NONLINEAR REGRESSION WITH AUTOREGRESSIVE ERRORS

by

A. R. GALLANT and J. J. GOEBEL

Institute of Statistics
Mimeograph Series #986
Raleigh - March 1975

ABSTRACT

The article sets forth an estimator of the parameters of a nonlinear time series regression when the statistical behavior of the disturbances can be reasonably approximated by an autoregressive model. The sampling distribution of the estimator and relavent statistics is investigated both theoretically and using Monte-Carlo simulations.

NONLINEAR REGRESSION WITH AUTOREGRESSIVE ERRORS

A. R. Gallant and J. J. Goebel*

*A. R. Gallant is assistant professor of statistics and economics, Department of Statistics, North Carolina State University, Raleigh, N. C. J. J. Goebel is assistant professor of statistics, Department of Statistics, Iowa State University, Ames, Iowa 50010.

1. INTRODUCTION

This article examines estimation of the unknown parameter θ^* of the nonlinear time series regression model

$$y_t = f(x_t,\theta^*) + u_t \qquad (t = 1, 2, \ldots, n) .$$

The process $\{u_t\}_{t=+\infty}^{\infty}$ generating the realized disturbances $\{u_t\}_{t=1}^{n}$ is assumed to be a covariance stationary time series. This is to say that the covariances $\mathrm{Cov}(u_t,u_{t+h})$ of the time series depend only on the gap h and not on the position t in time. In consequence, the variance-covariance matrix Γ_n of the disturbance vector

$$u = (u_1, u_2, \ldots, u_n)' \qquad (n \times 1)$$

will have a banded structure with typical element $\gamma_{ij} = \gamma(i-j)$ where $\gamma(h)$ is the autocovariance function of the process

$$\gamma(h) = \mathrm{Cov}(u_t,u_{t+h}) \qquad (h = 0, \pm1, \pm2, \ldots) .$$

The appropriate estimater of θ^*, were Γ_n known, is the generalized nonlinear least squares estimator. Specifically, one would estimate θ^* by θ minimizing

$$[y - f(\theta)]'\Gamma_n^{-1}[y - f(\theta)]$$

where

$$y = (y_1, y_2, \ldots, y_n)' \qquad (n \times 1)$$

and

$$f(\theta) = [f(x_1,\theta), \ldots, f(x_n,\theta)]' \qquad (n \times 1) .$$

When Γ_n is not known, as we assume here, the obvious approach is to substitute an estimator of Γ_n in the formula above.

Recently, Hannan [6] and Goebel [5] have obtained estimators for θ^* employing this approach. They use a circular symmetric matrix [4, 1971, Ch. 4] to approximate Γ_n. The approximating matrix has easily computed eigen vectors which do not depend on any unknown parameters. This fact considerably reduces computational difficulties and avoids inversion of a matrix of the same order as the sample size; see, e.g., [3]. The eigen values of the approximating matrix are proportional to the spectral density of the time series $\{u_t\}_{t=-\infty}^{\infty}$ at appropriate frequencies. Hannan and Goebel estimate the spectral density from ordinary nonlinear least squares residuals to obtain consistent estimates of the eigen values of the approximating matrix. Their estimators differ in choice of procedures for estimating the spectral density but share the same desirable asymptotic properties under fairly unrestrictive assumptions on the time series generating the disturbances. Moreover, the Monte-Carlo simulations reported by Goebel indicate that this type of estimater has better efficiency in small samples than the ordinary nonlinear least squares estimator.

Presently, we shall demonstrate by an example that the approximating matrix used by Hannan and Goebel can be very inaccurate in small samples. This is not due to sampling variation but due to the form of a circular symmetric matrix which, in effect, imposes the unrealistic condition $\gamma(h) = \gamma(n-h)$ $(h = 1, 2, \ldots, n)$ on the approximating autocovariances. Often, the serially correlated disturbances appear to satisfy more restrictive assumptions than Hannan and Goebel were willing to assume. In these cases, it would seem possible to exploit these assumptions and obtain a better approximating matrix. One would naturally expect that the use of a better approximating matrix in the

generalized nonlinear least squares formula would lead to better small sample efficiency. In view of Hannan and Goebels theoretical results, we would not expect to obtain better large sample efficiency.

An assumption which is frequently satisfied in applications - at least to within errors which can reasonably be ascribed to sampling variation - is that the disturbances can be reduced to a white noise process by using a short linear filter. To be more specific, the time series $\{u_t\}_{t=-\infty}^{\infty}$ is assumed to satisfy the equations

$$u_t + a_1 u_{t-1} + a_2 u_{t-2} + \ldots + a_q u_{t-q} = e_t \qquad (t = 0, \pm 1, \pm 2, \ldots)$$

where $\{e_t\}_{t=-\infty}^{\infty}$ is a sequence of independently and identically distributed random variables each with mean zero and variance σ^2. In addition, we assume the roots of the polynomial

$$m^q + a_1 m^{q-1} + a_2 m^{q-2} + \ldots + a_q$$

have roots less than one in absolute value. A time series $\{u_t\}_{t=-\infty}^{\infty}$ which satisfies these assumptions is called an autoregressive process of order q.

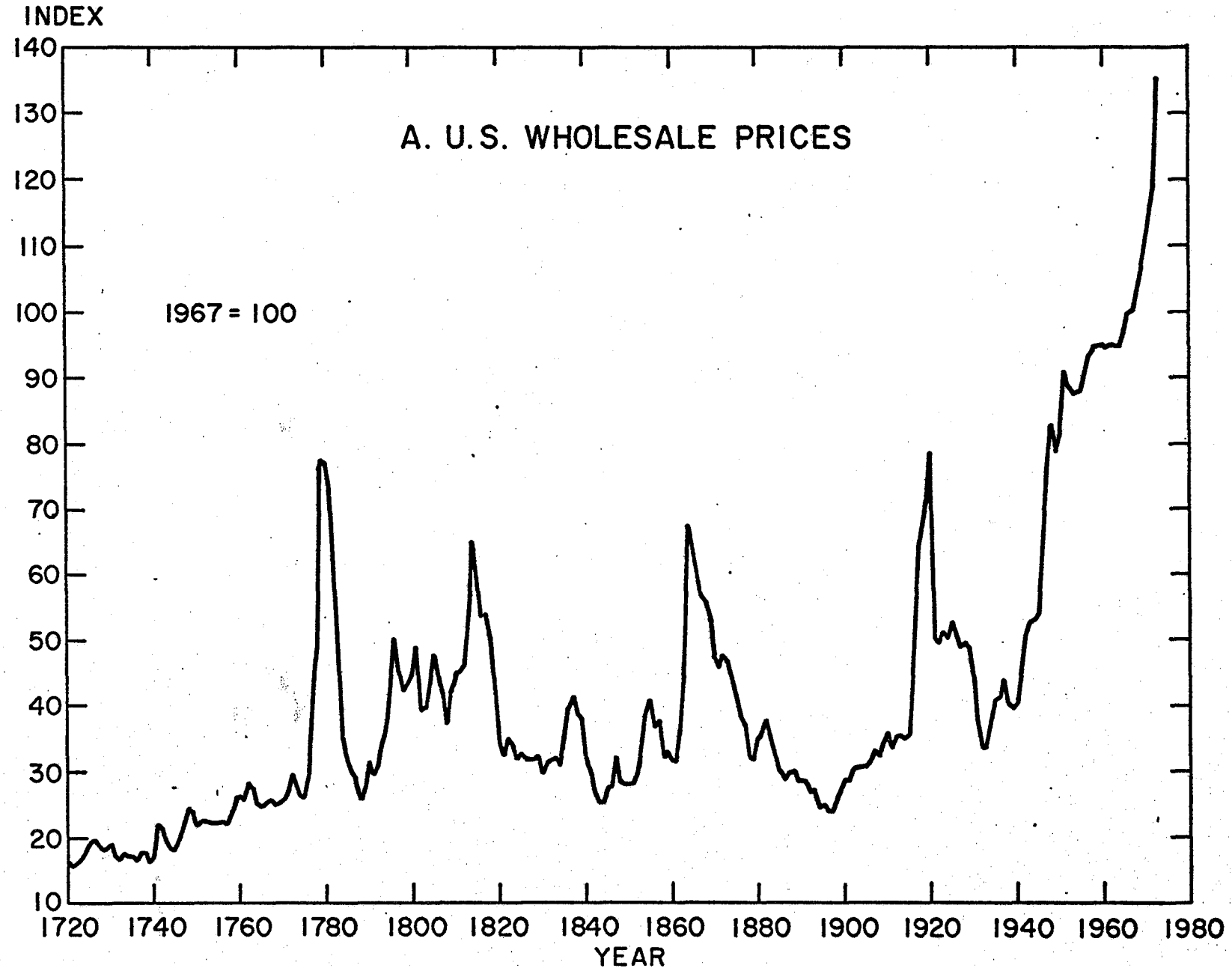
The Wholesale Price Index for the years 1720 through 1973 plotted as Figure A appears to satisfy these assumptions. To these data we have fitted an exponential growth model

