JOURNAL OF MATHEMATICAL ANALYSIS AND APPLICATIONS 59, 376–391 (1977)

Consistency and Asymptotic Normality of Least Squares Estimates Used in Linear Systems Identification*

D. O. NORRIS AND L. E. SNYDER

Department of Mathematics, Ohio University, Athens, Ohio 45701

Submitted by Masanao Aoki

Least squares estimation of the parameters of a single input–single output linear autonomous system is considered where both plant noise and observation noise are present. It is shown that under fairly general conditions the estimates converge almost surely to the true system parameters and that the estimates are asymptotically normal.

1. INTRODUCTION

In 1943 Mann and Wald published an article [9] dealing with the problem of estimating the parameters of a linear stochastic difference equation. In recent years this problem has again attracted attention, in particular in the realm of stochastic control theory for linear systems [1, 2, 6, 7, 11, 12]. The survey article [11] has an extensive bibliography. Two methods widely used to identify the parameters of a discrete time linear system are linear least squares and maximum likelihood. Each of these methods has a serious drawback. In the linear least squares case, the parameter estimates have an asymptotic bias if plant noise is present or if the observation noise is not white. The major drawback to use of the maximum likelihood method is that the distributions of the noise sequences must be known. Invariably it is assumed that the noise processes are Gaussian processes and in this case the properties of the Gaussian density make it relatively easy to derive maximum likelihood estimates.

Two important questions which arise when identifying the parameters of a linear system are concerned with consistency of the estimates and asymptotic normality of the estimates. As mentioned above, linear least squares estimates may fail to be consistent. The instrumental variable method as described by Wong and Polak [12] overcomes this difficulty by use of a weighting matrix. Aoki and Yue [1] showed that under the assumption that the noise processes

* This research was supported by the United States Air Force under Contract F33615-73-Q-4009.

ISSN 0022-247X

are Gaussian the parameter estimates are both consistent and asymptotically normal.

In this paper, we present a technique for estimating the parameters of a linear system which is based on an impulse response model. That is, the observed output can be described by

$$y_k = \sum_{j=0}^{\infty} h_j u_{k-j} + \text{noise},$$

where $\{u_k\}$ is a known input sequence and $\{h_k\}$ is the transfer function for a linear system with rational z-transform, viz.,

$$\sum_{k=0}^{\infty} h_k z^k = (b_1 z + \cdots + b_n z^n)/(1 + a_1 z + \cdots + a_n z^n).$$

The estimates for $\boldsymbol{\theta} = (a_1, \ldots, a_n, b_1, \ldots, b_n)$ are obtained by minimizing the mean square deviation between observed output and model output, i.e.,

$$\text{minimize} \sum_{k=0}^{N-1} \left(y_k - \sum_{j=0}^{\infty} h_j u_{k-j} \right)^2 \Big/ N.$$

Thus, the method is a least squares method based on the impulse response model. However, it does not provide a linear least squares estimate of $\boldsymbol{\theta}$.

In [10] the authors of this paper showed that the least squares estimates converge almost surely to the true system parameters if the plant noises and observation noises are both sequences of independent, identically distributed random variables with zero means and finite variances and if almost every sample sequence of plant noises is a bounded sequence. This last assumption of boundedness is quite severe. For example, it does not allow for Gaussian plant noise. In this paper we have simplified the proofs of some of the results leading to the almost sure convergence and we have removed the hypothesis that almost all sample sequences of the plant noises are bounded. Also, we show that the least squares estimates are asymptotically normal if the noises have finite third moments. Computational experience with the method is reported in [13] and the method appears to be quite robust.

