JOURNAL OF MULTIVARIATE ANALYSIS 36, 175–198 (1991)

Maximum Likelihood Estimation for Noncausal Autoregressive Processes

F. JAY BREIDT* AND RICHARD A. DAVIS*

Colorado State University

KEH-SHIN LII†

University of California, Riverside

AND

MURRAY ROSENBLATT†

University of California, San Diego

Communicated by the Editors

We discuss a maximum likelihood procedure for estimating parameters in possibly noncausal autoregressive processes driven by i.i.d. non-Gaussian noise. Under appropriate conditions, estimates of the parameters that are solutions to the likelihood equations exist and are asymptotically normal. The estimation procedure is illustrated with a simulation study for AR(2) processes. © 1991 Academic Press, Inc.

1. INTRODUCTION

In this paper we discuss maximum likelihood estimation for possibly noncausal autoregressive (AR) processes. We assume that $\{X_t\}$ satisfies the difference equations

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

where $\{Z_t\}$ is an independent and identically distributed (i.i.d.) sequence of random variables with mean zero, variance σ^2, and common probability

Received October 31, 1989; revised June 26, 1990.

AMS 1980 subject classifications: primary, 62M09, secondary, 62E20, 60F05, 62F10.

Key words and phrases: maximum likelihood estimates, asymptotic normality, autoregressive process, nonminimum phase, noncausal.

* Research supported by NSF Grant 8802559.

† Research supported by ONR Contract N00014-85-k-0468 and Grant N00014-90-J-1371 and NSF Grant DMS 83-12106.

0047-259X/91 $3.00

683/36/2-3

density function f_σ, σ a scale parameter. A unique stationary solution to these difference equations exists provided the autoregressive polynomial $\phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p$ has no roots on the unit circle ($\phi(z) \neq 0$ for $|z| = 1$). This solution is said to be *causal* (or *minimum phase*) if $\phi(z)$ has no roots inside the unit circle ($\phi(z) \neq 0$ for $|z| \leqslant 1$), since then X_t can be expressed as a function of only the present and past of the noise process, $Z_t, Z_{t-1}, \ldots$. The solution is said to be *noncausal* (or *nonminimum phase*) if $\phi(z)$ has any roots inside the unit circle. More specifically, we will say that $\{X_t\}$ is *purely noncausal* if $\phi(z)$ has all of its roots inside the unit circle ($\phi(z) \neq 0$ for $|z| \geqslant 1$); in this case X_t is a function of only the future of the noise process, $Z_{t+1}, Z_{t+2}, \ldots$. Finally, if $\phi(z)$ has roots both inside and outside the unit circle, we will say that the solution $\{X_t\}$ is *mixed* in the sense that X_t is then a function of both the future and the past of the noise process. These ideas are made precise in Section 2.

Now if $\{X_t\}$ is a noncausal AR(p) driven by the i.i.d. sequence $\{Z_t\}$ with mean zero and variance σ^2, then $\{X_t\}$ can be reexpressed as a causal (or purely noncausal) AR(p) driven by a new white noise sequence $\{\tilde{Z}_t\}$ with mean zero and variance $\tilde{\sigma}^2$ (Brockwell and Davis [3, p. 125]). (Note that in the non-Gaussian case, $\{\tilde{Z}_t\}$ is uncorrelated, but not independent (Breidt and Davis [2].) In any of these representations, the second-order structure of $\{X_t\}$—namely its autocovariance function—remains unchanged. Thus any estimation method based solely on the second-order properties of the system will be unable to distinguish among causal and noncausal models. In particular, moment estimation techniques such as Yule–Walker estimation will always yield causal models.

Nonidentifiability of causal and noncausal models also appears in Gaussian maximum likelihood estimation. Classically in time series analysis, estimation of the parameters has been carried out for causal models using a Gaussian likelihood (Rosenblatt [10], Brockwell and Davis [3]). Since in the Gaussian case the probabilistic structure of $\{X_t\}$ is wholly determined by its autocovariance function, causal and noncausal systems cannot be distinguished, and so it is conventional to restrict the parameter space to the causal region. On the other hand, for $\{Z_t\}$ non-Gaussian, causal and noncausal models are identifiable from the likelihood function.

We propose a maximum likelihood procedure for estimating σ and the parameters of the autoregressive polynomial in possibly noncausal AR(p) processes driven by i.i.d. noise with mean zero, variance σ^2, and common probability density function f_σ, σ a scale parameter. We show that under appropriate conditions estimates of these parameters which are solutions to the likelihood equations exist and are asymptotically normal, and we derive the form of the asymptotic covariance matrix. Inherent in the estimation procedure is the identification of the order of causality of the model—whether the process $\{X_t\}$ will be modeled as causal, purely noncausal, or mixed. We demonstrate the effectiveness of this procedure with a simulation study for AR(2) processes.

A natural criticism is that the practicality of this estimation procedure is limited by the assumption that the probability density function of the noise process $\{Z_t\}$ is known to within the value of a scale parameter. Of course, this criticism applies equally to much of classical parametric inference. It seems that a reasonable first step is to consider the case of f_σ known. Consideration of the case of f_σ unknown is the next step, which we are currently taking. It involves developing methods of obtaining reasonably efficient initial estimates of the order of causality and of the parameters in the model. Such estimates might then be used in an adaptive estimation procedure, like those described by Beran [1] and Kreiss [6]. One approach to obtaining these estimates is based on the use of higher order cumulant spectra (Lii and Rosenblatt [8], Nikias and Raghuveer [9]). Preliminary simulation results indicate that maximizing an appropriate non-Gaussian likelihood, such as the likelihood obtained when Z_1 has a Laplace density, may be a more efficient approach; these results are discussed briefly in Section 4. This next step in the problem, however, is not our focus here.

2. Approximating the Likelihood

In this section we derive an approximation to the likelihood of a possibly noncausal AR(p) process and calculate the asymptotic covariance matrix of the partial derivatives of the likelihood.

Let $\{X_t\}$ be the mean zero AR(p) process satisfying the difference equations

$$X_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p} = Z_t, \tag{2.1}$$

where $\phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p \neq 0$ for $|z| = 1$, $\phi_p \neq 0$, and $\{Z_t\}$ is an i.i.d. sequence of random variables with mean zero, variance σ^2, and common probability density function f_σ, σ a scale parameter. Specifically, we assume $f_\sigma(x) = \sigma^{-1} f(x/\sigma)$ for some probability density function (pdf) $f(x)$. It is well known (see, for example, Brockwell and Davis [3, p. 88]) that there exists a unique stationary solution to (2.1) given by the two-sided moving average

$$X_t = \sum_{j=-\infty}^{\infty} \psi_j Z_{t-j},$$

where ψ_j is the coefficient of z^j in the Laurent series expansion of $1/\phi(z)$, viz.,

$$\phi(z)^{-1} = \sum_{j=-\infty}^{\infty} \psi_j z^j \tag{2.2}$$

which exists in some annulus $d < |z| < d^{-1}$, $d < 1$. If $\phi(z) \neq 0$ for $|z| \leqslant 1$, then $\psi_j = 0$ for $j < 0$ and we call $\{X_t\}$ *causal*, since it now is a causal function of $\{Z_t\}$, i.e.,

$$X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j}.$$

On the other hand, if $\phi(z) \neq 0$ for $|z| \geqslant 1$ then

$$X_t = \sum_{j=0}^{\infty} \psi_{-j} Z_{t+j}$$

is a function of the future values of $\{Z_t\}$. We call such a process *purely noncausal*. In this case, the coefficients ψ_j satisfy

$$(1 - \phi_1 z - \cdots - \phi_p z^p)(\psi_0 + \psi_{-1} z^{-1} + \cdots) = 1$$

which implies

$$\psi_0 = \psi_{-1} = \cdots = \psi_{1-p} = 0, \qquad \psi_{-p} = -\phi_p^{-1} \tag{2.3}$$

so that X_t is independent of Z_s, $s \leqslant t + p - 1$.

In the causal and purely noncausal cases, it is rather straightforward to approximate the likelihood by the conditional likelihood (conditional on the first p observations in the causal case and the last p observations in the noncausal case). However, in the *mixed case*, when $\phi(z)$ has zeros lying inside and outside the unit circle, approximating the likelihood is more difficult since X_t now depends on both the future and past values of $\{Z_t\}$. To handle the mixed case, we first reparameterize the model by decomposing the autoregressive polynomial, $\phi(z)$, into its causal and purely noncausal components, and then analyze the corresponding AR processes that arise from this decomposition.

