The Annals of Statistics
1974, Vol. 2, No. 1, 99-108

ESTIMATION OF MODELS OF AUTOREGRESSIVE SIGNAL PLUS WHITE NOISE[1]

BY MARCELLO PAGANO

State University of New York at Buffalo

If $x(\cdot)$ is a time series which may be written as $x(t) = s(t) + n(t)$ where t is an integer, $s(\cdot)$ an autoregressive signal of order q and $n(\cdot)$ white noise, then the model has $q + 2$ parameters. These are (i) the q autoregressive parameters (ii) the residual variance of the autoregressive scheme and (iii) the variance of the white noise. A method is proposed to estimate the $q + 2$ parameters. This method is based on analogies with regression theory and in the case of a normal series yields strongly consistent efficient estimators.

1. Introduction. The model in which we are interested is, for integer t,

$$x(t) = s(t) + \eta(t), \qquad \sum_{j=0}^{q} \beta(j)s(t-j) = \varepsilon(t), \qquad q \geqq 1 \tag{1}$$

where the

(2) (i) $\{s(\cdot)\}$ and $\{\eta(\cdot)\}$ are independent,

(ii) $\{n(\cdot)\}$ are independent identically distributed $N(0, \sigma_n^2)$ and $\{\varepsilon(\cdot)\}$ i.i.d. $N(0, \sigma_\varepsilon^2)$,

(iii) $\beta(0) = 1$, $B(q) \neq 0$ and $\{s(\cdot)\}$ stationary.

This scheme has $q + 2$ parameters, viz. $\beta(1), \cdots, \beta(q), \sigma_n^2, \sigma_\varepsilon^2$. We have available a sample $x(1), \cdots, x(N)$ from which we wish to draw inference about the parameters of the scheme.

For some instances where this model is of importance, see Parzen (1967). Anderson, Kleindorfer, Kleindorfer and Woodroofe (1969) consider the vector analogue of the model (1) with $q = 1$, obtain strongly consistent estimators of the parameters of the system and study the effect of using these estimators in the Kalman prediction formulas. Since the estimators we obtain are strongly consistent, the effect of using our estimators in the Kalman filtering formulas can be obtained from Anderson *et al.* (1969).

Our method of solving the problem is to first introduce $(q + 1)$ new parameters and obtain initial estimates of these new parameters and the original $(q + 2)$ parameters. The reason for doing this is that by introducing the new parameters the model is then transformed into a mixed autoregressive-moving average model (see Whittle (1963) page 35), for which the solution to the estimation problem is known. Having obtained these initial estimates, we obtain "better" estimates by expressing these new parameters in terms of the original parameters.

Received April 1972; revised April 1973.

[1] Research supported by National Science Foundation Grant No. GP 29069.

AMS 1970 *subject classifications.* Primary 6285; Secondary 6255.

Key words and phrases. Autoregressive-moving average, non-linear regression, Gauss-Newton, spectral averages.

Define the series $\{y(\cdot)\}$ by

(3) $$y(t) = \sum_{j=0}^{q} \beta(j)x(t-j) .$$

LEMMA 1. *There exists a sequence of* i.i.d. $\mathscr{N}(0, \sigma_\eta^2)$ *random variables* $\{\eta(\cdot)\}$, *defined on the same sample space as* $\{x(\cdot)\}$ *and constants* $\alpha(1), \cdots, \alpha(q)$, *satisfying*

(4) $$y(t) = \eta(t) + \sum_{j=1}^{q} \alpha(j)\eta(t-j) .$$

PROOF. From (1) and (3), $y(t) = \varepsilon(t) + \sum_{j=0}^{q} \beta(j)n(t-j)$. Clearly $\{y(\cdot)\}$ is stationary. Define

$$R_Y(v) = \mathscr{E}y(t)y(t+v) \qquad \forall v \in Z ,$$

which, from (3), is zero for $|v| > q$. Therefore $\{y(\cdot)\}$ is a moving average scheme of order less than or equal to q. Thus the existence of the $\{\eta(\cdot)\}$ sequence.

The normality of the η's follows from the fact that the $\{y(\cdot)\}$ process is normal; thus the η's may be chosen to be independent and if independent they must be normal, from a theorem of Cramér (Loève page 272). □

Define the complex polynomials

$$h(z) = \sum_{k=0}^{q} \beta(k)z^k \qquad \text{and} \qquad g(z) = \sum_{j=0}^{q} \alpha(j)z^j , \qquad \alpha(0) \equiv 1 .$$

LEMMA 2. *The* $\{y(\cdot)\}$ *sequence defined by* (3) *is a moving average scheme of order* q (*i.e.* $\alpha(q) \neq 0$). *Moreover,* $\{x(\cdot)\}$ *is a proper mixed autoregressive-moving average scheme of order* (q, q).