2. Notation and Assumptions

The systems that are treated in this article are of the form

$$\begin{aligned} x_k + a_1^0 x_{k-1} + \cdots + a_n^0 x_{k-n} &= b_1^0 u_{k-1} + \cdots + b_n^0 u_{k-n} \\ &\quad + d_1^0 \xi_{n-1} + \cdots + d_n^0 \xi_{k-n}, \qquad (1) \\ y_k &= x_k + \eta_k, \end{aligned}$$

$k = 0, 1, 2, \ldots$, where quantities with negative subscripts are assumed to be zero. The sequences $\xi_0, \xi_1, \xi_2, \ldots$ and $\eta_0, \eta_1, \eta_2, \ldots$ are the plant noises and observation noises, respectively; $a_1^0, \ldots, a_n^0, b_1^0, \ldots, b_n^0$ are the parameters to be identified and are denoted by the vector $\boldsymbol{\theta}^0$.

Following the notation used in [1, 10], for any sequence $w_0, w_1, \ldots$, we let $\mathbf{w}_N = (w_0, w_1, \ldots, w_{N-1})^{\mathrm{T}}$ and

$$S_N = \begin{bmatrix} 0 & & \cdots & & 0 \\ 1 & & 0 & \cdots & 0 \\ 0 & & & & \vdots \\ \vdots & \ddots & & \ddots & \vdots \\ 0 & \cdots & 0 & 1 & 0 \end{bmatrix},$$

and we rewrite (1) as

$$\begin{aligned} A_N(\boldsymbol{\theta}^0)\, \mathbf{x}_N &= B_N(\boldsymbol{\theta}^0)\, \mathbf{u}_N + D_N \boldsymbol{\xi}_N, \\ \mathbf{y}_N &= \mathbf{x}_N + \boldsymbol{\eta}_N, \end{aligned} \tag{2}$$

where

$$A_N(\boldsymbol{\theta}) = \sum_{j=0}^{n} a_j S_N^j, \qquad a_0 = 1,$$

$$B_N(\boldsymbol{\theta}) = \sum_{j=1}^{n} b_j S_N^j,$$

$$D_N = \sum_{j=1}^{n} d_j^0 S_N^j.$$

Stochastic Assumptions

ASSUMPTION R. Throughout the sequel $\{\xi_k\}$ and $\{\eta_k\}$ will each denote sequences of independent, identically distributed random variables with zero means and finite variances, $\sigma_1^2 = \mathrm{Var}[\xi_0]$ and $\sigma_2^2 = \mathrm{Var}[\eta_0]$, respectively. The state variables and output variables then will be sample functions of stochastic processes and will be expressed as functions of ω for ω belonging to a probability space Ω.

The least squares estimate for $\boldsymbol{\theta}^0$ will be denoted by $\hat{\boldsymbol{\theta}}_N(\omega)$. For a given input vector $\mathbf{u}_N$ and output vector $\mathbf{y}_N(\omega)$, $\hat{\boldsymbol{\theta}}_N(\omega)$ is chosen so as to minimize

$$F(N, \boldsymbol{\theta}, \omega) = (1/N)\, \| \mathbf{y}_N(\omega) - A_N(\boldsymbol{\theta})^{-1} B_N(\boldsymbol{\theta})\, \mathbf{u}_N \|^2. \tag{3}$$

From (2) it follows that

$$\begin{aligned} F(N, \boldsymbol{\theta}, \omega) = (1/N)\, \| &(A_N(\boldsymbol{\theta}^0)^{-1} B_N(\boldsymbol{\theta}^0) - A_N(\boldsymbol{\theta})^{-1} B_N(\boldsymbol{\theta}))\, \mathbf{u}_N \\ &+ A_N(\boldsymbol{\theta}^0)^{-1} D_N \boldsymbol{\xi}_N(\omega) + \boldsymbol{\eta}_N(\omega) \|^2. \end{aligned} \tag{4}$$