$$y_t = \theta_1 e^{\theta_2 \cdot t} + u_t \qquad (t = 1, \ldots, n = 254)$$

by ordinary nonlinear least squares to obtain residuals $\{\hat{u}_t\}_{t=1}^{254}$. From these residuals we have estimated the autocovariances using

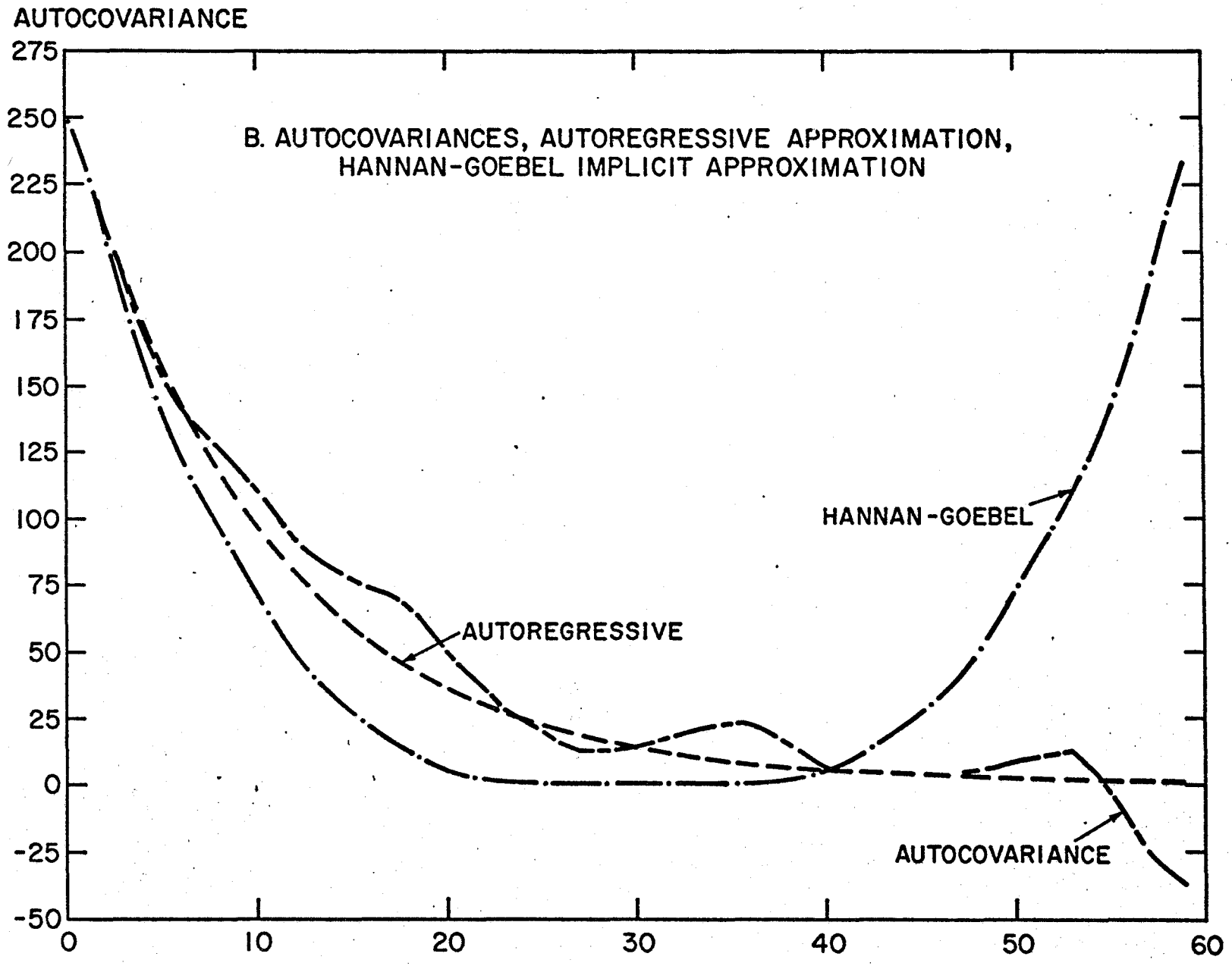
$$\hat{\gamma}(h) = (1/n) \sum_{t=1}^{n-h} \hat{u}_t \hat{u}_{t+h} \qquad (h = 0, 1, \ldots, 60)$$

and plotted them as Plot AUTOCOVARIANCE in Figure B.


INDEX
140
130
120
110
100
90
80
70
60
50
40
30
20
10
A. U.S. WHOLESALE PRICES
1967 = 100
1720
1740
1760
1780
1800
1820
1840
1860
1880
1900
1920
1940
1960
1980
YEAR

FOOTNOTE TO FIGURE A

Source: Composite derived from: Wholesale Prices for Philadelphia, 1720 to 1861, Series E82, [14]; Wholesale Prices, All Commodaties, 1749 to 1890, Series E1, [14]; Wholesale Prices, All Commodaties, 1890 to 1951, Series E13, [14]; Wholesale Prices, All Commodoties, 1929 - 1971, [12]; Wholesale Prices, All Commodoties, 1929 - 1973, [13].


AUTOCOVARIANCE
275
250
225
200
175
150
125
100
75
50
25
0
-25
-50
0
10
20
30
40
50
60
B. AUTOCOVARIANCES, AUTOREGRESSIVE APPROXIMATION,
HANNAN-GOEBEL IMPLICIT APPROXIMATION
HANNAN-GOEBEL
AUTOREGRESSIVE
AUTOCOVARIANCE

Using the methods set forth in Section 3 and assuming $q = 2$ we estimate $(a_1, a_2, \sigma^2) = (-1.048, 0.1287, 34.09)$. Using the Yule-Walker equations and these estimates, we obtain the approximating autocovariances shown as Plot AUTOREGRESSIVE in Figure B.