PROOF. From Lemma 1, $R_Y(q) = \sigma_\eta^2\alpha(q) = \sigma_n^2\beta(q)$ which by hypothesis is not zero. Since from Lemma 1 the $\{y(\cdot)\}$ is a moving average of order less than or equal to q, $\alpha(q) \neq 0$ implies that the order is q.

From (3) and (4) $f_x(w) = \sigma_\eta^2|g(e^{iw})|^2/2\pi|h(e^{iw})|^2$, and from Lemma 1

(5) $$\sigma_\eta^2 g(z)g(z^{-1}) = \sigma_\varepsilon^2 + \sigma_n^2 h(z)h(z^{-1}) .$$

Therefore, if z_0 is a root of $h(\cdot)$ then neither z_0 nor z_0^{-1} is a root of $g(\cdot)$. Thus, $f_x(\cdot)$ is the modulus squared of the ratio of two qth degree polynomials with no common roots and so $\{x(\cdot)\}$ is a proper mixed autoregressive-moving average scheme of order (q, q). □

LEMMA 3. *The* $\alpha(1), \cdots, \alpha(q)$ *may be chosen such that the roots of* $g(\cdot)$ *all lie outside the unit circle.*

PROOF. Define the complex function $\rho(z) = \sum_{j=-q}^{q} R_Y(j)z^j$. Then from (4) the α's must be such that $\rho(z) = \sigma_\eta^2 g(z)g(z^{-1})$. Now from (5), $\sigma_\eta^2|g(e^{iw})|^2 = \sigma_\varepsilon^2 + \sigma_n^2|h(e^{iw})|^2$ for $\omega \in [0, 2\pi)$. Therefore, no root of $g(\cdot)$ has unit modulus; thus the same holds true for $\rho(\cdot)$. Using the symmetry of $\rho(\cdot)$, it seems clear which "square root," $g(\cdot)$, to choose to satisfy the lemma (see Wilson (1969)). □

Thus our sample $x(1), \cdots, x(N)$ may be viewed as a sample from a mixed autoregressive-moving average scheme

(6) $$\sum_{j=0}^{q} \beta(j)x(t-j) = \sum_{j=0}^{q} \alpha(j)\eta(t-j) ,$$

which has $2q + 1$ parameters, $\beta(1), \cdots, \beta(q), \alpha(1), \cdots, \alpha(q)$ and σ_η^2. Equivalently, the parameter set $\beta(1), \cdots, \beta(q), R_Y(0), \cdots, R_Y(q)$ defines the model. Furthermore, these parameters are such that the roots of $g(\cdot)$ and the roots of $h(\cdot)$ lie outside the unit circle (see Pagano (1971)) and $g(\cdot)$ and $h(\cdot)$ have no common roots. These conditions being satisfied the problem of estimating $\beta(1), \cdots, \beta(q)$ and $R_Y(0), \cdots, R_Y(q)$ has been solved (see Hannan (1970) Chapter VI or Parzen (1971)).

2. The information matrix. We have available an efficient set of estimates for the set of parameters $\beta(1), \cdots, \beta(q), R_Y(0), \cdots, R_Y(q)$ of our enlarged model (6). That these estimates are not efficient for the parameters $\beta(1), \cdots, \beta(q), \sigma_n^2$ and σ_ε^2 of our original model (1) seems clear, especially when one considers the fact that the R's are functions of the β's and the information matrix of the $\beta(1), \cdots, \beta(q), R_Y(0), \cdots, R_Y(q)$ is not block diagonal.

In this section we obtain the information matrix of the parameters $\beta(1), \cdots, \beta(q), \sigma_n^2$ and σ_ε^2 in the original model (1).

Since the $\{x(\cdot)\}$ are normally distributed we can use a formula from Clevenson (1970, Theorem 1.13) to obtain the required information matrix. If we denote the parameters of the scheme by $\mu_1, \cdots, \mu_{q+2}$ and the information matrix by $\mathscr{I}$, then

$$\mathscr{I}_{j,k} = \frac{1}{4\pi} \int_{-\pi}^{\pi} \left(\frac{\partial}{\partial \mu_j} \log f_x(\omega)\right)\left(\frac{\partial}{\partial \mu_k} \log f_x(\omega)\right) d\omega$$

for $j, k = 1, \cdots, q + 2$. So in our case, if we let

$$\boldsymbol{\theta}^T = \left(\frac{\partial}{\partial \beta(1)}, \cdots, \frac{\partial}{\partial \beta(q)}, \frac{\partial}{\partial \sigma_n^2}, \frac{\partial}{\partial \sigma_\varepsilon^2}\right) \log f_x(\omega)$$

then $\mathscr{I} = \int_{-\pi}^{\pi} \boldsymbol{\theta}\boldsymbol{\theta}^T \, d\omega/4\pi$. To obtain $\boldsymbol{\theta}$, we have from (3) and (4) that

$$f_x(\omega) = (\sigma_\varepsilon^2 + \sigma_n^2 |h(e^{i\omega})|^2)/2\pi |h(e^{i\omega})|^2 \,. \tag{7}$$