For notational convenience we let

$$Q_N(\boldsymbol{\theta}) = A_N(\boldsymbol{\theta}^0)^{-1} B_N(\boldsymbol{\theta}^0) - A_N(\boldsymbol{\theta})^{-1} B_N(\boldsymbol{\theta}),$$

and

$$P_N = A_N(\boldsymbol{\theta}^0)^{-1} D_N ;$$

note that the matrices $Q_N(\boldsymbol{\theta})$ and P_N are of the form

$$Q_N(\boldsymbol{\theta}) = \sum_{j=0}^{N-1} q_j(\boldsymbol{\theta}) S_N{}^j,$$

$$P_N = \sum_{j=0}^{N-1} p_j S_N{}^j.$$

Then expanding (4), we have

$$\begin{aligned} F(N, \boldsymbol{\theta}, \omega) = {} & (1/N) \| Q_N(\boldsymbol{\theta}) \mathbf{u}_N \|^2 \\ & + (1/N) \| P_N \boldsymbol{\xi}_N(\omega) \|^2 + (1/N) \| \boldsymbol{\eta}_N(\omega) \|^2 \\ & + (2/N) \langle Q_N(\boldsymbol{\theta}) \mathbf{u}_N , P_N \boldsymbol{\xi}_N(\omega) \rangle \\ & + (2/N) \langle Q_N(\boldsymbol{\theta}) \mathbf{u}_N , \boldsymbol{\eta}_N(\omega) \rangle \\ & + (2/N) \langle P_N \boldsymbol{\xi}_N(\omega), \boldsymbol{\eta}_N(\omega) \rangle, \end{aligned} \tag{5}$$

where we are using the inner product notation $\langle \mathbf{v}, \mathbf{w} \rangle = \mathbf{v}^{\mathrm{T}} \mathbf{w}$.

System Assumptions

Assumption S. The set $\boldsymbol{\Theta}$ of admissible parameters $\boldsymbol{\theta}$ is a compact subset of R^{2n} which contains the true system parameter $\boldsymbol{\theta}^0$ in its interior and for each $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, with $\boldsymbol{\theta} = (a_1, ..., a_n, b_1, ..., b_n)^{\mathrm{T}}$,

$$\begin{aligned} &\text{(i)} \quad A(z, \boldsymbol{\theta}) = 1 + \sum_{j=1}^{n} a_j z^j \text{ has no zeros in } |z| \leqslant 1, \\ &\text{(ii)} \quad B(z, \boldsymbol{\theta}) = \sum_{j=1}^{n} b_j z^j \text{ has no zeros in common with } A(\cdot, \boldsymbol{\theta}), \end{aligned} \tag{6}$$

and not all $b_j = 0$. These assumptions assure that the system is asymptotically stable and completely controllable.

It follows from (6) that $A(\cdot, \boldsymbol{\theta})^{-1}$, the reciprocal of $A(\cdot, \boldsymbol{\theta})$, is analytic on $|z| \leqslant 1$ and has a series representation

$$A(z, \boldsymbol{\theta})^{-1} = \sum_{j=0}^{\infty} g_j(\boldsymbol{\theta}) z^j \qquad \text{valid on} \quad |z| \leqslant 1.$$

Also,

$$A_N(\boldsymbol{\theta})^{-1} = A(S_N, \boldsymbol{\theta})^{-1} = \sum_{j=0}^{\infty} g_j(\boldsymbol{\theta})\, S_N{}^j.$$

The matrix S_N is nilpotent of order N; hence $A_N(\boldsymbol{\theta})^{-1} = \sum_{j=0}^{N-1} g_j(\boldsymbol{\theta})\, S_N{}^j$. Similarly it follows that

$$Q_N(\boldsymbol{\theta}) = \sum_{j=0}^{N-1} q_j(\boldsymbol{\theta})\, S_N{}^j, \qquad \text{where} \quad \sum_{j=0}^{\infty} | q_j(\boldsymbol{\theta})| < \infty,$$

and

$$P_N = \sum_{j=0}^{N-1} p_j S_N{}^j, \qquad \text{where} \quad \sum_{j=0}^{\infty} | p_j | < \infty.$$