As we mentioned earlier, the entries on the diagonal bands of the variance-covariance matrix Γ_n are the autocovariances $\gamma(h)$ of the process; that is, $\gamma(h) = \gamma_{ij}$ where $h = |i-j|$. Thus, the approximating circular symmetric matrix used by Hannan and Goebel implicity defines an estimator of the autocovariance function; see Box and Jenkins [2, p. 31] for a schematic illustration. Due to the nature of a circular symmetric matrix, the autocovariance function implicitly defined will vary with the sample size n and with the choice of an estimator for the spectral density. We used $n = 60$ and estimated the spectral density using a Tukey-lag window [9, p. 243 ff.] with $M = 25$ to obtain the approximating autocovariances shown as Plot HANNAN-GOEBEL in Figure B.

As can be seen from Figure B, the approximation of the autocovariance function obtained using the autoregressive assumption is considerably more realistic than the approximation implied by the Hannan-Goebel procedure. Correspondingly, for these data we would expect that a generalized nonlinear least squares estimator using the autoregressive assumption to obtain the approximating variance-covariance matrix would perform better in small samples than a Hannan- Goebel estimator. In general, when the ordinary nonlinear squares residuals $\{\hat{u}_t\}_{t=1}^{n}$ appear to satisfy the assumption that they were generated by a low order autoregressive process we would expect that a generalized nonlinear least squares estimator which exploits this assumption would have better efficiency in small samples than a Hannan-Goebel estimator.

This article investigates this hypothesis. In Section 2 we describe an estimation procedure. In Section 3 we summarize the consequences of the autoregressive assumption which motivate the procedure. In Section 4 we set forth the asymptotic properties of the estimator. In Section 5 we investigate the small sample properties of the estimator using a Monte-Carlo simulation and compare them with Goebel's results.

We conclude, on the basis of the Monte-Carlo evidence, that the estimator suggested here will have better efficiency in small samples than the Hannan-Goebel estimator when the autoregressive assumption is at least approximately satisfied. Gains in efficiency relative to ordinary least squares can be dramatic when the autoregressive assumption is exactly satisfied and q is correctly identified.

2. ESTIMATION PROCEDURE

Our objective is to estimate the unknown parameter θ^* appearing in the nonlinear regression equations

$$y_t = f(x_t,\theta^*) + u_t \qquad (t = 1, 2, \ldots, n)$$

when we assume that the disturbance terms are an autoregressive process of order q. A detailed description of such a process is deferred to the next section where methods of estimating q from the data are discussed. The inputs x_t are known k by 1 vectors and the unknown parameter θ^* is a p by 1 vector known to be contained in the parameter space Θ. The regression equations may be written in vector form as

$$y = f(\theta) + u$$

using the notation of the previous section.

The first step of the procedure is to compute the ordinary least squares estimator $\hat{\theta}$ which minimizes

$$[y - f(\theta)]'[y - f(\theta)]$$

over Θ using, for example, Hartley's [7] modified Gauss-Newton method or Marquardt's [11] algorithm. A program implementing at least one of these algorithms is usually found at computing centers with a statistical program library.

The second step is to compute the ordinary least squares residuals

$$\hat{u} = y - f(\hat{\theta})$$

and from these estimate the autocovariances up to lag q of the disturbances using

$$\hat{\gamma}(h) = (1/n) \sum_{t=1}^{n-|h|} \hat{u}_t \hat{u}_{t+|h|} \qquad (h = 0, 1, \ldots, q) .$$

For the third step, let

$$\hat{\Gamma}_q = \begin{bmatrix} \hat{\gamma}(0) & \hat{\gamma}(1) & \cdots & \hat{\gamma}(q-1) \\ \hat{\gamma}(1) & \hat{\gamma}(0) & \cdots & \hat{\gamma}(q-2) \\ \vdots & \vdots & & \\ \hat{\gamma}(q-1) & \hat{\gamma}(q-2) & \cdots & \hat{\gamma}(0) \end{bmatrix} \qquad (q \times q)$$

$$\hat{\gamma}_q = [\hat{\gamma}(1), \hat{\gamma}(2), \ldots, \hat{\gamma}(q)]' \qquad (q \times 1)$$

and compute

$$\hat{a} = -\hat{\Gamma}_q \hat{\gamma}_q \qquad (q \times 1) ,$$

$$\hat{\sigma}^2 = \hat{\gamma}(0) + \hat{a}'\hat{\gamma}_q .$$

Factor $\hat{\Gamma}_q^{-1} = \hat{P}_q'\hat{P}_q$ using, e.g., Cholesky's method [8, p. 158] and set

$$\hat{P} = \begin{bmatrix} \sqrt{\hat{\sigma}^2}\,\hat{P}_q & \vdots & 0 \\ \hat{a}_q \quad \hat{a}_{q-1} \quad \cdots \quad \hat{a}_1 & & 1 \\ & \hat{a}_q \quad \hat{a}_{q-1} \quad \cdots \quad \hat{a}_1 \quad 1 & \\ & \ddots \quad \ddots \quad \ddots & \\ & \hat{a}_q \quad \hat{a}_{q-1} \quad \cdots \quad \hat{a}_1 \quad 1 & \end{bmatrix} \begin{matrix} \} q \text{ rows} \\ \\ \\ \} n - q \text{ rows} \end{matrix}$$

Note that to perform the multiplication $\hat{P}w$ on a digital computer it is not necessary to store $\hat{P}$; only $\hat{P}_q$ and $\hat{a}_1, \hat{a}_2, \ldots, \hat{a}_q$ are needed.

Define:

$$\nabla' f(x,\theta) = [(\partial/\partial\theta_1)f(x,\theta), \ldots, (\partial/\partial\theta_p)f(x,\theta)] \qquad (1 \times p) ,$$

$F(\theta)$ = the n by p matrix whose t^{th} row is $\nabla' f(x_t,\theta)$.

The fourth step is to compute $\tilde{\theta}$ minimizing

$$Q_n(\theta) = (1/n)[\hat{P}y - \hat{P}f(\theta)]'[\hat{P}y - \hat{P}f(\theta)]$$

and from this value obtain

$$\tilde{\sigma}^2 = [\hat{P}y - \hat{P}f(\tilde{\theta})]'[\hat{P}y - \hat{P}f(\tilde{\theta})]/(n-p) ,$$

and

$$\tilde{C} = [(1/n)F'(\tilde{\theta})\hat{P}'\hat{P}F(\tilde{\theta})]^{-1} .$$

As shown in Section 4, $\sqrt{n}\,(\tilde{\theta} - \theta^*)$ is asymptotically normally distributed with a variance-covariance matrix for which $\tilde{\sigma}^2\,\tilde{C}$ is a strongly consistent estimator.

Either Hartley's [7] or Marquardt's [11] algorithm may be used for the final minimization to obtain $\tilde{\theta}$. Put $z = \hat{P}y$ and $g(\theta) = \hat{P}f(\theta)$. The problem

is then to minimize $\Sigma_{t=1}^{n} [z_t - g_t(\theta)]^2$. The derivatives $\nabla' g_t(\theta)$ needed by the program may be obtained as the $t^{\underline{th}}$ row of $\hat{P}F(\theta)$. As mentioned above, this matrix multiplication does not require the storage by the n by n matrix $\hat{P}$. The matrix printed by the program

$$(A'A)^{-1} = [\Sigma_{t=1}^{n} \nabla g_t(\tilde{\theta}) \nabla' g_t(\tilde{\theta})]^{-1}$$

will satisfy $\tilde{C} = n(A'A)^{-1}$.

The estimation procedure may be iterated by returning to the second step with $\tilde{\theta}$ replacing $\hat{\theta}$. The asymptotic properties of this two-stage estimator do not differ from those of the one-stage estimator $\tilde{\theta}$. Intuitively, it would seem that the two-stage estimator would be more efficient but the Monte-Carlo evidence, Section 5, suggests otherwise. The only case we find improved performance is when the autoregressive assumption is exactly satisfied.