For $j = 1, \cdots, q$ define $\Psi_j(\omega) = e^{ij\omega}h(e^{-i\omega}) + e^{-ij\omega}h(e^{i\omega})$, then the derivative of $\log f_x(\omega)$ with respect to: σ_n^2 is $(2\pi f_x(\omega))^{-1}$, σ_ε^2 is $(2\pi f_x(\omega)|h(e^{i\omega})|^2)^{-1}$ and $\beta(j)$ is

$$\sigma_\varepsilon^2 \Psi_j(\omega)/2\pi f_x(\omega)|h(e^{i\omega})|^4 = \Psi_j(\omega)(\sigma_n^2(2\pi f_x(\omega))^{-1} - 1)/|h(e^{i\omega})|^2 \,.$$

We now have $\boldsymbol{\theta}$ and thus $\mathscr{I}$, the required information matrix.

3. The efficient estimators. Our aim is to obtain efficient estimators of $\beta(1), \cdots, \beta(q), \sigma_n^2, \sigma_\varepsilon^2$. By this we mean that the asymptotic covariance matrix of the estimators is the same as that for the maximum likelihood estimators, namely, $N^{-1}\mathscr{I}^{-1}$.

We recognize the fact that the R_Y's are functions of the $\beta(1), \cdots, \beta(q), \sigma_n^2$ and σ_ε^2. Indeed from Lemma 1

$$R_Y(k) = \sigma_\varepsilon^2 \delta_{0,k} + \sigma_n^2 \textstyle\sum_{j=0}^{q} \beta(j)\beta(j+k) \,, \qquad k = 0, 1, \cdots, q \,. \tag{8}$$

If we define $\mathbf{R}_Y = (R_Y(0), \cdots, R_Y(q))^T$ and $\boldsymbol{\beta} = (\beta(0), \cdots, \beta(q))^T$ then we may denote the relationships in (8) by $\mathbf{R}_Y = \mathbf{R}_Y(\boldsymbol{\beta}, \sigma_n^2, \sigma_\varepsilon^2)$.

Our method of solving the problem is to regress the estimators of $\boldsymbol{\beta}$ and $\mathbf{R}_Y$ on $\boldsymbol{\beta}$, σ_n^2 and σ_ε^2 using (8). To then solve the regression problem by finding the least squares estimators of $\boldsymbol{\beta}$, σ_n^2 and σ_ε^2. And then show that the estimators found in this fashion are indeed efficient and strongly consistent.

This is a non-linear regression and the theory required is developed in Section 6. To enable us to use this theory, we must first lay some groundwork.

Denote the efficient estimators of $\mathbf{R}_Y$ and $\boldsymbol{\beta}$ in the enlarged model obtained at the end of Section 1 by $\hat{\mathbf{R}}_Y$ and $\hat{\boldsymbol{\beta}}$. Then we have from Parzen (1971) that $(\hat{\boldsymbol{\beta}}^T, \hat{\mathbf{R}}_Y{}^T) \to (\boldsymbol{\beta}^T, \mathbf{R}_Y{}^T)$ in mean square as $N \to \infty$ and

$$\mathscr{L}\left\{N^{\frac{1}{2}}\left(\begin{pmatrix}\hat{\boldsymbol{\beta}}\\ \hat{\mathbf{R}}_Y\end{pmatrix} - \begin{pmatrix}\boldsymbol{\beta}\\ \mathbf{R}_Y\end{pmatrix}\right)\right\} \to \mathscr{N}(\mathbf{0}; E^{-1}) \tag{9}$$

where (see Parzen (1971)) the information matrix

$$E = \begin{pmatrix} E^{\beta\beta} & E^{\beta R} \\ E^{\beta\beta} & E^{RR}\end{pmatrix}, \qquad E^{\beta R} = (E^{R\beta})^T \tag{10}$$

and

$$(E^{\beta\beta})_{jk} = \int_{-\pi}^{\pi} \frac{\Psi_j(\omega)\Psi_k(\omega)}{4\pi|h(e^{i\omega})|^4}\, d\omega$$

for $j, k = 1, \cdots, q$ and Ψ defined in Section 2,

$$(E^{\beta R})_{jk} = -\int_{-\pi}^{\pi} \frac{\Psi_j(\omega)W_k(\omega)}{2\sigma_\eta^2|h(e^{i\omega})|^2|g(e^{i\omega})|^2}\, d\omega$$

with $W_0(\omega) = 1/2\pi$ and $W_k(\omega) = (\cos k\omega)/\pi$, $k = 1, 2, \cdots$ for $j = 1, \cdots, q$ and $k = 0, \cdots, q$ and

$$(E^{RR})_{jk} = \pi \int_{-\pi}^{\pi} \frac{W_j(\omega)W_k(\omega)}{\sigma_\eta^4|g(e^{i\omega})|^4}\, d\omega$$

for $j, k = 0, \cdots, q$. Parzen (1971) also gives an estimator $\hat{E}_N$ of E.