Factor the autoregressive polynomial as

$$\phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p = \phi^\dagger(z)\, \phi^*(z), \tag{2.4}$$

where

$$\phi^\dagger(z) = 1 - \theta_1 z - \cdots - \theta_r z^r \neq 0 \qquad \text{for} \quad |z| \leqslant 1,$$
$$\phi^*(z) = 1 - \theta_{r+1} z - \cdots - \theta_p z^s \neq 0 \qquad \text{for} \quad |z| \geqslant 1,$$

and $r, s \geqslant 0$, with $r + s = p$. In other words, if $m_1, ..., m_r, m_{r+1}, ..., m_p$ are the p zeros of $\phi(z)$ with $|m_i| > 1$, $i = 1, ..., r$, and $|m_i| < 1$, $i = r+1, ..., p$, then

$$\phi^\dagger(z) = \prod_{i=1}^{r} (1 - m_i^{-1} z) \qquad \text{and} \qquad \phi^*(z) = \prod_{i=r+1}^{p} (1 - m_i^{-1} z) \tag{2.5}$$

and the ϕ_j's can be determined from the θ_j's through the equations

$$\phi_j = \begin{cases} \theta_j - \sum_{i=1}^{j} \theta_{j-i}\theta_{r+i}, & j = 1, ..., r, \\ -\sum_{i=j-r}^{j} \theta_{j-i}\theta_{r+i}, & j = r+1, ..., p, \end{cases} \tag{2.6}$$

where we set $\theta_0 = -1$ and $\theta_j = 0$ whenever $j \notin \{0, ..., p\}$.

Now define the causal and purely noncausal AR processes by

$$U_t = \phi^*(B)X_t \qquad \text{and} \qquad V_t = \phi^\dagger(B)X_t,$$

respectively, where B is the backwards shift operator $(B^kX_t = X_{t-k}$, $k = 0, \pm 1, ...)$. Since $\phi^\dagger(B)\,\phi^*(B)X_t = Z_t$,

$$\phi^\dagger(B)\,U_t = Z_t \qquad \text{and} \qquad \phi^*(B)\,V_t = Z_t,$$

and hence

$$U_t = \sum_{j=0}^{\infty} \alpha_j Z_{t-j}, \qquad V_t = \sum_{j=s}^{\infty} \beta_j Z_{t+j}, \tag{2.7}$$

where

$$\phi^\dagger(z)^{-1} = \sum_{j=0}^{\infty} \alpha_j z^j \qquad \text{and} \qquad \phi^*(z)^{-1} = \sum_{j=s}^{\infty} \beta_j z^{-j}. \tag{2.8}$$

Since U_t is independent of V_{t-s+1}, the pdf of the random vector $(U_1, ..., U_n, V_{n-s+1}, ..., V_n)'$ is

$$h_U(U_1, ..., U_r) \prod_{t=r+1}^{n} f_\sigma(U_t - \theta_1 U_{t-1} - \cdots - \theta_r U_{t-r})\, h_V(V_{n-s+1}, ..., V_n),$$

where h_U and h_V are the joint pdf's of $(U_1, ..., U_r)'$ and $(V_{n-s+1}, ..., V_n)'$, respectively. The joint pdf of $(U_1, ..., U_s, X_1, ..., X_n)'$ is obtained via the transformation

$$\begin{bmatrix} U_1 \\ \vdots \\ U_s \\ U_{s+1} \\ \vdots \\ U_n \\ V_{n-s+1} \\ \vdots \\ V_n \end{bmatrix} = \begin{bmatrix} U_1 \\ \vdots \\ U_s \\ X_{s+1} - \theta_{r+1}X_s - \cdots - \theta_p X_1 \\ \vdots \\ X_n - \theta_{r+1}X_{n-1} - \cdots - \theta_p X_{n-s} \\ X_{n-s+1} - \theta_1 X_{n-s} - \cdots - \theta_r X_{n-s+1-r} \\ \vdots \\ X_n - \theta_1 X_{n-1} - \cdots - \theta_r X_{n-r} \end{bmatrix} = T \begin{bmatrix} U_1 \\ \vdots \\ U_s \\ X_1 \\ \vdots \\ X_n \end{bmatrix},$$

where T is an $(n+s)\times(n+s)$ matrix. The joint pdf of $(U_1, ..., U_s, X_1, ..., X_n)'$ is then

$$h_U(U_1, ..., U_r)\left(\prod_{t=r+1}^{n} f_\sigma(X_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p})\right)$$
$$\times h_V(\phi^\dagger(B)X_{n-s+1}, ..., \phi^\dagger(B)X_n)\ |\det T|,$$

where U_j is replaced by $\phi^*(B)X_j$ for $s < j \leqslant r$. From the form of the transformation, it follows for $s > 0$ that $\ln |\det T| \sim \ln |\theta_p|^{n-p}$ which suggests approximating the log-likelihood by

$$L(\theta_1, ..., \theta_{p,\sigma}) = \sum_{t=p+1}^{n} (\ln f_\sigma(U_t - \theta_1 U_{t-1} - \cdots - \theta_r U_{t-r}) + \ln |\theta_p|)$$
$$= \sum_{t=p+1}^{n} g_t(\boldsymbol{\theta}), \tag{2.9}$$

where

$$g_t(\boldsymbol{\theta}) = \ln f_\sigma(U_t - \theta_1 U_{t-1} - \cdots - \theta_r U_{t-r}) + \ln |\theta_p|$$
$$= \ln f_\sigma(V_t - \theta_{r+1} V_{t-1} - \cdots - \theta_p V_{t-s}) + \ln |\theta_p|$$

and $\boldsymbol{\theta} = (\theta_1, ..., \theta_{p+1})'$ denotes the parameter vector with $\theta_{p+1} = \sigma$. Note that in the causal case ($s = 0$), the $\ln |\theta_p|$ term does not appear in (2.9).

In order to calculate the asymptotic covariance matrix of the partials of L, we make the following assumptions on f:

A1. $f(x) > 0$ for all x.

A2. $f \in C^2$.

A3. $f' \in L^1$ with $\int f'(x)\, dx = f(x)|_{-\infty}^{\infty} = 0$.

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

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

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

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

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

Evaluating the partials of g_t at the true values, we obtain

$$\frac{\partial g_t}{\partial \theta_i} = \begin{cases} -U_{t-i}\dfrac{f_\sigma'(Z_t)}{f_\sigma(Z_t)}, & i = 1, ..., r \\[2ex] -V_{t+r-i}\dfrac{f_\sigma'(Z_t)}{f_\sigma(Z_t)}, & i = r+1, ..., p-1 \\[2ex] -V_{t-s}\dfrac{f_\sigma'(Z_t)}{f_\sigma(Z_t)} + \dfrac{1}{\theta_p}, & i = p \\[2ex] -\sigma^{-1}\left(Z_t \dfrac{f_\sigma'(Z_t)}{f_\sigma(Z_t)} + 1\right), & i = p+1. \end{cases} \tag{2.10}$$

The assumptions on f imply

$$EZ_s \frac{f_\sigma'(Z_t)}{f_\sigma(Z_t)} = \begin{cases} 0, & \text{if} \quad s \neq t, \\ -1, & \text{if} \quad s = t, \end{cases} \tag{2.11}$$

and hence, using the representation of (2.7) and applying (2.3) to the β_j sequence, we have

$$E \frac{\partial g_t}{\partial \theta_i} = 0, \qquad i = 1, ..., p+1. \tag{2.12}$$

Next, we determine the limiting covariance matrix of $(n-p)^{-1/2} \sum_{t=p+1}^{n} (\partial g_t / \partial \boldsymbol{\theta})$. From (2.11) and the assumptions on f, we have

$$\operatorname{Cov}\left(Z_{t-i} \frac{f_\sigma'(Z_t)}{f_\sigma(Z_t)}, Z_{k-j} \frac{f_\sigma'(Z_k)}{f_\sigma(Z_k)} \right) = \begin{cases} \sigma^2 \tilde{J}, & \text{if} \quad t = k, i = j = 0, \\ \sigma^2 \tilde{I}, & \text{if} \quad t = k, i = j \neq 0, \\ 1, & \text{if} \quad t \neq k, i = t-k, j = k-t, \\ 0, & \text{otherwise}, \end{cases} \tag{2.13}$$

where

$$\tilde{I} = \sigma^{-2} \int (f'(x))^2 / f(x) \, dx, \qquad \tilde{J} = \sigma^{-2} \left(\int x^2 (f'(x))^2 / f(x) \, dx - 1 \right).$$

Let $\gamma_U(\cdot)$ and $\gamma_V(\cdot)$ denote the autocovariance functions of $\{U_t\}$ and $\{V_t\}$. Then from the representations of U_t and V_t given in (2.7) together with (2.13), we obtain

$$\operatorname{Cov}\left(U_{t-i} \frac{f_\sigma'(Z_t)}{f_\sigma(Z_t)}, U_{k-j} \frac{f_\sigma'(Z_k)}{f_\sigma(Z_k)} \right) = \begin{cases} \gamma_U(i-j) \tilde{I}, & k = t, 1 \leqslant i \leqslant j \leqslant r, \\ 0, & k \neq t, 1 \leqslant i \leqslant j \leqslant r, \end{cases}$$

$$\operatorname{Cov}\left(V_{t+r-i} \frac{f_\sigma'(Z_t)}{f_\sigma(Z_t)}, V_{k+r-j} \frac{f_\sigma'(Z_k)}{f_\sigma(Z_k)} \right)$$

$$= \begin{cases} \gamma_V(i-j) \tilde{I}, & k = t, r < i \leqslant j \leqslant p, \ i \neq p, \\ \gamma_V(0) \tilde{I} + \beta_s^2 \sigma^2 (\tilde{J} - \tilde{I}), & k = t, \ i = j = p, \\ 0, & k \neq t, \ r \leqslant i \leqslant j \leqslant p, \end{cases}$$