3. AUTOREGRESSIVE PROCESSES

The autoregressive process $\{u_t\}_{t=-\infty}^{\infty}$ of order q is defined implicitly as a sequence of random variables which satisfies the difference equation

$$u_t + a_1 u_{t-1} + a_2 u_{t-2} + \ldots + a_q u_{t-q} = e_t$$

where the process $\{e_t\}_{t=-\infty}^{\infty}$ is a sequence of independently and identically distributed random variables each with mean zero, finite variance σ^2 , and finite $4^{\underline{th}}$ moment. The parameters a_i $(i = 1, \ldots, q)$ of the defining equation are assumed to satisfy the condition that the roots of the polynomial

$$m^q + a_1 m^{q-1} + a_2 m^{q-2} + \ldots + a_q$$

are less than one in absolute value.

The process $\{u_t\}$ can be given an explicit moving average representation in terms of weights $\{w_j\}_{j=0}^{\infty}$ defined by the equations

$$w_0 = 1$$

$$w_1 = -a_1 w_0$$

$$w_2 = -a_1 w_1 - a_2 w_0$$

$$\vdots$$

$$w_{q-1} = -a_1 w_{q-2} - a_2 w_{q-3} - \cdots - a_{q-1} w_0$$

and

$$w_j = -a_1 w_{j-1} - a_2 w_{j-2} - \cdots - a_q w_{j-q}$$

for $j = q, q+1, \ldots$. These weights are absolutely summable $(\Sigma_{j=0}^{\infty} |w_j| < 0)$ and

$$u_t = \Sigma_{j=0}^{\infty} w_j e_{t-j}$$

almost surely [4].

The autocovariances of the process $\{u_t\}$ are defined by

$$\gamma(h) = \mathrm{Cov}(u_t, u_{t+h}) \qquad (h = 0, \pm 1, \pm 2, \ldots) .$$

From the moving average representation we have that the autocovariances are given by

$$\gamma(h) = \sigma^2 \Sigma_{j=0}^{\infty} w_j w_{j+|h|} \qquad (h = 0, \pm 1, \pm 2, \ldots)$$

and satisfy the Yule-Walker equations

$$\gamma(j) + a_1\gamma(j-1) + \ldots + a_q\gamma(j-q) = \begin{cases} w_j\sigma^2 & j \leq 0 \\ 0 & j > 0 \end{cases}$$

In addition, the autocovariances are absolutely summable $(\Sigma_{h=-\infty}^{\infty}|\gamma(h)|<\infty)$.

The Yule-Walker equation may be used to compute the parameters $a_1, a_2, \ldots, a_q, \sigma^2$ defining the process $\{u_t\}$ from knowledge of the first q autocovariances $\gamma(h)$ $(h = 0, 1, \ldots, q)$. Define

$$\Gamma_s = \begin{bmatrix} \gamma(0) & \gamma(1) & \ldots & \gamma(s-1) \\ \gamma(1) & \gamma(0) & \ldots & \gamma(s-2) \\ \vdots & \vdots & & \\ \gamma(s-1) & \gamma(s-2) & \ldots & \gamma(0) \end{bmatrix} \qquad (s \times s),$$

$$\gamma_s' = [\gamma(1), \gamma(2), \ldots, \gamma(s)] \qquad (s \times 1),$$

and

$$a' = (a_1, a_2, \ldots, a_q) \qquad (q \times 1).$$

The parameters are computed from the equations

$$a = -\Gamma_q^{-1}\gamma_q$$

$$\sigma^2 = \gamma(0) + a_q'\gamma_q .$$

The observed portion $\{u_t\}_{t=1}^{n}$ of a realization of the process $\{u_t\}_{t=-\infty}^{\infty}$ has variance-covariance matrix Γ_n. For n larger than q a matrix P which diagonalizes Γ_n may be obtained as follows. Factor Γ_q^{-1} as $\Gamma_q^{-1} = P_q'P_q$ and set

$$P = \begin{bmatrix} \sqrt{\sigma^2}\, P_q & \vdots & 0 \\ \hline a_q \quad a_{q-1} \quad \cdots \quad a_1 & 1 & \\ \quad a_q \quad a_{q-1} \quad \cdots \quad a_1 & 1 & \\ \quad \ddots \quad \ddots \quad \ddots & \ddots & \ddots \\ \quad a_q \quad a_{q-1} \quad \cdots \quad a_1 & 1 & \end{bmatrix} \begin{matrix} \} \text{ q rows} \\ \\ \} \text{ n-q rows} \end{matrix}$$

Using the Yule-Walker equations and the equations defining the weights $\{w_j\}$, one can verify that $P\Gamma_n P' = I_n$. This fact motivates the use of $\hat{P}$ in the estimation procedure described in the previous section.

Estimators of the autocovariances may be obtained from $\{u_t\}_{t=1}^n$ as

$$\bar{\gamma}(h) = (1/n) \Sigma_{t=1}^{n-|h|} u_t u_{t-|h|} \qquad (h = 0, \pm 1, \ldots, \pm (n-q)) .$$

In the nonlinear regression context we cannot observe $\{u_t\}_{t=1}^n$ directly and must substitute the estimators $\hat{\gamma}(h)$ computed from the ordinary nonlinear least squares residuals $\{\hat{u}_t\}$. We will show in the next section that the estimators $\hat{\gamma}(h)$ and $\bar{\gamma}(h)$ are asymptotically equivalent under appropriate regularity conditions.

We have [4] that

$$\lim_{n\to\infty} n \operatorname{Cov}[\bar{\gamma}(h), \bar{\gamma}(\ell)]$$

$$= [\mathcal{E}(e_t^4)/\sigma^4 - 3]\gamma(h)\gamma(\ell)$$

$$+ \Sigma_{j=-\infty}^{\infty} [\gamma(j)\gamma(j-h+\ell) + \gamma(j+\ell)\gamma(j-h)]$$

and [5] that $\bar{\gamma}(h)$ converges almost surely to $\gamma(h)$. Using Chebischev's inequality the covariance formula implies that $\sqrt{n}\,[\bar{\gamma}(k) - \gamma(h)]$ is bounded in probability; that is, given $\delta > 0$ there is a finite bound B and an N such

that for all $n > N$

$$P[\sqrt{n}|\bar{\gamma}(h) - \gamma(h)| > B] < \delta \ .$$

If we estimate the parameters $a_1, \ldots, a_q, \sigma^2$ by substituting $\bar{\gamma}(h)$ $(h = 0, 1, \ldots, q)$ in the equations

$$a = -\Gamma_q^{-1}\gamma_q$$

$$\sigma^2 = \gamma(0) + a'\gamma_q$$

to obtain

$$\bar{a} = -\bar{\Gamma}_q^{-1}\bar{\gamma}_q$$

$$\bar{\sigma}^2 = \bar{\gamma}(0) + \bar{a}'\bar{\gamma}_q$$

we have as consequences of the facts set forth in the previous paragraph that each $\bar{a}_i$ converges almost surely to a_i and that $\sqrt{n}\,(\bar{a}_i - a_i)$ is bounded in probability. The same is true for $\bar{\sigma}^2$. These equations may be solved subject to <u>a priori</u> restrictions that some of the $a_i = 0$. This will not disturb the properties of the estimators $\bar{a}_i$ provided that the restrictions are, in fact, true.

