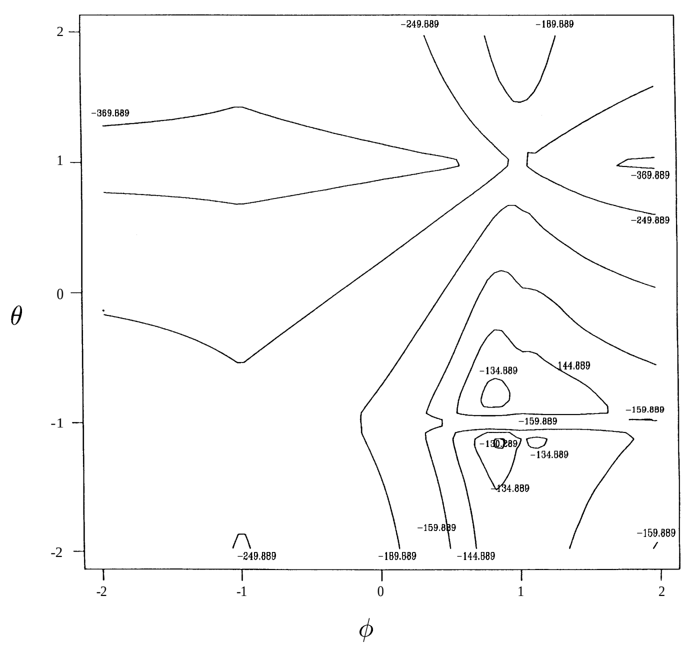

Figure 2. Approximate log-likelihood contour plot. The maximum is at $\hat{\phi} = 0.8745$, $\hat{\theta} = -1.125$ with the value -129.89.

the model (6.1) with $\phi = 0.9$ and $\theta = -1.1$ and $\{z_t\}$ independent identically distributed with density (6.2) and $\sigma = 1.0$. We then get $(\hat{s}, \hat{s}') = (0, 1)$, $\hat{\phi}_1 = 0.8745$, $\hat{\theta}_1 = -1.125$ and $\hat{\sigma} = 0.846$. The likelihood surface is plotted with a grid size 0.05 and $\phi, \theta \in (-1.975, 1.975)$. There are four local maxima in a neighborhood of $(1, -1)$.

For each of the following models, we applied the method just described to each of 200 independently generated time series of length n and recorded the number of times n_1 among the 200 independent repetitions that the procedure identified the correct s and s'. Sample means $(\overline{\phi}, \overline{\theta}, \overline{\sigma})$ and standard deviations (SD) of these n_1 estimates $(\hat{\phi}, \hat{\theta}, \hat{\sigma})$ were computed to assess the accuracy and were compared with the asymptotic standard deviations (ASD). Results of these simulations are given in Tables 2 and 3 where f indicates the density function used in Equation (5.2) for the calculation while the actual density function of $\{z_t\}$ used in the simulation of $\{x_t\}$ is indicated in the second row of the tables. Sample size of $\{x_t\}$ is indicated by n. The correct number of times (s, s') is correctly identified in 200 independent runs is given by n_1. These estimates are used to compute the sample mean $(\overline{\phi}, \overline{\theta}, \overline{\sigma})$ and sample standard deviations (indicated by SD). Asymptotic standard deviations (ASD) are computed from (1.5). The truncation is $q = 10$. For a sample size $n = 50$, the effective length in (5.2) is $n - 2q = 30$. From these tables it is seen that the identification of (s, s') is moderately good even for $n = 50$ and as the sample size increases the accuracy increases also. Estimates of parameters are quite accurate when compared with the corresponding standard deviation. Estimated standard deviations are in close agreement with the asymptotics also. When the roots are moved farther away from the unit circle the general accuracy is further improved as expected. Other cases with parameter combinations of (ϕ^{-1}, θ), (ϕ, θ^{-1}) and (ϕ^{-1}, θ^{-1}) were carried out also. They all exhibit qualitatively the same features as indicated in Tables 2 and 3. When the true density function of the input z_t is unknown, one could use least absolute deviation (which we will be formally led to by assuming the Laplace density function which strictly speaking does not satisfy all conditions in A). Results using Laplace density when the input has a t-distribution are given in the tables also. They are comparable to the results obtained when the true density is known.

We make a few additional remarks on computation. An approximation of z_t for the range $t = q, \ldots, n - q$ can be computed recursively as follows. Since $\xi_t = \theta^*(B) z_t$ we have

$$z_t = -\theta_{p'}^{-1} (\theta_{p'-1} z_{t+1} + \cdots + \theta_{r'+1} z_{t+s'-1} - z_{t+s'} + \xi_{t+s'}) \,. \tag{6.3}$$

Since $\theta^*(B)$ is minimum phase this recursion backward in time is stable and initial values will be washed out in time. For $n \geq t \geq 1$ we can approximate

z_t using (6.3) and setting initial values $\{z_{n+s'}, \ldots, z_{n+1}\}$ equal to zero if ξ_t's are available for $n + s' \geq t \geq s'$. We obtain ξ_t's by the following recursion. From $\theta^+(B)\xi_t = \phi(B)x_t$ we have

$$\xi_t = \theta_1\xi_{t-1} + \cdots + \theta_{r'}\xi_{t-r'} + x_t - \phi x_{t-1} - \cdots - \phi_p x_{t-p} .$$

This can be computed for $n \geq t \geq 1$ with initial values $\xi_{-r} = \cdots = \xi_0 = 0 = x_0 = \cdots = x_{-p}$ and the recursion is stable since $\theta^+(B)$ is minimum phase. This may save some numerical computation.