$$\operatorname{Cov}\left(U_{t-i} \frac{f_\sigma'(Z_t)}{f_\sigma(Z_t)}, V_{k+r-j} \frac{f_\sigma'(Z_k)}{f_\sigma(Z_k)} \right)$$

$$= \sum_{b=s}^{\infty} \sum_{a=0}^{\infty} \alpha_a \beta_b \operatorname{Cov}\left(Z_{t-i-a} \frac{f_\sigma'(Z_t)}{f_\sigma(Z_t)}, Z_{k+r-j+b} \frac{f_\sigma'(Z_k)}{f_\sigma(Z_k)} \right)$$

$$= \begin{cases} \alpha_{t-k-i} \beta_{t-k+j-r}, & \text{if} \quad t > k, \ 1 \leqslant i \leqslant r < j \leqslant p, \\ 0, & \text{if} \quad t \leqslant k, \ 1 \leqslant i \leqslant r < j \leqslant p, \end{cases}$$

and

$$\operatorname{Cov}\left(\frac{\partial g_t}{\partial \theta_i}, \frac{\partial g_k}{\partial \theta_{p+1}}\right)=\begin{cases}0, & \text{if } \ i=1, \ldots, p-1 \text{ or } t \neq k, \\ -\theta_p^{-1} \sigma \tilde{J}, & \text{if } \ i=p, t=k, \\ \tilde{J}, & \text{if } \ i=p+1, \ t=k.\end{cases}$$

Also, for $1 \leqslant i \leqslant r$ and $r<j \leqslant p$,

$$\begin{aligned}
(n-p) & \operatorname{Cov}\left(\frac{1}{n-p} \sum_{t=p+1}^n \frac{\partial g_t}{\partial \theta_i}, \frac{1}{n-p} \sum_{k=p+1}^n \frac{\partial g_k}{\partial \theta_j}\right) \\
& =\frac{1}{n-p} \sum_{t=p+1}^n \sum_{k=p+1}^n \operatorname{Cov}\left(\frac{\partial g_t}{\partial \theta_i}, \frac{\partial g_k}{\partial \theta_j}\right) \\
& =\frac{1}{n-p} \sum_{k=p+1}^{n-1} \sum_{t=k+1}^n \alpha_{t-k-i} \beta_{t-k+j-r} \\
& =\frac{1}{n-p} \sum_{k=p+1}^{n-1} \sum_{t=0}^{n-k-i} \alpha_t \beta_{t+i+j-r} \\
& \rightarrow \sum_{t=0}^{\infty} \alpha_t \beta_{t+i+j-r},
\end{aligned}$$

the convergence holding due to the geometric decay of $\{\alpha_t\}$ and $\{\beta_t\}$. Combining the preceding results, we conclude that

$$(n-p)^{-1} \operatorname{Cov}\left(\sum_{t=p+1}^n \frac{\partial g_t}{\partial \boldsymbol{\theta}},\left(\sum_{k=p+1}^n \frac{\partial g_k}{\partial \boldsymbol{\theta}}\right)^{\prime}\right) \rightarrow \Sigma, \tag{2.14}$$

where, if $s>0$, Σ is the symmetric matrix with (i, j) element $(i \leqslant j)$,

$$\sigma_{ij}=\begin{cases}\tilde{I} \gamma_U(i-j), & 1 \leqslant i \leqslant j \leqslant r, \\ \tilde{I} \gamma_V(i-j), & r<i \leqslant j \leqslant p, i \neq p, \\ \tilde{I} \gamma_V(0)+\beta_s^2 \sigma^2(\tilde{J}-\tilde{I}), & i=j=p, \\ \sum_{k=0}^{\infty} \alpha_k \beta_{k+i+j-r}, & 1 \leqslant i \leqslant r<j \leqslant p, \\ -\theta_p^{-1} \sigma \tilde{J}, & i=p, j=p+1, \\ \tilde{J}, & i=j=p+1, \\ 0, & \text{otherwise.}\end{cases} \tag{2.15}$$

If $s=0$ then

$$\sigma_{ij}=\begin{cases}\tilde{I} \gamma_U(i-j), & 1 \leqslant i \leqslant j \leqslant p, \\ \tilde{J}, & i=j=p+1, \\ 0, & \text{otherwise.}\end{cases}$$

Remark. In view of (2.14), we call Σ the information matrix. Note that in the $s>0$ case, only the four elements in the southeast corner of Σ depend on the scale parameter σ. Of course, in the causal case ($r=p$, $s=0$), $\sigma_{p+1,p+1}$ is the only entry which depends on σ.

PROPOSITION 1. *Let f be a non-normal pdf satisfying the conditions* A1–A8. *Then the matrix Σ is positive definite.*

Proof. First partition Σ as

$$\Sigma=\begin{bmatrix} A & B \\ B' & C \end{bmatrix},$$

where A is $r\times r$, C is $(s+1)\times(s+1)$, and B is $r\times(s+1)$. Next consider the random vectors, $\mathbf{R}=(R_1, ..., R_r)'$ and $\mathbf{S}=(S_1, ..., S_{s+1})'$ defined by

$$R_t=U_{-t}\frac{f_\sigma'(Z_0)}{f_\sigma(Z_0)}=\sum_{k=0}^{\infty}\alpha_k Z_{-t-k}\frac{f_\sigma'(Z_0)}{f_\sigma(Z_0)}, \qquad \text{for} \quad t=1, ..., r,$$

$$S_t=\sum_{j=s}^{\infty}\beta_j Z_{t-j}\frac{f_\sigma'(Z_0)}{f_\sigma(Z_0)}+\begin{cases}0, & 1\leqslant t<s,\\ -\theta_p^{-1}, & t=s,\end{cases} \qquad \text{for} \quad t=1, ..., s,$$

and

$$S_{s+1}=\sigma^{-1}\left(Z_0\frac{f_\sigma'(Z_0)}{f_\sigma(Z_0)}+1\right).$$

It is readily verified that the covariance matrices, Σ_{RR} and Σ_{SS}, of $\mathbf{R}$ and $\mathbf{S}$ are equal to A and C, respectively. Also, from (2.13), we have for $j\leqslant s$,

$$\begin{aligned}\operatorname{Cov}(R_i, S_j)&=\sum_{a=0}^{\infty}\sum_{b=s}^{\infty}\alpha_a\beta_b E\left(Z_{-i-a}Z_{j-b}\left(\frac{f_\sigma'(Z_0)}{f_\sigma(Z_0)}\right)^2\right)\\ &=\sum_{a=0}^{\infty}\beta_{a+i+j}\sigma^2\tilde{I}\end{aligned}$$

and

$$\operatorname{Cov}(R_i, S_{s+1})=0,$$

whence

$$\Sigma_{RS}=\operatorname{Cov}(\mathbf{R}, \mathbf{S})=\sigma^2\tilde{I}B.$$

Moreover,

$$\sigma^2\tilde{I}>1, \tag{2.16}$$

since, by the Cauchy–Schwarz inequality and the assumptions on f,

$$1 = \left| E\left[-Z_0 \frac{f_\sigma'(Z_0)}{f_\sigma(Z_0)} \right] \right|^2 \leqslant \sigma^2 \tilde{I}$$

with equality holding if and only if f is normal.

The matrices A and C are positive definite since there is no linear dependence within the vectors $\mathbf{R}$ and $\mathbf{S}$. Thus, to prove positive definiteness of Σ, it suffices to show that

$$C - B'A^{-1}B$$

is positive definite. The matrix $C - (\sigma^2 \tilde{I})^2 B'A^{-1}B$, being the covariance matrix of $\mathbf{S} - \Sigma_{SR}\Sigma_{RR}^{-1}\mathbf{R}$, is non-negative definite and hence for a nonzero vector $\mathbf{a} \in \mathbb{R}^{s+1}$ with $B\mathbf{a} \neq \mathbf{0}$, we have, by (2.16),

$$\mathbf{a}'(C - B'A^{-1}B)\mathbf{a} > \mathbf{a}'(C - (\sigma^2\tilde{I})^2 B'A^{-1}B)\mathbf{a} \geqslant 0.$$

On the other hand, if $B\mathbf{a} = \mathbf{0}$, then

$$\mathbf{a}'(C - B'A^{-1}B)\mathbf{a} = \mathbf{a}'C\mathbf{a} > 0$$

by the positive definiteness of C. This concludes the proof. ∎

PROPOSITION 2. *If f satisfies the assumptions* A1–A8, *then*

$$(n-p)^{-1/2} \sum_{t=p+1}^{n} \frac{\partial g_t}{\partial \boldsymbol{\theta}} \xrightarrow{d} N(\mathbf{0}, \Sigma),$$

where Σ is the matrix given in (2.15).