To determine the order q of the process from the data, we consider a test of

$$H: \ a_q = 0 \quad \text{against} \quad A: \ a_q \neq 0$$

at the α-level of significance. We may use a two-sided "t-test" by comparing

$$\bar{t} = \frac{\sqrt{n}|\bar{a}_q|}{s\sqrt{\bar{\gamma}^{qq}}}$$

where $\bar{s}^2 = [\bar{\gamma}(0) + \bar{a}'\bar{\gamma}_q]/(n-q)$ and $\bar{\gamma}^{qq}$ is the $q^{\underline{th}}$ diagonal element of $\bar{\Gamma}_q^{-1}$ to the two-sided α-level critical point c of a t-variate with n-q degrees of freedom. We have [1, Sec. 5.6.3] that

$$\lim_{n\to\infty} P[\bar{t} > c|H] = \alpha$$

and

$$\lim_{n\to\infty} P[\bar{t} > c|A] = 1 .$$

This test is employed to determine the order q of an autoregressive process using procedures analogous to those used to determine the appropriate degree of a polynomial used in polynomial regression analysis. One may, for example, test sequentially upward or, alternatively, start from a very high order and use Anderson's [1, Sec. 3.2.2] downward selection procedure.

For a sequence of weights $\{c_t(\theta)\}_{t=1}^{\infty}$ depending on a p-dimensional parameter $\theta \in \Theta$ there is a Uniform Strong Law [5, 6] which, in the present context, states that

$$S_n(\theta) = (1/n) \Sigma_{t=1}^{n} c_t(\theta)u_t$$

converges almost surely to zero uniformly for θ in Θ provided that:

(i) The process $\{u_t\}$ is as described above.

(ii) Each $c_t(\theta)$ is a continuous function for θ in Θ.

(iii) The set Θ is compact.

(iv) The sum $(1/n) \Sigma_{t=1}^{n} c_t(\alpha)c_t(\beta)$ converges uniformly for all α and β in Θ as n tends to infinity.

4. ASYMPTOTIC PROPERTIES OF THE ESTIMATOR

The assumptions used by Goebel [5] to show that $\hat{\theta}$ converges almost surely to θ^* and that $\sqrt{n}\,(\hat{\theta} - \theta^*)$ is bounded in probability may be stated in the present context as:

Assumptions: The errors $\{u_t\}$ are an autoregressive process of order q as described in the previous section. The true parameter value θ^* is contained in an open sphere S, which is, in turn, contained in Θ; the parameter space Θ is a compact set. The response function $f(x,\theta)$ and its first and second partial derivatives in θ are continuous over the parameter space. The sequence of inputs $\{x_t\}$ are chosen so that the functions

$$\bar{g}(h,\alpha,\beta) = (1/n)\,\Sigma_{t=1}^{n-|h|}\, g_1(x_t,\alpha)g_2(x_{t+|h|},\beta) \qquad (h = 0, \pm 1, \ldots, \pm q)$$

converge uniformly for all α, β in Θ as n tends to infinity where $g_1(x,\theta)$ and $g_2(x,\theta)$ may variously be $f(x,\theta)$, $(\partial/\partial\theta_i)f(x,\theta)$, or $(\partial^2/\partial\theta_i\partial\theta_j)f(x,\theta)$. The function $\lim_{n\to\infty}(1/n)\,\Sigma_{t=1}^{n}\,\delta^2(x_t,\theta)$ where

$$\delta(x,\theta) = f(x,\theta^*) - f(x,\theta)$$

is assumed to be non-zero on Θ except at the point θ^*. The $p \times p$ matrix $V(h)$ with typical element

$$\bar{v}_{ij}(h) = \lim_{n\to\infty}(1/n)\Sigma_{t=1}^{n-|h|}(\partial/\partial\theta_i)f(x_t,\theta^*)(\partial/\partial\theta_j)f(x_{t+|h|},\theta^*)$$

$$(h = 0, \pm 1, \ldots, \pm q)$$

is assumed to be non-singular when $h = 0$.

These assumptions are sufficient to prove that the autocovariance estimators $\hat{\gamma}(h)$ computed from ordinary nonlinear least squares residuals are asymptotically equivalent to the estimators $\bar{\gamma}(h)$ computed from the unobservable disturbance terms.

Theorem 1. Under the assumptions above,

$$\hat{\gamma}(h) = \bar{\gamma}(h) + \alpha_n \qquad (h = 0, \pm 1, \ldots, \pm q)$$

where α_n converges almost surely to zero and $\sqrt{n}\,\alpha_n$ converges in probability to zero.

Proof. Let $\dot{\theta} = \hat{\theta}$ when $\hat{\theta}$ is in S and let $\dot{\theta} = \theta^*$ when $\hat{\theta}$ is not in S. Since, by Goebel's results, $\hat{\theta} = \dot{\theta}$ almost surely for n sufficiently large it will suffice to prove the theorem with $\dot{\theta}$ replacing $\hat{\theta}$. This allows the use of Taylor's theorem in the proof.

The estimator with $\dot{\theta}$ replacing $\hat{\theta}$ may be written as

$$\hat{\gamma}(h) = (1/n)\,\Sigma_{t=1}^{n-|h|}[u_t + \delta(x_t,\dot{\theta})][u_{t+|h|} + \delta(x_{t+|h|},\dot{\theta})]$$

$$= \bar{\gamma}(h) + (1/n)\,\Sigma_{t=1}^{n-|h|}[u_t\delta(x_{t+|h|},\dot{\theta}) + u_{t+|h|}\delta(x_t,\dot{\theta})$$

$$+ \delta(x_t,\dot{\theta})\delta(x_{t+|h|},\dot{\theta})] = \bar{\gamma}(h) + \alpha_n .$$

Using Taylor's theorem we may write

$$\alpha_n = \{(1/n)\Sigma_{t=1}^{n}[u_t\nabla' f(x_{t+|h|},\bar{\theta}) + u_{t+|h|}\nabla' f(x_t,\bar{\theta})$$

$$+ \delta(x_t,\hat{\theta})\nabla' f(x_{t+|h|},\bar{\theta})]\} \quad (\dot{\theta} - \theta^*)$$

where the $\bar{\theta}$ indicate points on the line segment joining $\dot{\theta}$ to θ^*. The term in braces converges almost surely to zero by the Uniform Strong Law, the assumption that $(1/n)\Sigma_{t=1}^{n}\,\delta(x_t,\alpha)\nabla' f(x_{t+|h|},\beta)$ converges uniformly, and the almost sure convergence of $\dot{\theta}$ to θ^*. Thus, α_n converges almost surely to zero by th almost sure convergence of $\dot{\theta}$; $\sqrt{n}\,\alpha_n$ converges in probability to zero because $\sqrt{n}\,(\dot{\theta} - \theta^*)$ is bounded in probability. ∎

The following additional assumptions are needed to derive the asymptotic properties of $\tilde{\theta}$.

Assumptions. (continued) The $q \times q$ matrix with typical element $\bar{\delta}(i-j,\theta)$ where $\bar{\delta}(h,\theta) = \lim_{n\to\infty}(1/n)\Sigma_{t=1}^{n-|h|}\delta(x_t,\theta)\delta(x_{t+|h|},\theta)$ is assumed to be non-singular on Θ except at the point θ^* . The matrix

$$C = \Sigma_{i=0}^{q}\Sigma_{j=0}^{q} a_i a_j \frac{1}{2}[V(i-j) + V'(i-j)]$$

is non-singular where the a_i are the coefficients of the stochastic difference equation defining $\{u_t\}$ and $a_0 = 1$.

Theorem 2. Under the assumptions above, $\tilde{\theta}$ converges almost surely to θ^* and $\tilde{\sigma}^2$ converges almost surely to σ^2 .