Table 2. $\phi = 0.9$, $\theta = -1.1$, $\sigma = 1.0$

	$f \sim$ Laplace				$f \sim t(4)$	
	$z_t \sim$ i.i.d. Laplace		$z_t \sim$ i.i.d. $t(4)$		$z_t \sim$ i.i.d. $t(4)$	
n	50	100	50	100	50	100
n_1	100	139	81	112	92	111
$\bar{\phi}$	0.8525	0.8779	0.8040	0.8616	0.8288	0.8662
SD	0.0807	0.0459	0.1068	0.0153	0.1071	0.0456
ASD	0.0799	0.0489	0.0951	0.0582	0.0951	0.0582
$\bar{\theta}$	−1.2250	−1.1753	−1.2275	−1.1625	−1.2005	−1.1545
SD	0.1215	0.0645	0.1312	0.0673	0.1049	0.0617
ASD	0.0925	0.0566	0.1099	0.0673	0.1099	0.0673
$\bar{\sigma}$	0.9071	0.9049	0.8863	0.9726	0.9652	0.9775
SD	0.1531	0.0998	0.1556	0.1056	0.1515	0.1037
ASD	0.2010	0.1231	0.1979	0.1212	0.1979	0.1212

Table 3. $\phi = 0.8$, $\theta = -1.2$, $\sigma = 1.0$

	$f \sim$ Laplace				$f \sim t(4)$	
	$z_t \sim$ Laplace		$z_t \sim t(4)$		$z_t \sim t(4)$	
n	50	100	50	100	50	100
n_1	135	153	95	127	115	145
$\bar{\phi}$	0.7617	0.7773	0.7182	0.7715	0.7426	0.7689
SD	0.1011	0.0567	0.1171	0.0646	0.1274	0.0616
ASD	0.1118	0.0685	0.1349	0.0826	0.1349	0.0826
$\bar{\theta}$	−1.2485	−1.2372	−1.2584	−1.2170	−1.2357	−1.2069
SD	0.1647	0.0722	0.1710	0.0905	0.1411	0.0760
ASD	0.1483	0.0908	0.1790	0.1096	0.1790	0.1096
$\bar{\sigma}$	0.9587	0.9659	0.9397	0.9951	0.9861	1.0097
SD	0.1849	0.1050	0.1783	0.1178	0.1537	0.1218
ASD	0.2205	0.1350	0.2268	0.1389	0.2268	0.1389

Acknowledgement

We wish to thank a referee for helpful remarks leading to an improved version of Section 2. The research of the first author was supported by ONR grant N00014-92-J-1086. The research of the second author was supported by ONR grant N00014-90-J-1371.

References

Bartelt, H., Lohmann, A. and Wirnitzer, B. (1984). Phase and amplitude recovery from bispectra. Appl. Optics **23**, 3121-3129.

Breidt, F. J., Davis, R. A., Lii, K. S. and Rosenblatt, M. (1991). Maximum likelihood estimation for noncausal autoregressive processes. J. Multivariate. Anal. **36**, 175-198.

Donoho, D. (1981). On minimum entropy deconvolution in Applied Time Series Analysis (Edited by D. F. Findley), 565-608.

Kay, S. and Sengupta, D. (1991). Recent advances in non-Gaussian autoregressive process. Advances in Spectrum Analysis and Array Processing (Edited by S. Haykin), 141-210, Prentice-Hall.

Kreiss, J.-P. (1987). On adaptive estimation in stationary ARMA processes. Ann. Statist. **15**, 112-133.

Lehmann, E. L. (1983). Theory of Point Estimation. John Wiley, New York.

Lii, K. S. and Rosenblatt, M. (1982). Deconvolution and estimation of transfer function phase and coefficients for nonGaussian linear processes. Ann. Statist. **10**, 1195-1208.

Lii, K.-S. and Rosenblatt, M. (1992). An approximate maximum likelihood estimation for nonGaussian non-minimum phase moving average processes. J. Multivariate. Anal. **43**, 272-299.

Lohmann, A., Weigelt, G. and Wirnitzer, B. (1983). Speckle masking in astronomy: triple correlation theory and applications. Appl. Optics **23**, 3121-3129.

Mendel, J. M. (1991). Tutorial on higher-order statistics (spectra) in signal processing and systems theory: theoretical results and some applications. Proc. IEEE **79**, 278-305.

Rosenblatt, M. (1985). Stationary Sequences and Random Fields. Birkhäuser, Boston.

Wiggins, R. A. (1978). Minimum entropy deconvolutions. Geoexpoloration **17**, 21-35.

Wiggins, R. A. (1985). Entropy guided deconvolution. Geophysics **50**, 2720-2716.

Department of Statistics, University of California, Riverside, CA 92521, U.S.A.

Department of Mathematics, University of California, San Diego, La Jolla, CA 92093, U.S.A.

(Received September 1992; accepted December 1994)