To enable us to use the results obtained in Section 5, we require strongly convergent estimators of $\boldsymbol{\beta}$, $\mathbf{R}_Y$ and E. Hannan (1970, page 392 and pages 409 *et seq.*) does give strongly convergent estimators of the parameters of the mixed scheme, but he uses a different parametrization of the scheme, namely, in terms of the β's and α's in equation (6). We can use Hannan's results, make the appropriate transformations from his parametrization to ours and verify that we have strongly convergent estimators. The asymptotic normality in (9) is obtained by a straightforward application of the delta method (Rao, 1965, page 322). Alternatively, one can verify that Parzen's (1971) estimators are indeed strongly convergent. To this end we notice that, since $\{x(\cdot)\}$ is mixing and thus ergodic, the sample covariances converge almost surely (see Hannan, 1970, Chapter IV). Thus the only source of difficulty might be the spectral averages involved. To see that these do not in fact present any difficulty we have Theorem 2.

Thus we have random variables $\hat{\boldsymbol{\beta}}$, $\hat{\mathbf{R}}_Y$ which converge almost surely to $\boldsymbol{\beta}$ and $\mathbf{R}_Y(\boldsymbol{\beta}, \sigma_n^2, \sigma_\varepsilon^2)$, respectively. Furthermore, we also have $\hat{E}_N$ which is a strongly convergent estimator of E.

Define the $(2q + 1) \times (q + 2)$ matrix

$$A = \begin{pmatrix} I_q & \mathbf{0} & \mathbf{0} \\ A_\beta & \mathbf{a}_n & \mathbf{a}_\varepsilon \end{pmatrix} \tag{11}$$

where I_q is the q-dimensional identity matrix,

$$a_\varepsilon = (1, 0, \cdots, 0)^T ,$$
$$\mathbf{a}_n = (a_n(0), \cdots, a_n(q))^T$$

where

$$a_n(j) = \sum_{k=0}^{q-j} \beta(k)\beta(k+j)$$

and $A_\beta = \partial \mathbf{R}_Y / \partial \boldsymbol{\beta}$, i.e. $(A_\beta)_{kj} = \sigma_n^2\{[\beta(j+k)] + [\beta(j-k)]\}$, with

$$\begin{aligned} [\beta(j)] &= \beta(j) \qquad j = 0, \cdots, q , \\ &= 0 \qquad \text{otherwise} \end{aligned}$$

for $j = 1, \cdots, q$ and $k = 0, \cdots, q$. Note that A is in fact the derivative of $(\boldsymbol{\beta}^T, \mathbf{R}_Y{}^T)^T$ with respect to $(\boldsymbol{\beta}^T, \sigma_n^2, \sigma_\varepsilon^2)^T$. Furthermore, from (14), (22) and the chain rule for vector derivatives

$$\mathscr{I} = A^T E A , \tag{12}$$

and since $\beta(q) \neq 0$ this implies that A is of full rank.

Now if we solve the non-linear regression (using the theory of Section 6)

$$\begin{pmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\mathbf{R}}_Y \end{pmatrix} = \begin{pmatrix} \boldsymbol{\beta} \\ \mathbf{R}_Y(\boldsymbol{\beta}, \sigma_n^2, \sigma_\varepsilon^2) \end{pmatrix} + \mathbf{z}_N$$

with

$$\mathbf{z}_N \to 0 \quad \text{a.s.} \qquad \text{as } N \to \infty$$

and

$$\mathscr{L}(N^{\frac{1}{2}}\mathbf{z}_N) \to \mathscr{N}(\mathbf{0}; E)$$

and denote the least squares estimators by $\tilde{\boldsymbol{\beta}}$, $\tilde{\sigma}_n^2$, $\tilde{\sigma}_\varepsilon^2$, we see from Section 6, that as $N \to \infty$

$$(\tilde{\boldsymbol{\beta}}^T, \tilde{\sigma}_n^2, \tilde{\sigma}_\varepsilon^2) \to (\boldsymbol{\beta}^T, \sigma_n^2, \sigma_\varepsilon^2) \quad \text{a.s.} ,$$

$$\mathscr{L}\left\{N^{\frac{1}{2}}\left(\begin{pmatrix} \tilde{\boldsymbol{\beta}} \\ \tilde{\sigma}_n^2 \\ \tilde{\sigma}_\varepsilon^2 \end{pmatrix} - \begin{pmatrix} \boldsymbol{\beta} \\ \sigma_n^2 \\ \sigma_\varepsilon^2 \end{pmatrix}\right)\right\} \to \mathscr{N}(\mathbf{0}, \mathscr{I}^{-1})$$

by (12) and

$$A^T(\tilde{\boldsymbol{\beta}}, \tilde{\sigma}_n^2)\hat{E}_N A(\tilde{\boldsymbol{\beta}}, \tilde{\sigma}_n^2) \to \mathscr{I} \quad \text{a.s.}$$