Proof. Let $\hat{\gamma}_q^{ij}$ be a typical element of $\hat{\Gamma}_q^{-1}$. Then

$$Q_n(\theta) = (\hat{\sigma}^2/n)\Sigma_{i=1}^{q}\Sigma_{j=1}^{q}[u_i + \delta(x_i,\theta)]\hat{\gamma}_q^{ij}[u_j + \delta(x_j,\theta)]$$

$$+ (1/n)\Sigma_{t=q+1}^{n}\{\Sigma_{j=0}^{q}\hat{a}_j[u_{t-j} + \delta(x_{t-j},\theta)]\}^2$$

$$= a_n(\theta)/n + b_n(\theta)/n .$$

Now $a_n(\theta)/n$ converges almost surely to zero uniformly in θ because $\hat{\gamma}_q^{ij}$ converges almost surely to γ_q^{ij} and the continuous functions $\delta(x_i,\theta)$ are bounded over the compact set Θ . Let $\bar{\gamma}^*(i-j) = (1/n)\Sigma_{t=q+1}^{n}u_{t-j}u_{t-i}$; this estimator differs from $\bar{\gamma}(i-j)$ by only a finite number of terms, so they share asymptotic properties. The second term may be written

$$b_n(\theta)/n = \Sigma_{i=0}^{q}\Sigma_{j=0}^{q}\hat{a}_i\hat{a}_j[\bar{\gamma}^*(i-j) + (1/n)\Sigma_{t=q+1}^{n}\delta(x_{t-i},\theta)\delta(x_{t-j},\theta)]$$

$$+ \Sigma_{i=0}^{q}\Sigma_{j=0}^{q}\hat{a}_i\hat{a}_j(1/n)\Sigma_{t=q+1}^{n}[u_{t-i}\delta(x_{t-j},\theta) + u_{t-j}\delta(x_{t-i},\theta)]$$

which converges almost surely to

$$\Sigma_{i=0}^{q}\Sigma_{j=0}^{q} a_i a_j \gamma(i-j) + \Sigma_{i=0}^{q}\Sigma_{j=0}^{q} a_i a_j \bar{\delta}(i-j,\theta)$$

uniformly in θ by the Uniform Strong Law, the definition of $\bar{\delta}(h,\theta)$, and the almost sure convergence of $\hat{a}_i$ and $\bar{\gamma}^*(h)$. Using the Yule-Walker equations we have

$$\Sigma_{i=0}^{q} a_i \Sigma_{j=0}^{q} a_j \gamma(i-j) = a_0 w_0 \sigma^2 = \sigma^2$$

whence $Q_n(\theta)$ converges to

$$\bar{Q}(\theta) = \sigma^2 + \Sigma_{i=0}^{q} \Sigma_{j=0}^{q} a_i a_j \bar{\delta}(i-j,\theta)$$

almost surely uniformly in θ. The term on the right is the limit of a sum of squares and is therefore positive or zero. By our non-singularity assumption, $\bar{Q}(\theta) = \sigma^2$ only when $\theta = \theta^*$.

Consider a sequence of points $\{\tilde{\theta}_n\}_{u=1}^{\infty}$ minimizing $Q_n(\theta)$ over Θ corresponding to a realization of the process $\{u_t\}_{t=-\infty}^{\infty}$. Since Θ is compact there is at least one limit point θ° and one subsequence $\{\tilde{\theta}_{n_m}\}_{m=1}^{\infty}$ such that $\lim_{m\to\infty} \tilde{\theta}_{n_m} = \theta^\circ$. Unless, this realization belongs to the exceptional set E, $Q_{n_m}(\theta)$ converges uniformly to $\bar{Q}(\theta)$ whence

$$\sigma^2 \le \bar{Q}(\theta^\circ) = \lim_{m\to\infty} Q_{n_m}(\tilde{\theta}_{n_m}) \le \lim_{m\to\infty} Q_{n_m}(\theta^*) = \sigma^2$$

which implies $\theta^\circ = \theta^*$. Thus, the sequence $\{\tilde{\theta}_n\}$ has one limit point θ^* except for realizations of the process $\{u_t\}$ in E, where $P(E) = 0$.

Since $\tilde{\sigma}^2 = [n/(n-p)]Q_n(\tilde{\theta})$ we have the almost sure convergence of $\tilde{\sigma}^2$ to σ^2. □

Theorem 3. Under the assumptions above, $\sqrt{n}\,(\tilde{\theta} - \theta^*)$ converges in distribution to a p-variate normal with mean vector zero and variance-covariance

matrix $\sigma^2 C^{-1}$. The elements of the matrix

$$\tilde{C} = (1/n)F'(\tilde{\theta})\hat{P}'\hat{P}F(\tilde{\theta})$$

converge almost surely to the corresponding elements of C.

Proof. Let $\dot{\theta} = \tilde{\theta}$ if $\tilde{\theta}$ is in S and let $\dot{\theta} = \theta^*$ if $\tilde{\theta}$ is not in S . Since $\sqrt{n}\,(\dot{\theta} - \tilde{\theta})$ converges almost surely to the zero vector by Theorem 2, it will suffice to prove the theorem for $\dot{\theta}$.

By Taylor's theorem we may write

$$\delta(x_t,\dot{\theta}) = -\nabla' f(x_t,\theta^*)(\dot{\theta} - \theta^*) - \frac{1}{2}(\dot{\theta} - \theta^*)'\nabla^2 f(x_t,\bar{\theta})(\dot{\theta} - \theta^*)$$

where $\bar{\theta}$ varies with t and $\dot{\theta}$ and lies on the line segment joining θ^* to $\dot{\theta}$. Similarly, the vector

$$\Sigma_{t=q+1}^{n} u_{t-r}\nabla f(x_{t-s},\dot{\theta}) = \Sigma_{t=q+1}^{n} u_{t-r}\nabla f(x_{t-s},\theta^*) + D(r,s)(\dot{\theta} - \theta^*)$$

where the matrix D(r,s) has typical element

$$d_{ij}(r,s) = \Sigma_{t=q+1}^{n} u_{t-r}(\partial^2/\partial\theta_i\partial\theta_j) f(x_{t-s},\bar{\theta}) .$$

As a consequence of the Uniform Strong Law, $(1/n)d_{ij}(r,s)$ converges almost surely to zero.

Using these expressions we can write

$$(-\sqrt{n}/2)\nabla Q_n(\dot{\theta}) = -\tfrac{1}{2}(\hat{\sigma}^2/\sqrt{n})\nabla\Sigma_{i=1}^{q}\Sigma_{j=1}^{q}[u_i + \delta(x_i,\dot{\theta})]\hat{\gamma}_q^{ij}[u_j + \delta(x_j,\dot{\theta})]$$

$$+ \Sigma_{i=0}^{q}\Sigma_{j=0}^{q}\hat{a}_i\hat{a}_j(1/\sqrt{n})\Sigma_{t=q+1}^{n}\nabla f(x_{t-i},\theta^*)u_{t-j}$$

$$- [\Sigma_{i=0}^{q}\Sigma_{j=0}^{q}\hat{a}_i\hat{a}_j A_n(i,j)]\sqrt{n}\,(\dot{\theta} - \theta^*)$$

$$= -\alpha_n + \beta_n - \gamma_n$$

where

$$A_n(i,j) = (1/n)\Sigma_{t=q+1}^{n} \nabla f(x_{t-i},\dot{\theta}) \nabla' f(x_{t-j},\theta^*)$$

$$+ (1/n)\Sigma_{t=q+1}^{n} \nabla f(x_{t-i},\dot{\theta})(\dot{\theta} - \theta^*)' \nabla^2 f(x_{t-j},\bar{\theta})$$

$$- (1/n)D(j,i) \ .$$

Note that if $i \geq j$ then $A_n(i,j)$ converges almost surely to $V(i-j)$ and if $i < j$ $A_n(i,j)$ converges almost surely to $V'(i-j)$.