Proof. By the Cramér–Wold device, it suffices to show that

$$(n-p)^{-1/2} \sum_{t=p+1}^{n} \mathbf{a}' \frac{\partial g_t}{\partial \boldsymbol{\theta}} \xrightarrow{d} N(0, \mathbf{a}'\Sigma\mathbf{a}) \tag{2.17}$$

for all $\mathbf{a} \in \mathbb{R}^{p+1}$. Define the sequence of $(p+1)$ dimensional random vectors $\{\mathbf{Y}_{tm}, t = 0, \pm 1, ...\}$ to be the partials defined in (2.10), but with the U_t's and V_t's replaced by the sums in (2.7) truncated at a large positive integer m. In addition, let Σ_m be the matrix corresponding to Σ, obtained by truncating the U_t's and V_t's. Then the stationary sequence $\{\mathbf{Y}_{tm}, t = 0, \pm 1, ...\}$ is $m+p$ dependent and it follows easily that (2.14) holds with $\partial g_t/\partial \boldsymbol{\theta}$ replaced by $\mathbf{Y}_{tm}$ and limit covariance matrix given by Σ_m. Applying a standard central limit theorem for finite dependent stationary sequences (for example, Theorem 6.4.2 in [3], we obtain

$$(n-p)^{-1/2} \sum_{t=p+1}^{n} \mathbf{a}'\mathbf{Y}_{tm} \xrightarrow{d} N(0, \mathbf{a}'\Sigma_m\mathbf{a}).$$

Now $\Sigma_m \to \Sigma$ as $m \to \infty$ and since

$$\lim_{m\to\infty} \lim_{n\to\infty} \operatorname{Var}\left((n-p)^{-1/2} \sum_{t=p+1}^{n} \left(\mathbf{a}'\mathbf{Y}_{tm} - \mathbf{a}' \frac{\partial g_t}{\partial \boldsymbol{\theta}}\right)\right) = 0,$$

the convergence in (2.17) is immediate from Proposition 6.3.9 in [3]. ∎

3. Asymptotic Normality

In this section, we show that there exists a sequence of solutions, $\hat{\boldsymbol{\theta}}_n$, to the *likelihood equations*,

$$\frac{\partial L(\boldsymbol{\theta})}{\partial \theta_j} = 0, \qquad j = 1, ..., p+1, \tag{3.1}$$

L given in (2.9), which is consistent and asymptotically efficient in the sense that

$$n^{1/2}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}) \xrightarrow{d} N(\mathbf{0}, \Sigma^{-1}),$$

where Σ is the Fisher information matrix computed in (2.15). We temporarily assume that L does not depend on s, the number of zeros of the autoregressive polynomial which lie inside the unit circle. In addition to the assumptions A1–A8 of Section 2, we assume that

$$h'(x) := \frac{d}{dx} h(x) = \frac{d}{dx}\left(\frac{f'(x)}{f(x)}\right) = h_1(x) - h_2(x), \tag{3.2}$$

where h_1 and h_2 are nondecreasing functions with

$$h_i(x) = O(|x|^k) \qquad \text{as} \qquad |x| \to \infty$$

for some $k \geqslant 0$ satisfying

$$E\,|Z_0|^{2+k} < \infty.$$

In particular, this condition implies

$$E\,|Z_0|^j\,|h'(\sigma_0^{-1} Z_0)| \leqslant E\,|Z_0|^j\,|h_1(\sigma_0^{-1} Z_0)| + E\,|Z_0|^j\,h_2(\sigma_0^{-1} Z_0)| < \infty$$

for $j = 0, 1, 2$.

To establish the existence of a consistent sequence of estimators, $\hat{\boldsymbol{\theta}}_n$, satisfying (3.1), we follow the argument given on p. 430 of Lehmann [7]. Since we are assuming here that s is fixed, the parameter space for the model is

$$\begin{aligned} \Omega_s = \{\boldsymbol{\theta} \in \mathbb{R}^{p+1} : 1 - \theta_1 z - \cdots - \theta_r z^r \neq 0 \qquad &\text{for} \quad |z| \leqslant 1, \\ 1 - \theta_{r+1} z - \cdots - \theta_p z^s \neq 0 \qquad &\text{for} \quad |z| \geqslant 1, \\ \theta_r \neq 0, \theta_p \neq 0, \text{ and } \theta_{p+1} = \sigma > 0\}. \end{aligned}$$

Let $\boldsymbol{\theta}_0 = (\theta_{01}, ..., \theta_{0p}, \sigma_0)' \in \Omega_s$ be the true parameter value and let Q_ε denote the closed ball of radius ε centered at $\boldsymbol{\theta}_0$, i.e., $Q_\varepsilon =$

$\{\boldsymbol{\theta} \in \mathbb{R}^{p+1}: |\boldsymbol{\theta} - \boldsymbol{\theta}_0| \leqslant \varepsilon\}$, where $|\cdot|$ represents the max norm on $\mathbb{R}^{p+1}$. The parameter space Ω_s is open and for ε small, there exists a $d < 1$ such that for all $\boldsymbol{\theta} \in Q_\varepsilon$

$$\phi^\dagger(z) = 1 - \theta_1 z - \cdots - \theta_r z^r \neq 0 \qquad \text{for} \quad |z| < d^{-1},$$

$$\phi^*(z) = 1 - \theta_{r+1} z - \cdots - \theta_p z^s \neq 0 \qquad \text{for} \quad |z| > d,$$

and

$$\phi(z) = \phi^\dagger(z)\,\phi^*(z) = 1 - \phi_1 z - \cdots - \phi_p z^p \neq 0 \qquad \text{for} \quad d \leqslant |z| \leqslant d^{-1}.$$

It follows from these relations and (2.5) and (2.6) that there exists a constant $C > 0$ such that

$$\begin{aligned} \sup_{\boldsymbol{\theta} \in Q_\varepsilon} |\phi_j - \phi_{0j}| &\leqslant C\varepsilon, & j &= 1, ..., p, \\ \sup_{\boldsymbol{\theta} \in Q_\varepsilon} |\psi_j| &\leqslant C d^{|j|}, & j &= 0, \pm 1, ..., \\ \sup_{\boldsymbol{\theta} \in Q_\varepsilon} |\psi_j - \psi_{0j}| &\leqslant C\varepsilon d^{|j|}, & j &= 0, \pm 1, ..., \end{aligned} \tag{3.3}$$

where the $\{\psi_j\}$ and $\{\psi_{0j}\}$ are the coefficients in the power series (2.2) with parameter $\boldsymbol{\theta}$ and $\boldsymbol{\theta}_0$, respectively.

Expanding $L(\boldsymbol{\theta})$ in a neighborhood of $\boldsymbol{\theta}_0$, we have

$$\begin{aligned} &\frac{1}{n-p}(L(\boldsymbol{\theta}) - L(\boldsymbol{\theta}_0)) \\ &\quad = \frac{1}{n-p} \sum_{j=1}^{p+1} A_j(\boldsymbol{\theta}_0)(\theta_j - \theta_{0j}) \\ &\qquad + \frac{1}{2(n-p)} \sum_{j=1}^{p+1} \sum_{k=1}^{p+1} B_{jk}(\boldsymbol{\theta}_0)(\theta_j - \theta_{0j})(\theta_k - \theta_{0k}) \\ &\qquad + \frac{1}{2(n-p)} \sum_{j=1}^{p+1} \sum_{k=1}^{p+1} (B_{jk}(\boldsymbol{\theta}^*) - B_{jk}(\boldsymbol{\theta}_0))(\theta_j - \theta_{0j})(\theta_k - \theta_{0k}) \\ &\quad = S_1 + S_2 + S_3, \end{aligned} \tag{3.4}$$

where

$$A_j(\boldsymbol{\theta}) = \sum_{t=p+1}^{n} \frac{\partial g_t(\boldsymbol{\theta})}{\partial \theta_j},$$

$$B_{jk}(\boldsymbol{\theta}) = \sum_{t=p+1}^{n} \frac{\partial^2 g_t(\boldsymbol{\theta})}{\partial \theta_j\, \partial \theta_k},$$

and $\boldsymbol{\theta}^*$ is between $\boldsymbol{\theta}_0$ and $\boldsymbol{\theta}$. By the ergodic theorem and (2.12),

$$S_1 = \sum_{j=1}^{p+1} \frac{1}{n-p} \sum_{t=p+1}^{n} \frac{\partial g_t(\boldsymbol{\theta}_0)}{\partial \theta_j}(\theta_j - \theta_{0j}) \to \sum_{j=1}^{p+1} E\frac{\partial g_1(\boldsymbol{\theta}_0)}{\partial \theta_j}(\theta_j - \theta_{0j}) = 0 \quad \text{a.s.} \tag{3.5}$$

To analyze the terms involving mixed partials, we begin with some preliminaries. In what follows, a tilde over a random variable will indicate dependence of that random variable on the parameter $\boldsymbol{\theta}$, while a tildeless random variable will depend only on the true model ($\boldsymbol{\theta}=\boldsymbol{\theta}_0$). For example,

$$
\begin{aligned}
\tilde{U}_t &= X_t-\theta_{r+1}X_{t-1}-\cdots-\theta_pX_{t-s}, & U_t &= X_t-\theta_{0,r+1}X_{t-1}-\cdots-\theta_{0p}X_{t-s},\\
\tilde{Z}_t &= \tilde{U}_t-\theta_1\tilde{U}_{t-1}-\cdots-\theta_r\tilde{U}_{t-r}, & Z_t &= U_t-\theta_{01}U_{t-1}-\cdots-\theta_{0r}U_{t-r},\\
&= X_t-\phi_1X_{t-1}-\cdots-\phi_pX_{t-p}, & &= X_t-\phi_{01}X_{t-1}-\cdots-\phi_{0p}X_{t-p},\\
\tilde{X}_t &= \sum_{j=-\infty}^{\infty}\psi_jZ_{t-j}, & X_t &= \sum_{j=-\infty}^{\infty}\psi_{0j}Z_{t-j}.
\end{aligned}
\tag{3.6}
$$