These results hold provided that $(\boldsymbol{\beta}^T, \sigma_n^2, \sigma_\varepsilon^2)$ is an interior point of the parameter space. That this is always true follows from the following:

LEMMA 4. *Under assumption* (iii) (*Section* 1) *the parameter space of* $(\boldsymbol{\beta}^T, \sigma_n^2, \sigma_\varepsilon^2)$ *is an open subset of* R^{q+2}.

PROOF. The parameter space may be written $B \times (0, \infty) \times (0, \infty)$. To show that B is an open subset of R^q write

$$h(z) = \prod_{j=1}^{q} (1 - r_j z)$$

where the r_j^{-1} are the roots of $h(z) = 0$. From the stationarity assumption (iii) we obtain that the permissible parameter space for the r_j is (see Pagano (1971))

$$C = \{r_j: |r_j| < 1, j = 1, \cdots, q\} .$$

Thus C is an open set in R^q and since the mapping from C to B is continuous, B is open in R^q. $\square$

We thus have our efficient estimators of $\boldsymbol{\beta}$, σ_n^2 and σ_ε^2. To initialize the Gauss–Newton iterations we use $\hat{\boldsymbol{\beta}}$ and the consistent estimators of σ_n^2 and σ_ε^2 which are obtained in the next section.

4. Initial consistent estimators for σ_n^2 and σ_ε^2. From Lemma 2 we have that

$$\sigma_\eta^2|g(e^{i\omega})|^2 = \sigma_\varepsilon^2 + \sigma_n^2|h(e^{i\omega})|^2 ,$$

or equivalently

$$f_Y(\omega) = \sigma_\varepsilon^2 + \sigma_n^2|h(e^{i\omega})|^2 ,$$

for all $\omega \in [0, 2\pi)$ and in particular for $\omega_j = \pi j/q$, $j = 0, \cdots, q$.

For $j = 0, \cdots, q$ let

$$\xi_j = \hat{R}_Y(0) + 2 \sum_{k=1}^q \hat{R}_Y(k) \cos k\omega_j$$

and

$$\nu_j = |\sum_{j=0}^q \hat{\beta}(j)e^{i\omega}j|^2 .$$

Consider as estimators of σ_ε^2 and σ_n^2 the solution to the normal equations:

$$\begin{pmatrix} (q+1) & \sum_{j=0}^q \nu_j \\ \sum_{j=1}^q \nu_j & \sum_{j=0}^q \nu_j^2 \end{pmatrix} \begin{pmatrix} \hat{\sigma}_\varepsilon^2 \\ \hat{\sigma}_n^2 \end{pmatrix} = \begin{pmatrix} \sum_{j=0}^q \xi_j \\ \sum_{j=0}^q \nu_j \xi_j \end{pmatrix} .$$

These are strongly consistent estimators of σ_ε^2 and σ_n^2 (see Parzen (1961) and Theorem 2).

5. Convergence of spectral averages. In this section we prove two theorems which show that under certain conditions spectral averages converge strongly.

Given a sample $x(1), \cdots, x(N)$ from a stationary time series $\{x(\cdot)\}$ with spectral density $f(\cdot)$, define the periodogram

$$f_N(\lambda) = \frac{1}{2\pi N} |\sum_{j=1}^N x(j)e^{-ij\lambda}|^2 .$$

THEOREM 1. *Suppose $h(\cdot)$, is a continuous function. Define*

$$J_N(h) = \frac{\pi}{N} \sum_{j=1-N}^{N-1} f_N(\lambda_j)h(\lambda_j)$$

where $\lambda_j = \pi j/N$, $j = 1 - N, \cdots, N - 1$, and

$$J(h) = \int_{-\pi}^{\pi} f(\lambda)h(\lambda)\, d\lambda .$$

If, as $N \to \infty$,

(13) $$J_N(e^{ik\omega}) \to J(e^{ik\omega}) \quad \text{a.s.}$$

for all integer k, then $J_N(h) \to J(h)$ a.s.

Before proving the theorem, we note that condition (13) is equivalent to asserting that the sample covariances strongly converge to the autocovariances of the scheme. Hannan and Robinson (1973) have a similar but slightly different theorem.