The vector $(-\sqrt{n}/2)\nabla Q_n(\dot{\theta})$ converges to the zero vector almost surely because $\dot{\theta}$ is a stationary point of $Q_n(\theta)$ when $\tilde{\theta}$ is in S . The vector α_n converges almost surely to zero because it is a finite sum of random variables which converge almost surely divided by $\sqrt{n}$.

We may write

$$\beta_n = (1/\sqrt{n})\Sigma_{i=0}^{q} a_i \Sigma_{t=q+1}^{n} e_t \nabla f(x_{t-i},\theta^*)$$

$$+ \Sigma_{i=0}^{q} \hat{a}_i \sqrt{n}\ (\hat{a}_j - a_j)(1/n)\Sigma_{t=q+1}^{n} u_{t-j} \nabla f(x_{t-i},\theta^*)$$

$$+ \Sigma_{i=0}^{q} a_j \sqrt{n}\ (\hat{a}_i - a_i)(1/n)\Sigma_{t=q+1}^{n} u_{t-j} \nabla f(x_{t-i},\theta^*)$$

which converges in distribution to a p-variate normal with mean vector zero and variance-covariance matrix $\sigma^2 C$ where

$$C = \Sigma_{i=0}^{q}\Sigma_{j=0}^{q} a_i a_j \lim_{n\to\infty} A_n(i,j)$$

$$= \Sigma_{i=0}^{q}\Sigma_{j=0}^{q} a_i a_j \tfrac{1}{2}[V(i-j) + V'(i-j)]$$

in consequence of Corollary 1 of Theorem 5 of [10], the Uniform Strong Law, and the fact that $\sqrt{n}\,(\hat{a}_i - a_i)$ is bounded in probability. Thus, $\sqrt{n}\,(\dot{\theta} - \theta^*)$ is asymptotically normally distributed with mean vector zero and variance-covariance matrix $\sigma^2 C^{-1}$.

The second conclusion is obtained by writing

$$\tilde{C} = (\hat{\sigma}^2/n)\Sigma_{i=1}^{q}\Sigma_{j=1}^{q}\hat{\gamma}_q^{ij}\,\nabla f(x_i,\tilde{\theta})\nabla' f(x_i,\tilde{\theta})$$

$$+ \Sigma_{i=0}^{q}\Sigma_{j=0}^{q}\hat{a}_i\hat{a}_j(1/n)\Sigma_{t=q+1}^{n}\nabla f(x_{t-i},\tilde{\theta})\nabla' f(x_{t-j},\tilde{\theta}) .$$

The two terms on the right converge almost surely to the zero matrix and C, respectively, using arguments similar to those employed above. □

5. MONTE-CARLO SIMULATIONS

The simulations summarized in Tables 1 and 2 were performed to gain information on three questions:

1. Does the autoregressive estimation procedure described in Section 3 have better small sample efficiency than the Hannan-Goebel estimator?
2. Is the small sample distribution of the "t-ratios" $t_i = \sqrt{n}\,(\tilde{\theta}_i - \theta_i^*)/(\tilde{\sigma}^2 c_{ii})^{1/2}$ $(i = 1, 2, \ldots, p)$ derived from the asymptotic theory approximated by the t-distribution with $n-p$ degrees freedom with sufficient accuracy for use in applications?
3. Is the two-stage autoregressive estimator an improvement over the one-stage autoregressive estimator in small samples?

The Monte-Carlo evidence presented here indicates that the answers are:

1. Yes.

1. MONTE-CARLO ESTIMATES: MEAN, VARIANCE, MEAN SQUARE ERROR, AND RELATIVE MEAN SQUARE ERROR EFFICIENCY FOR EXPONENTIAL MODEL WITH n = 60[a]

Error Structure	Ordinary Least Squares				Hannan-Goebel				Autoregressive, q = 2: One Stage				Autoregressive, q = 2: Two Stage			
	Mean	Var.	M.S.E.	Eff.	Mean	Var.	M.S.E.	Eff.	Mean	Var.	M.S.E.	Eff.	Mean	Var.	M.S.E.	Eff.
							Parameter θ_1									
IID	.751	.00087	.00087	1	.749	.00086	.00086	0.92	.751	.00090	.00090	.96	.751	.00091	.00091	.96
MA(4)	.758	.00753	.00759	1	.756	.00658	.00661	1.28	.755	.00534	.00536	1.42	.755	.00536	.00538	1.41
AR(1)	.751	.00518	.00518	1	.751	.00394	.00394	1.34	.751	.00317	.00317	1.63	.751	.00315	.00315	1.65
AR(2)	.775	.04444	.04507	1	----	------	------	----	.762	.00950	.00963	4.68	.761	.00759	.00772	5.84
							Parameter θ_2									
IID	1.150	.00022	.00022	1	1.151	.00023	.00023	0.91	1.150	.00023	.00023	.96	1.150	.00024	.00024	.95
MA(4)	1.149	.00151	.00151	1	1.149	.00127	.00127	1.34	1.149	.00102	.00102	1.48	1.149	.00103	.00103	1.47
AR(1)	1.151	.00099	.00099	1	1.151	.00072	.00072	1.38	1.151	.00058	.00058	1.71	1.151	.00057	.00057	1.73
AR(2)	1.152	.00757	.00757	1	------	-----	------	----	1.148	.00153	.00153	4.95	1.147	.00122	.00122	6.19

[a]Computed from two thousand trials except the Hannan-Goebel estimator which was computed from four hundred trials for MA(4) and two hundred trials in the remaining cases.

Source: The values for the Hannan-Goebel estimator were obtained from [5].

2. EMPERICAL DISTRIBUTION OF "t-RATIOS" COMPARED TO THE t-DISTRIBUTION[a]

Tabular Values		Emperical Distribution								
		IID		MA(4)		AR(1)		AR(2)		
c	P[t ≤c]	P[t_1≤c]	P[t_2≤c]	P[t_1≤c]	P[t_2≤c]	P[t_1≤c]	P[t_2≤c]	P[t_1≤c]	P[t_2≤c]	Std. Error
					Ordinary Least Squares					
-2.663	.005	.0060	.0050	.0445	.0320	.0765	.0455	.2205	.1810	.0016
-2.002	.025	.0290	.0270	.1035	.0785	.1330	.1005	.2750	.2430	.0035
-1.672	.050	.0515	.0615	.1385	.1180	.1830	.1435	.3045	.2875	.0049
-0.679	.250	.2590	.2555	.3085	.3265	.3590	.3300	.4045	.4185	.0097
0.0	.500	.5000	.5045	.4805	.5240	.5005	.4985	.4820	.5190	.0112
0.679	.750	.7485	.7465	.6555	.7075	.6540	.6690	.5700	.6150	.0097
1.672	.950	.9435	.9535	.8690	.8990	.8280	.8715	.6860	.7465	.0049
2.002	.975	.9770	.9775	.9105	.9420	.8700	.9045	.7240	.7860	.0035
2.663	.995	.9960	.9950	.9640	.9800	.9360	.9585	.7940	.8570	.0016
					One Stage Autoregressive					
-2.663	.005	.0115	.0075	.0105	.0080	.0290	.0160	.0470	.0610	.0016
-2.002	.025	.0330	.0320	.0340	.0340	.0615	.0515	.0860	.1155	.0035
-1.672	.050	.0630	.0695	.0625	.0685	.0890	.0855	.1210	.1480	.0049
-0.679	.250	.2590	.2655	.2635	.2735	.2905	.2755	.2980	.3325	.0097
0.0	.500	.4995	.5015	.4810	.5240	.5060	.5020	.4875	.5110	.0112
0.679	.750	.7380	.7365	.7310	.7505	.7215	.7205	.6685	.7070	.0097
1.672	.950	.9370	.9455	.9450	.9560	.9175	.9230	.8585	.8960	.0049
2.002	.975	.9720	.9755	.9735	.9755	.9535	.9470	.9000	.9300	.0035
2.663	.995	.9950	.9895	.9950	.9955	.9830	.9860	.9465	.9645	.0016