It follows easily from (3.3) and (3.6) that

$$
\begin{aligned}
&\sup_{\boldsymbol{\theta}\in Q_\varepsilon}|\tilde{U}_t-U_t|\leqslant\varepsilon(|X_{t-1}|+\cdots+|X_{t-s}|)\\
&\sup_{\boldsymbol{\theta}\in Q_\varepsilon}|\tilde{V}_t-V_t|\leqslant\varepsilon(|X_{t-1}|+\cdots+|X_{t-r}|)\\
&\sup_{\boldsymbol{\theta}\in Q_\varepsilon}|\tilde{X}_t-X_t|\leqslant\sup_{\boldsymbol{\theta}\in Q_\varepsilon}\sum_{j=-\infty}^{\infty}|\psi_j-\psi_{0j}|\,|Z_{t-j}|\leqslant C\varepsilon\sum_{j=-\infty}^{\infty}d^{|j|}|Z_{t-j}|.
\end{aligned}
\tag{3.7}
$$

Also, writing $\tilde{Z}_t=\tilde{Z}_t-Z_t+Z_t$, we have for all $\boldsymbol{\theta}\in Q_\varepsilon$

$$
Z_t-C\varepsilon\left(\sum_{i=1}^{p}|X_{t-i}|\right)\leqslant\tilde{Z}_t\leqslant Z_t+C\varepsilon\left(\sum_{i=1}^{p}|X_{t-i}|\right).
\tag{3.8}
$$

The mixed partials of $g_t(\boldsymbol{\theta})$ are calculated to be

$$
\frac{\partial^2 g_t(\boldsymbol{\theta})}{\partial\theta_j\theta_k}=
\sigma^{-2}\begin{cases}
\tilde{U}_{t-j}\tilde{U}_{t-k}h'(\sigma^{-1}\tilde{Z}_t), & 1\leqslant j\leqslant k\leqslant r,\\
\sigma X_{t-j-k+r}h(\sigma^{-1}\tilde{Z}_t)+\tilde{U}_{t-j}\tilde{V}_{t-k+r}h'(\sigma^{-1}\tilde{Z}_t), & 1\leqslant j\leqslant r<k\leqslant p,\\
\tilde{U}_{t-j}h(\sigma^{-1}\tilde{Z}_t)+\tilde{U}_{t-j}\sigma^{-1}\tilde{Z}_th'(\sigma^{-1}\tilde{Z}_t), & 1\leqslant j\leqslant r,\ k=p+1,\\
\tilde{V}_{t-j+r}\tilde{V}_{t-k+r}h'(\sigma^{-1}\tilde{Z}_t), & r<j\leqslant k\leqslant p,\ j<p,\\
\tilde{V}_{t-s}^2h'(\sigma^{-1}\tilde{Z}_t)-\sigma^2\theta_p^{-2}, & j=k=p,\\
\tilde{V}_{t-j+r}h(\sigma^{-1}\tilde{Z}_t)+\tilde{V}_{t-j+r}\sigma^{-1}\tilde{Z}_th'(\sigma^{-1}\tilde{Z}_t), & r<j\leqslant p,\ k=p+1,\\
\sigma^{-2}\tilde{Z}_t^2h'(\sigma^{-1}\tilde{Z}_t)+2\sigma^{-1}\tilde{Z}_th(\sigma^{-1}\tilde{Z}_t)+1, & j=k=p+1.
\end{cases}
$$

One verifies that

$$
E\frac{\partial^2 g_t(\boldsymbol{\theta}_0)}{\partial\theta_j\,\partial\theta_k}=-\sigma_{jk},
$$

where σ_{jk} is given in (2.15), and hence by the ergodic theorem,

$$S_2 = \frac{1}{2} \sum_{j=1}^{p+1} \sum_{k=1}^{p+1} \frac{1}{n-p} B_{jk}(\boldsymbol{\theta}_0)(\theta_j - \theta_{0j})(\theta_k - \theta_{0k})$$

$$\to -\frac{1}{2} (\boldsymbol{\theta} - \boldsymbol{\theta}_0)' \Sigma (\boldsymbol{\theta} - \boldsymbol{\theta}_0) \qquad \text{a.s.} \tag{3.9}$$

As for S_3, we show

$$\limsup_{n \to \infty} \sup_{\boldsymbol{\theta} \in Q_\varepsilon} \frac{1}{n-p} |B_{jk}(\boldsymbol{\theta}) - B_{jk}(\boldsymbol{\theta}_0)| \to 0 \qquad \text{a.s.} \tag{3.10}$$

as $\varepsilon \to 0$ for $j, k = 1, ..., p+1$. For $1 \leqslant j \leqslant k \leqslant r$, we have

$$\frac{1}{n-p} |B_{jk}(\boldsymbol{\theta}) - B_{jk}(\boldsymbol{\theta}_0)|$$

$$\leqslant \frac{1}{n-p} \sum_{t=p+1}^{n} |\sigma^{-2} \tilde{U}_{t-j} \tilde{U}_{t-k} h'(\sigma^{-1} \tilde{Z}_t) - \sigma_0^{-2} U_{t-j} U_{t-k} h'(\sigma_0^{-1} Z_t)|$$

$$\leqslant \frac{|\sigma^2 - \sigma_0^2|}{\sigma^2 \sigma_0^2} \frac{1}{n-p} \sum_{t=p+1}^{n} |U_{t-j} U_{t-k} h'(\sigma_0^{-1} Z_t)|$$

$$+ \sigma^{-2} \frac{1}{n-p} \sum_{t=p+1}^{n} |(U_{t-j} - \tilde{U}_{t-j})\, U_{t-k} h'(\sigma_0^{-1} Z_t)|$$

$$+ \sigma^{-2} \frac{1}{n-p} \sum_{t=p+1}^{n} |(U_{t-k} - \tilde{U}_{t-k})\, \tilde{U}_{t-j} h'(\sigma_0^{-1} Z_t)|$$

$$+ \sigma^{-2} \frac{1}{n-p} \sum_{t=p+1}^{n} |\tilde{U}_{t-j} \tilde{U}_{t-k} (h'(\sigma_0^{-1} Z_t) - h'(\sigma^{-1} \tilde{Z}_t))|$$

$$= T_1 + T_2 + T_3 + T_4.$$

For $\varepsilon < \sigma_0/2$, we have by the ergodic theorem

$$\sup_{\theta \in Q_\varepsilon} T_1 \leqslant \varepsilon 10 \sigma_0^{-3} \frac{1}{n-p} \sum_{t=p+1}^{n} |U_{t-j} U_{t-k} h'(\sigma_0^{-1} Z_t)|$$

$$\to \varepsilon 10 \sigma_0^{-3} E |U_{-j} U_{-k} h'(\sigma_0^{-1} Z_0)| \qquad \text{a.s.}$$

$$= \varepsilon 10 \sigma_0^{-3} E |U_{-j} U_{-k}|\, E |h'(\sigma_0^{-1} Z_0)| \to 0$$

as $\varepsilon \to 0$. Using the bounds in (3.7) and applying the ergodic theorem once again, we obtain

$$\limsup_{n \to \infty} \sup_{\boldsymbol{\theta} \in Q_\varepsilon} T_2 \leqslant 4 \sigma_0^{-2} \varepsilon E |(|X_{-j-1}| + \cdots + |X_{-j-s}|)\, U_{-k} h'(\sigma_0^{-1} Z_0))| \quad \text{a.s.}$$

$$\to 0 \qquad \text{as} \quad \varepsilon \to 0,$$

and, by a similar argument,

$$\limsup_{n\to\infty} \sup_{\boldsymbol{\theta}\in Q_\varepsilon} T_3 \to 0 \qquad \text{a.s. as } \varepsilon\to 0.$$

As for T_4, set $W_t = \sum_{i=1}^p |X_{t-i}|$ and define for $i = 1, 2$,

$$Y_{t,i}^\varepsilon = \begin{cases} h_i\left(\dfrac{Z_t + C\varepsilon W_t}{\sigma_0 - \varepsilon}\right) - h_i\left(\dfrac{Z_t - C\varepsilon W_t}{\sigma_0 + \varepsilon}\right), & \text{if } Z_t - C\varepsilon W_t > 0,\\ h_i\left(\dfrac{Z_t + C\varepsilon W_t}{\sigma_0 + \varepsilon}\right) - h_i\left(\dfrac{Z_t - C\varepsilon W_t}{\sigma_0 - \varepsilon}\right), & \text{if } Z_t + C\varepsilon W_t < 0,\\ h_i\left(\dfrac{Z_t + C\varepsilon W_t}{\sigma_0 - \varepsilon}\right) - h_i\left(\dfrac{Z_t - C\varepsilon W_t}{\sigma_0 - \varepsilon}\right), & \text{otherwise,} \end{cases}$$

where h_1 and h_2 are the nondecreasing functions specified in (3.2). Then by (3.7), (3.8), and the ergodic theorem,

$$\limsup_{n\to\infty} \sup_{\boldsymbol{\theta}\in Q_\varepsilon} T_4$$
$$\leqslant 4\sigma_0^{-2} E[(|U_{-j}| + \varepsilon W_{-j})(|U_{-k}| + \varepsilon W_{-k})(Y_{0,1}^\varepsilon + Y_{0,2}^\varepsilon)] \quad \text{a.s.} \tag{3.11}$$