PROOF. Since $h(\cdot)$ is a continuous function, from the Weierstrass theorem we can find a trigonometric polynomial $h_1(\cdot)$ such that

$$\max_{\lambda \in [-\pi,\pi]} |h(\lambda) - h_1(\lambda)| < \varepsilon$$

for arbitrary $\varepsilon > 0$. Now

$$|J_N(h - h_1)| \leqq J_N(|h - h_1|) < \varepsilon J_N(1) .$$

So from (14)

$$\limsup_{N\to\infty} |J_N(h) - J_N(h_1)| \leqq \varepsilon \limsup_{N\to\infty} J_N(1) = \varepsilon J(1) \quad \text{a.s.} ,$$

and by hypothesis $\lim_{N\to\infty} J_N(h_1) = J(h_1)$ a.s. and $|J(h_1) - J(h)| \leqq \varepsilon J(1)$. Now

$$\begin{aligned}\limsup_{N\to\infty} |J_N(h) - J(h)| &\leqq \limsup_{N\to\infty} |J_N(h) - J_N(h_1)| \\ &\quad + \limsup_{N\to\infty} |J_N(h_1) - J(h_1)| + |J(h_1) - J(h)| \\ &< 2\varepsilon J(1) \quad \text{a.s.}\end{aligned}$$

and since ε was arbitrary, the theorem follows. □

THEOREM 2. *If $h(\cdot)$ is a continuous function and $\{h_N(\cdot)\}$ is a sequence of random functions with*

$$\sup_{\lambda\in[-\pi,\pi]} |h_N(\lambda) - h(\lambda)| \to 0 \quad \text{a.s.} , \qquad N \to \infty$$

and the condition (13) *holds, then $J_N(h_N) \to J(h)$ a.s.*

PROOF. From Theorem 1, $J_N(h) \to J(h)$ a.s. The theorem follows from the inequality

$$|J_N(h_N) - J_N(h)| \leqq \sup_{\lambda\in[-\pi,\pi]} |h_N(\lambda) - h(\lambda)| J_N(1) .$$ □

6. Non-linear, weighted least squares. To show how to obtain efficient estimators of $\boldsymbol{\beta}$, σ_η^2 and σ_ε^2 we require the following non-linear, weighted least squares theory. The theory is closely related to that developed by Jennrich (1969), simpler in that we deal with fixed length vectors, but different in that our errors are not independent identically normally distributed.

Suppose $\{\mathbf{y}_n\}$ is a random sequence of $m \times 1$ vectors, $\boldsymbol{\theta}$ a $p \times 1$ vector ($p \leqq m$) and $\mathbf{f}: R^p \to R^m$, $\mathbf{f}$ being twice differentiable. Suppose furthermore that

(a) $\mathbf{y}_n \to \mathbf{f}(\boldsymbol{\theta}_0)$ a.s. as $n \to \infty$,
(b) $\mathbf{y}_n = \mathbf{f}(\boldsymbol{\theta}_0) + \mathbf{z}_n$

where

$$\mathscr{L}(n^{\frac{1}{2}}\mathbf{z}_n) \to \mathscr{N}(\mathbf{0}; V) , \qquad n \to \infty$$

and

(c) $\{V_n\}$ is a sequence of random matrices with $V_n \to V$ a.s. as $n \to \infty$.

Our aim is to show that we can use the theory and tools developed by Jennrich to prove that a strongly convergent least squares estimator of $\boldsymbol{\theta}_0$ can be found by the Gauss–Newton iteration technique.

Define Θ to be a compact subset of R^p, and assume that $\boldsymbol{\theta}_0$ is an interior point of Θ. Introducing the notation

$$||\mathbf{x} - \mathbf{y}||_D = (\mathbf{x} - \mathbf{y})^T D^{-1}(\mathbf{x} - \mathbf{y})$$

and

$$\langle \mathbf{x}, \mathbf{y} \rangle_D = \mathbf{x}^T D^{-1} \mathbf{y} ,$$

for positive definite D, define

$$Q(\boldsymbol{\theta}) = ||\mathbf{f}(\boldsymbol{\theta}_0) - \mathbf{f}(\boldsymbol{\theta})||_V \qquad \text{and} \qquad Q_n(\boldsymbol{\theta}) = ||\mathbf{y}_n - \mathbf{f}(\boldsymbol{\theta})||_{V_n} .$$

LEMMA (Jennrich). *Let Q be a real-valued function on $\Theta \times Y$ where Θ is a compact subset of a Euclidean space and Y is a measurable space. For each θ in Θ let $Q(\theta, y)$ be a measurable function of y and for each y in Y a continuous function of θ. Then there exists a measurable function $\hat{\theta}$ from Y into Θ such that for all y in Y,*

$$Q(\hat{\theta}(y), y) = \inf_\theta Q(\theta, y) .$$

In our case, the conditions of the lemma are satisfied. This ensures the existence of a least squares sequence $\{\hat{\boldsymbol{\theta}}_n\}$, i.e.

$$Q_n(\hat{\boldsymbol{\theta}}_n) = \inf_{\theta \in \Theta} Q_n(\boldsymbol{\theta}) .$$