				Two Stage Autoregressive						
-2.663	.005	.0115	.0075	.0100	.0075	.0290	.0155	.0350	.0480	.0016
-2.002	.025	.0330	.0330	.0350	.0330	.0615	.0485	.0725	.1000	.0035
-1.672	.050	.0635	.0700	.0610	.0665	.0905	.0810	.1075	.1360	.0049
-0.679	.250	.2600	.2650	.2630	.2710	.2870	.2715	.2805	.3260	.0097
0.0	.500	.5015	.5005	.4790	.5205	.5045	.5030	.4900	.5120	.0112
0.679	.750	.7370	.7350	.7350	.7505	.7255	.7220	.6695	.7230	.0097
1.672	.950	.9375	.9455	.9460	.9555	.9200	.9225	.8700	.9100	.0049
2.002	.975	.9720	.9745	.9750	.9760	.9570	.9485	.9155	.9425	.0035
2.663	.995	.9945	.9895	.9950	.9960	.9845	.9870	.9590	.9760	.0016

[a]Empirical distributions were computed from two thousand Monte-Carlo trials

2. No. On balance, the small sample distributions of the "t-ratios" have heavier tails than the t-distribution with n-p degrees freedom. Perhaps a reasonable procedure would be to enter tables of the t-distribution with n-p-q degrees of freedom in applications.

3. Ambiguous. Performance improved when the disturbances were, in fact, autoregressive and deteriorated slightly when they were not.

The details of the simulations are as follows. So as to be able to compare our results with Goebel's [5] we used his choice of a response function,

$$f(x,\theta) = \theta_1 e^{\theta_2 x} ,$$

his choice of inputs for $n = 60$ as shown in Table 3, his choice of parameters,

$$\theta = (.75,\ 1.15) ,$$

and his choice of error structures:

$$\text{IID} , \quad u_t = e_t ;$$

$$\text{MA(4)} , \quad u_t = 1.5e_t + 1.0e_{t-1} + .85e_{t-2} + .33e_{t-3} + .50e_{t-4} ;$$

$$\text{AR(1)} , \quad u_t - .735\, u_{t-1} = e_t ;$$

where in each case $e_t \sim \text{NID}(0,.25)$. We chose the autoregressive estimator with $q = 2$ for the simulations. With this choice, the autoregressive estimator is not exactly appropriate for any of these cases - IID, MA(4), and AR(1). However, one cannot really hope to exactly satisfy his assumptions in applications and we expect that a near miss such as used here is a more realistic imitation of an applied situation. To gain information on what would happen were our assumptions exactly satisfied we included a fourth case not considered by Goebel:

$$\text{AR(2)} \ , \quad u_t - 1.04\, u_{t-1} + .128\, u_{t-2} = e_t \ ;$$

where $e_t \sim \text{NID}(0,.25)$. The gains in mean square error efficiency relative to ordinary least squares are dramatic for this case as seen in Table 2.

Relying on the asymptotic theory for purposes of statistical inference - such as finding a confidence interval for θ_i - one would enter tables of the t-distribution with n-p degrees freedom using the statistic

$$t_i = \sqrt{n}\ (\tilde{\theta}_i - \theta_i^*)/(\tilde{\sigma}^2 \tilde{c}_{ii})$$

where $\tilde{c}_{ii}$ is the $i^{\underline{th}}$ diagonal element of $\tilde{C}$ as defined in Section 2. (Set $\hat{P} = I$ in the formulae for $\tilde{\sigma}^2$ and $\tilde{C}$ to obtain the appropriate quantities for ordinary least squares.) We see from Table 2 that if one uses ordinary least squares formulae when the disturbances are, in fact, autocorrelated his probability statements can be quite erroneous - confidence intervals, e.g., would be much too narrow. We could argue, on the basis of Table 2, that one should use the formulae of Section 2 not so much as to gain efficiency in non-linear time series regressions, but so as to be able to make reasonably accurate probability statements in applications.

The standard errors shown in Table 2 refer to the fact that if t_i does, in fact, follow the t-distribution then the Monte-Carlo estimate of $P[t_i \leq c]$ has a standard error of $\{P[t \leq c] \cdot P[t < c]/2000\}^{1/2}$.

3. INPUTS FOR THE SIMULATIONS

Inputs		
1.32040	2.12500	2.02300
2.42100	2.09400	2.00200
2.12300	2.98500	2.98600
3.00200	2.45300	1.33200
2.65200	1.54200	2.00123
1.03300	2.03600	2.54000
1.56300	2.65400	1.30000
2.10300	2.75400	1.65000
1.00330	1.23000	1.03300
2.45000	2.06680	2.03600
2.40000	2.00300	2.65400
1.56000	2.20300	2.75400
1.77000	1.00330	1.23000
1.23068	2.45000	2.06680
2.02000	2.40000	2.00300
2.75000	1.56000	1.32100
0.99800	1.77000	2.02300
1.65400	1.23068	2.42100
2.56800	2.02000	2.12300
2.12300	2.75000	3.00200

REFERENCES

[1] Anderson, T. W., The Statistical Analysis of Time Series, New York: John Wiley and Sons, Inc., 1971.

[2] Box, G. E. P. and Jenkins, G. M., Time Series Analysis: Forecasting and Control, San Francisco: Holden-Day, Inc., 1970.

[3] Duncan, D. B. and Jones, R. H., "Multiple Regression with Stationary Errors", Journal of the American Statistical Association, (December 1966), 917-928.

[4] Fuller, W. A. Introduction to Statistical Time Series, Ames, Iowa: University Bookstore, Iowa State University, 1971.

[5] Goebel, J. J., "Nonlinear Regression in the Presence of Autocorrelated Errors", Ph.D. Dissertation, Iowa State University, 1974.

[6] Hannan, E. J. "Nonlinear Time Series Regression", Journal of Applied Probability, 8 (December 1971), 767-80.

[7] Hartley, H. O. "The Modified Gauss-Newton Method for the Fitting of Nonlinear Regression Functions by Least Squares," Technometrics, 3 (May 1961), 269-80.

[8] I.B.M. Corporation, System/360 Scientific Subroutine Package, Version III, White Plains, New York: I.B.M. Corporation, Technical Publications Department, 1968.

[9] Jenkins, G. M. and Watts, D. G. Spectral Analysis and Its Applications, San Francisco: Holden-Day, Inc., 1968.

[10] Jennrich, R. J. "Asymptotic Properties of Nonlinear Least Squares Estimators", The Annals of Mathematical Statistics, 40 (March 1969), 533-43.

[11] Marquardt, D. W. "An Algorithm for Least Squares Estimation of Nonlinear Parameters", Journal of the Society for Industrial and Applied Mathematics, 11 (June 1963), 431-41.

[12] Office of the President, Economic Report of the President, Transmitted to Congress January 1972, Washington, D. C.: U. S. Government Printing Office, 1972.

[13] Office of the President, Economic Report of the President, Transmitted to Congress February 1974, Washington, D. C.: U. S. Government Printing Office, 1974.

[14] U. S. Bureau of the Census, Historical Statistics of the United States, Washington, D. C.: U.S. Government Printing Office, 1960.