Using the inequality

$$\left| h_i\left(\frac{Z_0 \pm C\varepsilon W_0}{\sigma_0 \pm \varepsilon}\right)\right|$$
$$\leqslant A_1 + A_2 \left|\frac{Z_0 \pm C\varepsilon W_0}{\sigma_0 \pm \varepsilon}\right|$$
$$\leqslant A_1 + A_3\left(|Z_0|^k + \sum_{i=1}^p |X_{-i}|^k\right)$$
$$\leqslant A_1 + A_3\left(|Z_0|^k + \sum_{i=1}^p \sum_{i_1} \cdots \sum_{i_k} |\psi_{i_1} \cdots \psi_{i_k}|\, |Z_{-i-i_1} \cdots Z_{-i-i_k}|\right),$$

where A_1, A_2, and A_3 are constants, the expectation in (3.11) is finite, from which it follows by the assumptions on h_i and dominated convergence that the limit of the right hand side in (3.11) is 0 as $\varepsilon \to 0$. This proves (3.10), at least for the $1 \leqslant j \leqslant k \leqslant r$ case. For the other cases, the arguments follow the same ideas as used above and hence are omitted.

Combining the results in (3.5), (3.9), (3.10), and Proposition 1, we conclude that for ε small

$$\sup(S_1 + S_2 + S_3) < 0 \qquad \text{a.s.}$$

as $n \to \infty$, where the sup is taken over $\boldsymbol{\theta}$ on the boundary of Q_ε. Consequently, for n large,

$$L(\boldsymbol{\theta}) < L(\boldsymbol{\theta}_0) \qquad \text{a.s.}$$

for all $\boldsymbol{\theta}$ on the boundary of Q_ε and so $L(\boldsymbol{\theta})$ must have a local maximum on the interior of Q_ε. Such a local maximum must satisfy the likelihood equations. Now as discussed in Lehmann [7], a sequence of local maxima can be chosen, independent of ε, so as to converge a.s. to $\boldsymbol{\theta}_0$.

Having established the existence of a consistent sequence of estimators, $\hat{\boldsymbol{\theta}}_n$, satisfying the likelihood equations (3.1), asymptotic normality of $\hat{\boldsymbol{\theta}}_n$ is practically immediate from Proposition 2. To see this, a Taylor series expansion of $\partial L(\boldsymbol{\theta})/\partial\boldsymbol{\theta}$ about $\boldsymbol{\theta}_0$ gives

$$0 = n^{-1/2}\frac{\partial L(\hat{\boldsymbol{\theta}}_n)}{\partial\boldsymbol{\theta}} = n^{-1/2}\sum_{t=p+1}^{n}\frac{\partial g_t(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}} + n^{-1}B(\boldsymbol{\theta}^*)\,n^{1/2}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0),$$

where $B(\boldsymbol{\theta})$ is the $(p+1)\times(p+1)$ matrix with entries $B_{jk}(\boldsymbol{\theta})$ and $\boldsymbol{\theta}^*$ is on the line segment joining $\boldsymbol{\theta}_0$ and $\hat{\boldsymbol{\theta}}_n$. Since $\boldsymbol{\theta}^* \to \boldsymbol{\theta}_0$ a.s., it follows from (3.10) and the ergodic theorem that

$$n^{-1}B(\boldsymbol{\theta}^*) = n^{-1}B(\boldsymbol{\theta}_0) + n^{-1}(B(\boldsymbol{\theta}^*) - B(\boldsymbol{\theta}_0)) \to -\Sigma$$

a.s. and hence, from Proposition 2, that

$$n^{1/2}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \xrightarrow{d} N(\mathbf{0}, \Sigma^{-1}).$$

We record the preceding results in the following theorem.

THEOREM 3. *Let $\{X_t\}$ be the mean zero* AR(p) *process satisfying*

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

where the autoregressive polynomial $\phi(z)$ has the factorization given in (2.4) *and $\{Z_t\}$ is an i.i.d. sequence of random variables with mean zero, variance σ^2 and common pdf, $\sigma^{-1}f(x/\sigma)$. Further suppose that f is a non-normal pdf which satisfies conditions* A1–A8 *of Section* 2 *and* (3.2). *Then there exists a sequence of solutions, $\hat{\boldsymbol{\theta}}_n$, to the likelihood equations* (3.1) *which is asymptotically normal with mean $\boldsymbol{\theta}_0$ and asymptotic covariance matrix $n^{-1}\Sigma^{-1}$, where Σ is specified in* (2.15).

From this theorem, it is also possible to get the limit distribution for the estimates of the original autoregressive parameters $\phi_1, \ldots, \phi_p$. In this case the resulting estimate, $\hat{\boldsymbol{\phi}}_n$, is computed by replacing the θ_j's in (2.6) by their estimated values. A standard argument shows that

$$n^{1/2}(\hat{\boldsymbol{\phi}}_n - \boldsymbol{\phi}) \xrightarrow{d} N(\mathbf{0}, R\Sigma_p^{-1}R'), \tag{3.12}$$

where Σ_p^{-1} is the $p\times p$ northwest submatrix of Σ^{-1} and

$$R = \left[\frac{\partial\phi_j}{\partial\theta_k}\right]_{j,k=1}^{p}.$$

R is quite trivial to compute from (2.6).

Theorem 3 and (3.12) remain valid if the likelihood function $L(\boldsymbol{\theta})$ in (2.9) is allowed to depend on s, the number of zeros of $\phi(z)$ inside the unit circle. The proof is the same as in the s fixed case, since the key observation is that if $|\boldsymbol{\phi}-\boldsymbol{\phi}_0|<\varepsilon$, ε small, then $\phi(z)$ and $\phi_0(z)$ have the same number of zeros inside the unit circle.

If there happens to be more than one solution to the likelihood equations with s either known or unknown, Theorem 3 does not indicate which of the local maxima to choose as the estimator. The obvious candidate is, of course, the maximum likelihood estimator, $\tilde{s}$, $\tilde{\boldsymbol{\theta}}_n$, found by maximizing $L(\boldsymbol{\theta})$ with respect to s and $\boldsymbol{\theta}$. Subject to mild restrictions, $\tilde{s}$, $\tilde{\boldsymbol{\theta}}_n$ will be consistent, from which the asymptotic normality of $\tilde{\boldsymbol{\theta}}_n$ ensues. An argument for this assertion is as follows.

It will be convenient here to work with the parameters $\boldsymbol{\phi}$ and σ rather than $\boldsymbol{\theta}$ so that the log-likelihood becomes

$$L(\boldsymbol{\phi}, \sigma)=\sum_{t=p+1}^{n} g_t(\phi, \sigma),$$

where

$$g_t(\boldsymbol{\phi}, \sigma)=\ln(f_\sigma(X_t-\phi_1 X_{t-1}-\cdots-\phi_p X_{t-p})\,|\theta_p|)$$

and from (2.5),

$$\theta_p^{-1}=\begin{cases}\prod_{i=r+1}^{p} m_i, & \text{if } s\neq 0,\\ 1, & \text{if } s=0.\end{cases}$$

Restrict the parameter space,

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

to any compact subset, Ω_c, containing the true parameter $\boldsymbol{\phi}_0, \sigma_0$. Now, using the same method of argument given for the consistency part of Theorem 3, it can be shown that with probability one,

$$\begin{aligned}\frac{1}{n-p} L(\boldsymbol{\phi}, \sigma) &\to E g_{p+1}(\boldsymbol{\phi}, \sigma)\\ &= E\ln(f_\sigma(X_{p+1}-\phi_1 X_p-\cdots-\phi_p X_1)\,|\theta_p|)\end{aligned}$$

uniformly on Ω_c. Provided the limit, $Eg_{p+1}(\boldsymbol{\phi}, \sigma)$, has a unique maximum at $\boldsymbol{\phi}_0, \sigma_0$, then, by a standard compactness argument, any maximum of $(1/(n-p))\,L(\boldsymbol{\phi}, \sigma)$ must converge to the maximum of the limit and, therefore, the mle is consistent.

4. Simulation Methods and Results

The standard recursion for simulating n observations from a causal AR(r) process $\{U_t\}$ is to begin far back in the past of the process, say at $t=-k$, to set $U_{-k-1}=U_{-k-2}=\cdots=U_{-k-r}=0$, and then to compute recursively

$$U_t=\phi_1 U_{t-1}+\phi_2 U_{t-2}+\cdots+\phi_r U_{t-r}+Z_t$$

for $t=-k, \quad -k+1, ..., 1, 2, ..., n$. Purely noncausal processes can be generated similarly by reversing the time scale; the recursion begins far forward in the future of the process, say at $t=n+m$, and computes U_t for $t=n+m,\ n+m-1, ..., n, n-1, ..., 1$.