LEMMA 5. *Under the assumptions* (a), (b) *and* (c)

$$\hat{\boldsymbol{\theta}}_n \to \boldsymbol{\theta}_0 \quad \text{a.s.} \qquad n \to \infty .$$

PROOF. We first note that from the positive definiteness of V,

$$Q(\boldsymbol{\theta}) = 0 \Leftrightarrow \boldsymbol{\theta} = \boldsymbol{\theta}_0 . \tag{15}$$

Secondly

$$\lim_{n\to\infty} Q_n(\boldsymbol{\theta}) = \lim_{n\to\infty} ||\mathbf{y}_n - \mathbf{f}(\boldsymbol{\theta})||_{V_n} = Q(\boldsymbol{\theta}) \quad \text{a.s.}$$

from (a) and (c). Now on a set of probability measure one,

$$\begin{aligned} \lim_{n\to\infty} Q_n(\hat{\boldsymbol{\theta}}_n) &= \lim_{n\to\infty} \inf_{\theta \in \Theta} ||\mathbf{y}_n - \mathbf{f}(\boldsymbol{\theta})||_{V_n} \\ &\leqq \lim_{n\to\infty} ||\mathbf{y}_n - \mathbf{f}(\boldsymbol{\theta})||_{V_n} , \qquad \forall \boldsymbol{\theta} \in \Theta \\ &= Q(\boldsymbol{\theta}) . \end{aligned}$$

Therefore $\lim_{n\to\infty} Q_n(\hat{\boldsymbol{\theta}}_n) \leqq \inf_{\theta \in \Theta} Q(\boldsymbol{\theta}) = 0$.

Therefore from (15), $\{\hat{\boldsymbol{\theta}}_n\}$ converges almost surely to $\boldsymbol{\theta}_0$ and the lemma is proved. □

Before proving further asymptotic properties of our estimators, we require the following lemma (see Jennrich, Theorem 4):

LEMMA 6. *If $\mathbf{l}: R^p \to R^m$ is continuous and if conditions* (a) *and* (c) *hold, then with $\{\mathbf{z}_n\}$ defined in* (b), *$\langle \mathbf{l}(\boldsymbol{\theta}), \mathbf{z}_n \rangle_{V_n}$ almost certainly converges to zero, uniformly for all $\boldsymbol{\theta} \in \Theta$.*

PROOF. Since $\mathbf{l}$ is continuous and $V_n \to V$ a.s., there is for every $\varepsilon > 0$ and

$\boldsymbol{\theta}^* \in \Theta$ a neighborhood M of $\boldsymbol{\theta}^*$ such that

$$||\mathbf{l}(\boldsymbol{\theta}) - \mathbf{l}(\boldsymbol{\theta}^*)||_{V_n} < \varepsilon \quad \text{a.s.}$$

for all $\boldsymbol{\theta}$ in M, and n sufficiently large. Now $\mathbf{z}_n \to \mathbf{0}$ a.s. and $|\langle \mathbf{l}(\boldsymbol{\theta}), \mathbf{z}_n \rangle_{V_n}| \leqq ||\mathbf{l}(\boldsymbol{\theta}) - \mathbf{l}(\boldsymbol{\theta}^*)||_{V_n}^{\frac{1}{2}} ||\mathbf{z}_n||_{V_n}^{\frac{1}{2}} + |\langle \mathbf{l}(\boldsymbol{\theta}^*), \mathbf{z}_n \rangle_{V_n}|$. Thus for almost all $(\mathbf{z}_n, V_n)$, Θ is covered by neighborhoods M such that

$$|\langle \mathbf{l}(\boldsymbol{\theta}), \mathbf{z}_n \rangle_{V_n}| < \varepsilon \qquad \text{for all} \quad \boldsymbol{\theta} \quad \text{in} \quad M,$$

for n sufficiently large. Since Θ is compact it is covered by a finite number of such neighborhoods and the lemma follows. □

(d) Define the matrix

$$A(\boldsymbol{\theta}) = \frac{\partial \mathbf{f}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$$

and assume $A(\boldsymbol{\theta}_0)$ is nonsingular.

THEOREM 3. *Under the assumptions* (a) *through* (d) *the sequence of least squares estimators* $\{\hat{\boldsymbol{\theta}}_n\}$ *are such that*

$$\mathscr{L}(n^{\frac{1}{2}}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)) \to \mathscr{N}(\mathbf{0}; (A^T V^{-1} A)^{-1})$$

and, furthermore,

$$A^T(\hat{\boldsymbol{\theta}}_n) V_n^{-1} A(\hat{\boldsymbol{\theta}}_n) \to A^T(\boldsymbol{\theta}_0) V^{-1} A(\boldsymbol{\theta}_0) \quad \text{a.s.}$$

PROOF. This is our analogue to Jennrich's Theorem 7. The proof is exactly the same as Jennrich's if we make use of Lemmas 5 and 6 and if we identify his inner products $(\cdot, \cdot)_n$, $(\cdot, \cdot)$ with our $\langle \cdot, \cdot \rangle_{V_n}$ and $\langle \cdot, \cdot \rangle_V$, respectively. □

To obtain the least squares estimators we use the Gauss–Newton iterative method, and to show that under certain conditions the method does converge, we call on Jennrich's Theorem 8.