Input Assumptions

ASSUMPTION I. The sequence of inputs $\{u_j\}$ will be assumed to satisfy the following conditions:

(i) $\{u_j\}$ is a bounded sequence,

(ii) $\lim_{N\to\infty}(1/N)\sum_{k=\min\{i,j\}}^{N-1} u_{k-i}u_{k-j} = \tilde{u}(i,j)$ exists for every integer $i \geqslant 0$ and $j \geqslant 0$,

(iii) $\lim_{N\to\infty}(1/N)\, U_{N,2n}^T U_{N,2n} = \tilde{U}$ is positive definite, where

$$U_{N,2n} = (S_N \mathbf{u}_N\,, S_N{}^2 \mathbf{u}_N\,,\ldots, S_N^{2n} \mathbf{u}_N).$$

In [2], a sequence of inputs satisfying conditions (i) to (iii) above was said to be "persistently exciting."

If X_N denotes a random variable for each $N = 1, 2,\ldots$, we write "$X_N \to^d X$" if $\{X_N\}$ converges in distribution to the random variable X. Also we write "$\mathbf{X}_N \to^d N(\mathbf{O}, M)$" if the sequence of random vectors $\{\mathbf{X}_N\}$ converges in distribution to the normal distribution with mean $\mathbf{O}$ and covariance matrix M.

3. CONSISTENCY OF THE ESTIMATES

In order to show that $\hat{\boldsymbol{\theta}}_N \to \boldsymbol{\theta}^0$ almost surely, we first establish that $F(N, \boldsymbol{\theta}, \cdot) \to J(\boldsymbol{\theta})$ almost surely as $N \to \infty$ for every $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, where

$$\begin{aligned} J(\boldsymbol{\theta}) &= \lim_{N\to\infty} E[F(N, \boldsymbol{\theta}, \cdot)] \\ &= \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} q_i(\boldsymbol{\theta})\, q_j(\boldsymbol{\theta})\, \tilde{u}(i, j) + \sigma_1{}^2 \sum_{j=0}^{\infty} p_j{}^2 + \sigma_2{}^2. \end{aligned}$$

For this the next three lemmas are useful. The proof of Lemma 1 can be found in [10; Lemma 1] and will not be repeated here.

LEMMA 1. *For any* θ^* *satisfying Assumption* S *and* $\epsilon > 0$, *there is an integer* N^* *and* $\delta > 0$ *such that*

$$\text{(i)} \quad \sum_{j=0}^{\infty} | q_j(\theta)| \leqslant \sum_{j=0}^{\infty} | q_j(\theta^*)| + 1$$

and

$$\text{(ii)} \quad \sum_{j=N^*+1}^{\infty} | q_j(\theta)| \leqslant \epsilon$$

for all $\theta \in \Theta$ *for which* $\| \theta - \theta^* \| < \delta$.

COROLLARY TO LEMMA 1. *There is a constant M such that*

(i) $\sum_{j=0}^{\infty} | q_j(\theta)| \leqslant M$ *for all* $\theta \in \Theta$ *and for every* $\epsilon > 0$, *there is an integer* N^* *such that*

(ii) $\sum_{j=N^*+1}^{\infty} | q_j(\theta)| \leqslant \epsilon$ *for all* $\theta \in \Theta$.

The proof of the corollary follows by using Lemma 1 together with a standard argument involving the compactness of Θ.

The next two lemmas are concerned with the convergence of sequences such as those which appear as terms in Eq. (5). Therefore, the functions F and G which appear in the statement of the lemmas will be analytic functions whose coefficients are $\{q_k(\theta)\}$ or $\{p_k\}$. In order to avoid proving several lemmas, all of which have nearly identical proofs, we leave it to the reader to choose the appropriate definitions of F and G in each application of the lemmas.