Of course, neither of these recursions will generate a series of mixed causality, since they force exclusive dependence on the past or on the future of the i.i.d. sequence $\{Z_t\}$. To simulate time series of mixed causality, we use the factorization (2.4) and a two-stage recursive procedure. We first simulate $n+m$ observations from the causal AR(r)$\{U_t\}$; as above, we begin at $t=-k$, far back in the past of the process. Then, since $\phi^*(B)X_t=U_t$, we have that

$$X_t=\theta_p^{-1}(X_{t+s}-\theta_{r+1}X_{t+s-1}-\cdots-\theta_{p-1}X_{t+1}-U_{t+s}). \qquad (4.1)$$

We set $X_{n+m+s}=X_{n+m+s-1}=\cdots=X_{n+m+1}=0$, compute (4.1) recursively for $t=n+m,\ n+m-1, ..., n, n-1, ..., 1$, and then retain the last n computed values as the simulated time series $\{X_t\}$, $t=1, 2, ..., n$.

Using the above recursions, we simulated causal and mixed AR(2) processes driven by non-Gaussian noise, emphasizing the mixed case. Noise distributions included the Laplace (two-sided exponential) distribution and the Student's t distribution; the underlying random number generator is discussed in Kahaner, Moler, and Nash [5]. We approximated the log-likelihood of each simulated series as in (2.9) and, allowing $L(\boldsymbol{\theta})$ to depend on s, the number of zeros of $1-\phi_1 z-\phi_2 z^2$ inside the unit circle, maximized this function using a non-linear optimizer described by Dennis and Schnabel [4].

In the Laplace case, the mle of σ can be expressed in closed form as a function of the data and of the parameters in $\phi(z)$. The approximate log-likelihood can thus be reduced to a function of ϕ_1, ϕ_2, and s, or in the parameterization of (2.5), as a function of m_1^{-1}, m_2^{-1}, and s. This parameterization is used in the surface and contour plots of Figs. 4.1 and 4.2, which show the reduced log-likelihood for 100 observations from the simulated AR(2) process

$$\begin{aligned} Z_t &= (1-m_1^{-1}B)(1-m_2^{-1}B)X_t \\ &= (1+0.9B)(1-1.1B)X_t, \end{aligned}$$

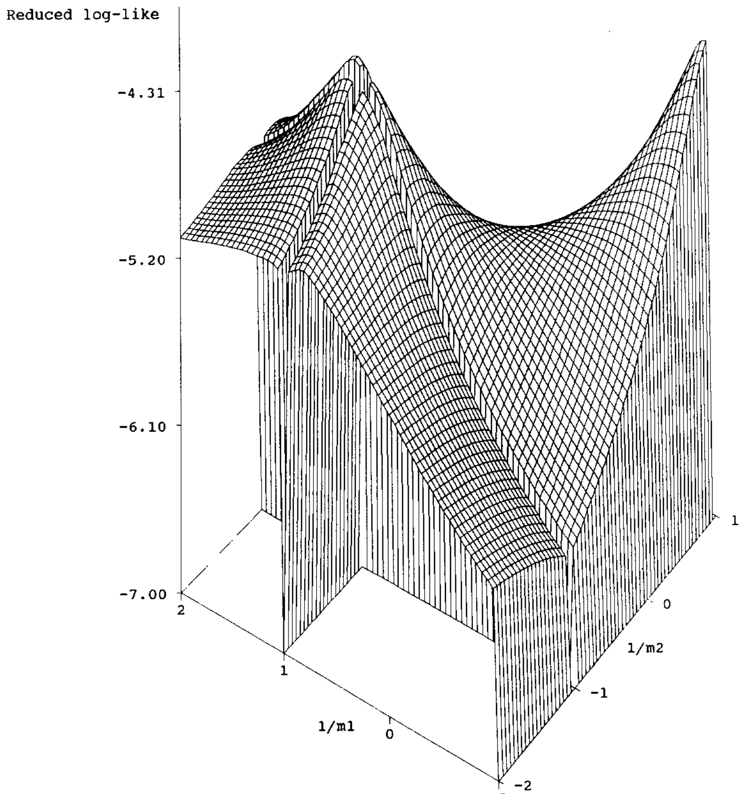

FIG. 4.1. Reduced log-likelihood surface plot.

where $\{Z_t\}$ is i.i.d. with the Laplace density $f(x) = \frac{1}{2}e^{-|x|}$. Here $s = 1$ and $\sigma = \sqrt{2}$. (Though the Laplace density does not, strictly speaking, meet assumption A2 of Section 2, we believe the results remain valid, as they do in the classical case for estimation of the location parameter in the Laplace density [7, p. 419].) These plots show sections of the causal ($s = 0$), mixed ($s = 1$), and purely noncausal ($s = 2$) regions. Note that the surface is unbounded along the lines $m_1^{-1} = \pm 1$ and $m_2^{-1} = \pm 1$, since these correspond to roots of $\phi(z)$ on the unit circle. Note also that because m_1^{-1} and m_2^{-1} commute in the reduced log-likelihood function, the surface is symmetric about the line $m_1^{-1} = m_2^{-1}$. In this example, the estimation procedure chose $\tilde{s} = 1$, $\tilde{m}_1^{-1} = -0.9270$, $\tilde{m}_2^{-1} = 1.1075$, and $\tilde{\sigma} = 1.1746$.

For the following simulation study, we applied the maximum likelihood estimation procedure to each of 1000 time series of length n and recorded the number of times R that the procedure chose the correct order of causality ($\tilde{s} = s$). To allow comparison with the asymptotic theory, we com-

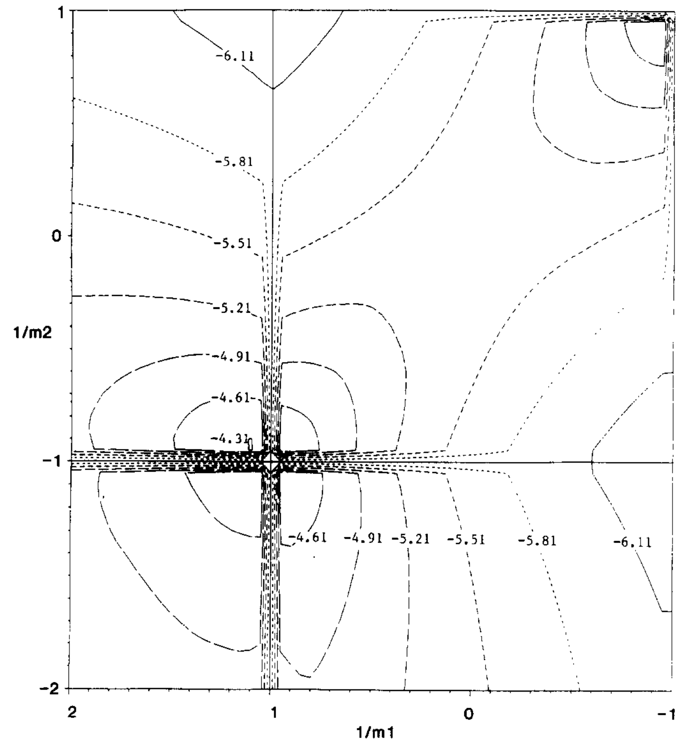

FIG. 4.2. Reduced log-likelihood contour plot.

puted sample means and standard deviations only for those R estimates. Results for the mixed case appear in Tables I–IV, while an example for the causal case and a few other numerical examples of interest appear below.

The tabled results for the mixed case show that the estimation procedure does an excellent job of choosing the order of causality of the model even when the roots m_1 and m_2 of $\phi(z)$ are near the unit circle (Tables I and II).

TABLE I

$\{Z_t\} \sim$ i.i.d. Laplace, $m_1^{-1} = -0.9$, $m_2^{-1} = 1.1$, $\sigma = \sqrt{2}$

		$\tilde{m}_1^{-1}$			$\tilde{m}_2^{-1}$			$\tilde{\sigma}$		
n	R	Mean	SD	ASD	Mean	SD	ASD	Mean	SD	ASD
50	613	−0.8671	0.0671	0.0436	1.1395	0.0830	0.0504	1.4309	0.2064	0.2103
100	805	−0.8935	0.0364	0.0309	1.1141	0.0505	0.0357	1.4096	0.1581	0.1487
200	935	−0.8887	0.0284	0.0218	1.1080	0.0285	0.0252	1.4174	0.1029	0.1051
400	1000	−0.8961	0.0176	0.0154	1.1026	0.0219	0.0178	1.4121	0.0674	0.0743

TABLE II

$\{Z_t\} \sim$ i.i.d. $t(4)$, $m_1^{-1} = -0.9$, $m_2^{-1} = 1.1$, $\sigma = \sqrt{2}$

		$\tilde{m}_1^{-1}$			$\tilde{m}_2^{-1}$			$\tilde{\sigma}$		
n	R	Mean	SD	ASD	Mean	SD	ASD	Mean	SD	ASD
50	684	−0.8752	0.0736	0.0517	1.1414	0.0845	0.0598	1.4730	0.2424	0.2029
100	791	−0.8810	0.0423	0.0366	1.1225	0.0550	0.0423	1.5077	0.1702	0.1435
200	925	−0.8913	0.0299	0.0259	1.1079	0.0351	0.0299	1.5073	0.1066	0.1014
400	980	−0.8973	0.0183	0.0183	1.1042	0.0221	0.0211	1.5089	0.0762	0.0717

When roots are near the unit circle—the boundary for causality—models with high likelihoods occur in the causal, mixed, and purely noncausal regions, as illustrated in Figs. 4.1 and 4.2, and the estimation procedure must choose among them. Not surprisingly, as the roots move away from the unit circle (Tables III and IV), the success rate for determining the order of causality is even higher.