THEOREM 4. *Let* $\{\hat{\boldsymbol{\theta}}_n\}$ *be a sequence of least squares estimators of* $\boldsymbol{\theta}_0$ *and let assumptions* (a) *through* (d) *hold. Then there exists a neighborhood* M *of* $\boldsymbol{\theta}_0$ *such that for almost every couplet* $(\mathbf{y}_n, V_n)$ *there is an* $m(\mathbf{y}_n, V_n)$, *such that the Gauss–Newton iteration will converge to* $\hat{\boldsymbol{\theta}}_n$ *from any starting value in* M *whenever* $n \geqq m$.

PROOF. With the modifications noted in Theorem 3 the proof is as in Jennrich (1969). □

7. Conclusions. We may generalize our model (1) by assuming that the signal is a mixed scheme, i.e.,

$$x(t) = s(t) + n(t)$$
$$\textstyle\sum_{j=0}^{q} \beta(j) s(t-j) = \sum_{k=0}^{p} \gamma(k) \varepsilon(t-k)$$

with $\gamma(0) = 1$ and $\gamma(p) \neq 0$, $p \geqq 1$. Our method of estimation may be generalized to this model on condition that $p < q$. The reason for this in the trivial case when $p = q = 0$ is that the $\{x(\cdot)\}$ is now a white noise sequence and σ_ε^2 and σ_n^2 are confounded. A similar thing happens assuming general p and q. We find that Lemmas 1, 2, and 3 hold and that $\{x(\cdot)\}$ is a proper mixed autoregressive-moving average scheme of order (q, ν) with $\nu = \max(p, q)$. But in our reparametrization, we have gone from the original $p + q + 2$ parameters to $q + \nu + 1$ parameters. Thus if $p \geqq q$ we have reparametrized to a lower dimensional space producing confounding. This is evidenced in the regression when we would have more parameters than observations. If $p < q$ the theory may be generalized in a straightforward manner.

This method of estimation can be easily used in other contexts when we have to perform constrained maxima. A simple example, which serves for illustration, is if we have a random sample from a normal (μ, μ^2) family. Form the sample mean and variance and regress these on μ and μ^2, having estimated the covariance matrix of the estimators.

8. Acknowledgments. For his suggestion of the problem, help in its solution, and Section 5, the author acknowledges Professor E. Parzen. The suggestions of the editor have been helpful.

REFERENCES

[1] Anderson, W. N., Kleindorfer, G. B., Kleindorfer, P. R. and Woodroofe, M. B. (1969). Consistent estimates of the parameters of a linear system. *Ann. Math. Statist.* **40** 2064–2075.

[2] Clevenson, M. L. (1970). Asymptotically efficient estimates of the parameters of a moving average time series. Statistics Department Technical Report No. 15, Stanford University.

[3] Hannan, E. J. (1970). *Multiple Time Series.* Wiley, New York.

[4] Hannan, E. J. and Robinson, P. H. (1973). Lagged regression with unknown lags. *J. Roy. Statist. Soc. Ser. B.* To appear.

[5] Jennrich, R. I. (1969). Asymptotic properties of non-linear least squares estimators. *Ann. Math. Statist.* **40** 633–643.

[6] Loève, M. (1963). *Probability Theory,* (3rd ed.). Van Nostrand, Princeton.

[7] Pagano, M. (1971). When is an autoregressive scheme stationary? Department of Statistics, Research Report No. 53, State Univ. of New York, Buffalo.

[8] Parzen, E. (1961). An approach to time series analysis. *Ann. Math. Statist.* **32** 951–989.

[9] Parzen, E. (1967). Time series analysis for models of signal plus white noise, in *Advanced Seminar on Spectral Analysis of Time Series,* (B. Harris, ed.) Wiley, New York.

[10] Parzen, E. (1971). Efficient estimation of stationary time series mixed schemes. Proc. 39th Session of I.S.I., Washington, D.C.

[11] Rao, C. R. (1965). *Linear Statistical Inference and Its Applications.* Wiley, New York.

[12] Whittle, P. (1963). *Prediction and Regulation by Linear Least-Square Methods.* The English Universities Press, London.

[13] Wilson, G. (1969). Factorization of the covariance generating function of a pure moving average process. *SIAM J. Numer. Anal.* **6** 1–7.

Statistical Science Division
State University of New York
Amherst, New York 14226