LEMMA 2. *Let G and F denote analytic functions in z on* $\{z: | z | \leqslant 1\} \times \Theta$. *Let* $F(z, \theta) = \sum_{k=0}^{\infty} f_k(\theta) z^k$ *and* $G(z, \theta) = \sum_{k=0}^{\infty} g_k(\theta) z^k$ *denote their respective power series representations where the coefficient sequences* $\{f_k(\theta)\}$ *and* $\{g_k(\theta)\}$ *satisfy the conclusions of the corollary to Lemma* 1. *Let* $\{w_k\}$ *and* $\{v_k\}$ *denote sequences of real numbers such that*

(i) $\lim_{N\to\infty}(1/N) \sum_{k=0}^{N-1} w_k v_{k+j}$ *exists for each* $j = 0, 1, 2,\ldots$,

(ii) $\lim_{N\to\infty}(1/N) \sum_{k=0}^{N-1} w_{k+j} v_k$ *exists for each* $j = 0, 1, 2,\ldots$,

(iii) $(1/N) \sum_{k=0}^{N-1} w_k^2 \leqslant M$ *for all* $N = 1, 2,\ldots$, *and*

(iv) $(1/N) \sum_{k=\theta}^{N-1} v_k^2 \leqslant M$ *for all* $N = 1, 2,\ldots$.

If $G_N(\boldsymbol{\theta}) = \sum_{k=0}^{N-1} g_k(\boldsymbol{\theta}) S_N{}^k$ *and* $F_N(\boldsymbol{\theta}) = \sum_{k=0}^{N-1} f_k(\boldsymbol{\theta}) S_N{}^k$, *then for every* $\epsilon > 0$, *there is an* N_0 *such that for* $N \geqslant N_0$ *and for every* $\boldsymbol{\theta} \in \boldsymbol{\Theta}$,

$$\left| (1/N)\langle G_N(\boldsymbol{\theta})\, \mathbf{w}_N , F_N(\boldsymbol{\theta})\, \mathbf{v}_N\rangle - \sum_{k=0}^{\infty} \sum_{j=0}^{\infty} g_k(\boldsymbol{\theta}) f_k(\boldsymbol{\theta})(wv)_{k,j} \right| < \epsilon,$$

where $(wv)_{k,j} = \lim_{N\to\infty}\langle S_N{}^k \mathbf{w}_N , S_N{}^j \mathbf{v}_N\rangle/N$.

Proof. Let M denote a constant chosen large enough to satisfy inequalities (iii) and (iv) in the hypothesis and such that for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, $\sum_{k=0}^{\infty} | f_k(\boldsymbol{\theta})| \leqslant M$ and $\sum_{k=0}^{\infty} | g_k(\boldsymbol{\theta})| \leqslant M$. Then we note that

$$\begin{aligned} |(1/N)\langle S_N{}^k \mathbf{w}_N , S_N{}^j \mathbf{v}_N\rangle| &\leqslant \|(1/N^{1/2})\, S_N{}^k \mathbf{w}_N\| \cdot \|(1/N^{1/2})\, S_N{}^j \mathbf{v}_N\| \\ &\leqslant \left[(1/N) \sum_{i=0}^{N-1} w_i{}^2\right]^{1/2} \cdot \left[(1/N) \sum_{i=0}^{N-1} v_i{}^2\right]^{1/2} \leqslant M \end{aligned} \tag{7}$$

for all $N = 1, 2,\ldots$ and all $k, j = 0, 1, 2,\ldots$. From (7) and the fact that the power series for G and F converge absolutely at $z = 1$, it follows that the double series in the conclusion of the lemma is absolutely convergent.