For a causal example, we simulated $n = 200$ observations from the series

$$Z_t = (1 - m_1^{-1}B)(1 - m_2^{-1}B)X_t$$
$$= (1 - .8B)(1 + .8B)X_t,$$

where $\{Z_t\}$ was i.i.d. $t(4)$ ($\sigma = \sqrt{2}$). Sample means and standard deviations (SDs) computed for the R estimates which fell in the correct region, as well as asymptotic standard deviations (ASDs) computed as in Theorem 3 and (2.15), are recorded here:

		$\tilde{m}_1^{-1}$			$\tilde{m}_2^{-1}$			$\tilde{\sigma}$		
n	R	Mean	SD	ASD	Mean	SD	ASD	Mean	SD	ASD
200	882	0.7971	0.0421	0.0455	−0.7930	0.0404	0.0455	1.5014	0.1188	0.0939

TABLE III

$\{Z_t\} \sim$ i.i.d. Laplace, $m_1^{-1} = -0.5$, $m_2^{-1} = 1.3$, $\sigma = \sqrt{2}$

		$\tilde{m}_1^{-1}$			$\tilde{m}_2^{-1}$			$\tilde{\sigma}$		
n	R	Mean	SD	ASD	Mean	SD	ASD	Mean	SD	ASD
50	753	−0.4898	0.1102	0.0884	1.3642	0.2102	0.1102	1.4656	0.2710	0.2332
100	921	−0.4936	0.0694	0.0625	1.2972	0.0931	0.0779	1.3968	0.1558	0.1649
200	994	−0.5023	0.0472	0.0442	1.3189	0.0696	0.0551	1.4292	0.1283	0.1166
400	1000	−0.4965	0.0384	0.0312	1.3065	0.0489	0.0390	1.4161	0.0949	0.0824

TABLE IV

$\{Z_t\} \sim$ i.i.d. $t(4)$, $m_1^{-1} = -0.5$, $m_2^{-1} = 1.3$, $\sigma = \sqrt{2}$

		$\tilde{m}_1^{-1}$			$\tilde{m}_2^{-1}$			$\tilde{\sigma}$		
n	R	Mean	SD	ASD	Mean	SD	ASD	Mean	SD	ASD
50	687	−0.4895	0.1071	0.1067	1.3548	0.2212	0.1331	1.5326	0.3084	0.2371
100	855	−0.4926	0.0805	0.0755	1.3321	0.1163	0.0941	1.5098	0.1965	0.1677
200	957	−0.4907	0.0550	0.0534	1.3113	0.0736	0.0665	1.5091	0.1355	0.1186
400	1000	−0.5012	0.0395	0.0377	1.3022	0.0507	0.0471	1.4998	0.0904	0.0838

In addition to computing sample means and standard deviations, we computed sample correlation matrices and compared them to the asymptotic theory. The following example is for $n = 200$ observations from the mixed AR(2)

$$\begin{aligned} Z_t &= (1 - m_1^{-1}B)(1 - m_2^{-1}B)X_t \\ &= (1 + .5B)(1 - 1.3B)X_t, \end{aligned}$$

where $\{Z_t\}$ was i.i.d. $t(4)$ ($\sigma = \sqrt{2}$). In this case, the asymptotic theory indicates that estimates of m_2^{-1} and σ should be quite highly correlated. Estimates of m_1^{-1} should be weakly correlated with the estimates of the other parameters. Here are the sample and asymptotic standard deviation-correlation matrices:

	Sample				Asymptotic		
	$\tilde{m}_1^{-1}$	$\tilde{m}_2^{-1}$	$\tilde{\sigma}$		$\tilde{m}_1^{-1}$	$\tilde{m}_2^{-1}$	$\tilde{\sigma}$
$\tilde{m}_1^{-1}$	0.055	0.283	0.208	$\tilde{m}_1^{-1}$	0.053	0.280	0.171
$\tilde{m}_2^{-1}$		0.074	0.636	$\tilde{m}_2^{-1}$		0.067	0.611
$\tilde{\sigma}$			0.136	$\tilde{\sigma}$			0.119

More results from this simulation are included in Table IV.

We also loked at normal probability plots for the optimized parameters to check the asymptotic normality of the estimates. In the mixed case, the asymptotic normal approximation to the distribution of the estimates of m_1^{-1}, the inverse of the root of $\phi(z)$ outside the unit circle, is good even for n as small as 50. The distributions of the estimates of m_2^{-1} and σ are, however, quite skewed for the smaller samples ($n = 50$ and 100); this skewness diminishes as n increases.

An example with the parameter set $m_1^{-1} = -0.5$ and $m_2^{-1} = 1.1$ and with $\{Z_t\}$ i.i.d. $t(10)$ ($\sigma = \sqrt{5/2}$) illustrates the identifiability problem we encounter as the noise distribution approaches a Gaussian distribution. As the degrees of freedom v in the $t(v)$ distribution are increased, $t(v)$ approaches the standard Gaussian distribution. Since in the Gaussian case the parameters m_1^{-1}, m_2^{-1}, and σ are not identifiable [3, pp. 123–125], processes driven by near-Gaussian noise will be similarly troublesome. In this example, the estimation procedure is still largely successful in choosing the order of causality, with a 70.5% success rate. These results should be compared with the results for the $t(4)$ in Table IV.

		$\tilde{m}_1^{-1}$			$\tilde{m}_2^{-1}$			$\tilde{\sigma}$		
n	R	Mean	SD	ASD	Mean	SD	ASD	Mean	SD	ASD
400	705	−0.5010	0.0452	0.0455	1.3106	0.0624	0.0567	1.2858	0.0799	0.0664

For this final example, we used approximate log-likelihoods corresponding to AR(2) series with Laplace error distributions to estimate the parameters in models for simulated AR(2) series with $t(4)$ error distributions. (These simulated series were the same as those corresponding to $n = 200$ in Table II, so the results below are directly comparable.) This procedure is analogous to estimating parameters in general ARMA(p, q) models by maximizing the Gaussian likelihood even when the process is known to be non-Gaussian, a standard estimation procedure in time series. The resulting estimators are, for noncausal models, the analogue of least absolute deviation estimators. In this example, maximizing the Laplace likelihood yields remarkably good estimates. Note that the asymptotic standard deviation (ASD) recorded here is for an AR(2) driven by $t(4)$ noise.

		$\tilde{m}_1^{-1}$			$\tilde{m}_2^{-1}$		
n	R	Mean	SD	ASD	Mean	SD	ASD
200	837	−0.8915	0.0331	0.0259	1.1090	0.0388	0.0299

This example suggests that maximizing the Laplace likelihood can give reasonably efficient estimates of the parameters in the model. It appears that these estimates might in fact be $\sqrt{n}$-consistent, in which case they could be used as initial values in an adaptive estimation procedure like those advocated by Beran [1] and Kreiss [6].

In Tables I–IV the sample mean and standard deviation (SD) are calculated from the R estimates out of the 1000 replications for which the estimation procedure chose the correct order of causality ($\tilde{s} = s$). The asymptotic standard deviation (ASD) as computed from Theorem 3 and (2.15) is also recorded.

Note added in proof. It should be noted that this paper contains detailed derivations of the results of the paper "Nonminimum Phase non-Gaussian Autoregressive Processes" which appeared in the *Proc. Natl. Acad. Sci. USA* **87**, 179–181.

REFERENCES

[1] BERAN, R. (1976). Adaptive estimates for autoregressive processes. *Ann. Inst. Statist. Math.* **28** 77–89.

[2] BREIDT, F. J., AND DAVIS, R. A. (1990). Time-reversibility, identifiability, and independence of innovations for stationary time series. Preprint.

[3] BROCKWELL, P. J., AND DAVIS, R. A. (1987). *Time Series: Theory and Methods.* Springer-Verlag, New York.

[4] DENNIS, J. E., AND SCHNABEL, R. B. (1983). *Numerical Methods for Unconstrained Optimization and Nonlinear Equations.* Prentice–Hall, New York.

[5] KAHANER, D., MOLER, C., AND NASH, S. (1988). *Numerical Methods and Software.* Prentice–Hall, New York.

[6] KREISS, J. (1987). On adaptive estimation in stationary ARMA processes. *Ann. Statist.* **15** 112–133.

[7] LEHMANN, E. L. (1983). *Theory of Point Estimation.* Wiley, New York.

[8] LII, K. S., AND ROSENBLATT, M. (1982). Deconvolution and estimation of transfer function phase and coefficients for non-Gaussian linear processes. *Ann. Statist.* **10**, 1195–1208.

[9] NIKIAS, C. L., AND RAGHUVEER, M. R. (1987). Bispectrum estimation: A digital signal processing framework. *Proc. IEEE* **75** 869–891.

[10] ROSENBLATT, M. (1985). *Stationary Sequences and Random Fields.* Birkhäuser, Boston.