Now let N^* be a fixed positive integer and suppose that $N > N^*$. Then

$$\begin{aligned} &\left| \sum_{k=0}^{\infty} \sum_{j=0}^{\infty} g_k(\boldsymbol{\theta}) f_j(\boldsymbol{\theta})(wv)_{k,j} - (1/N) \sum_{k=0}^{N-1} \sum_{j=0}^{N-1} g_k(\boldsymbol{\theta}) f_j(\boldsymbol{\theta})\langle S_N{}^k \mathbf{w}_N , S_N{}^j \mathbf{v}_N\rangle \right| \\ &\quad \leqslant \left| \sum_{k=0}^{N^*} \sum_{j=0}^{N^*} g_k(\boldsymbol{\theta}) f_j(\boldsymbol{\theta})[(wv)_{k,j} - (1/N)\langle S_N{}^k \mathbf{w}_N , S_N{}^j \mathbf{v}_N\rangle] \right| \\ &\quad + 2M \sum_{k=N^*+1}^{\infty} \sum_{j=0}^{\infty} | g_k(\boldsymbol{\theta})|\, | f_j(\boldsymbol{\theta})| + 2M \sum_{j=N^*+1}^{\infty} \sum_{k=0}^{\infty} | g_k(\boldsymbol{\theta})|\, | f_j(\boldsymbol{\theta})|. \end{aligned} \tag{8}$$

The last expression on the right side of inequality (8) is less than or equal to $2M^2 \cdot \sum_{j=N^*+1}^{\infty} | f_j(\boldsymbol{\theta})|$ for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$. By hypothesis this quantity can be made uniformly small for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ by choosing N^* sufficiently large. Similarly the second expression on the right side of (8) can be made uniformly small for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$. For fixed N^*, the first expression on the right side of (8) can be made uniformly small for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ by choosing $N_0 \geqslant N^*$ sufficiently large because of (i) and (ii) in the hypothesis. Q.E.D.

LEMMA 3. *Suppose that* G *and* G_N *are as in Lemma* 2 *and that* $\{w_k\}$ *and* $\{v_k\}$ *are sequences of real numbers such that*

(i) $\lim_{N\to\infty}(1/N) \sum_{j=0}^{N-1-k} w_j v_{j+k} = 0$ *for each* $k = 0, 1, 2,\ldots$,
(ii) $(1/N) \sum_{j=0}^{N-1} w_j{}^2 \leqslant M$ *for each* $N = 1, 2,\ldots$, *and*
(iii) $(1/N) \sum_{j=0}^{N-1} v_j{}^2 \leqslant M$ *for each* $N = 1, 2,\ldots$.

Then for every $\epsilon > 0$, *there is an integer* N_0 *such that* $N \geqslant N_0$ *implies that*

$$|(1/N)\langle G_N(\boldsymbol{\theta})\, \mathbf{w}_N\,, \mathbf{v}_N\rangle| < \epsilon \qquad \textit{for all} \quad \boldsymbol{\theta} \in \boldsymbol{\Theta}.$$

Proof. Let M be a constant chosen large enough to satisfy (i) and (ii) in the hypothesis and such that $\sum_{k=0}^{\infty} | g_k(\boldsymbol{\theta})| \leqslant M$ for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$. Let N^* be a fixed positive integer and suppose that $N > N^* + 1$. Then

$$\begin{aligned}|(1/N)\langle G_N(\boldsymbol{\theta})\, \mathbf{w}_N\,, \mathbf{v}_N\rangle| &= \left| (1/N) \sum_{k=0}^{N-1} g_k(\boldsymbol{\theta})\langle S_N{}^k \mathbf{w}_N\,, \mathbf{v}_N\rangle \right| \\ &\leqslant \sum_{k=0}^{N-1} | g_k(\boldsymbol{\theta})| \cdot \left| (1/N) \sum_{j=0}^{N-1-k} w_j v_{j+k} \right| \qquad (9)\\ &\leqslant \sum_{k=0}^{N^*} | g_k(\boldsymbol{\theta})| \cdot \left| (1/N) \sum_{j=0}^{N-1-k} w_j v_{j+k} \right| + \sum_{k=N^*+1}^{N-1} | g_k(\boldsymbol{\theta})| \cdot M\end{aligned}$$

(the fact that $|(1/N)\sum_{j=0}^{N-1-k} w_j v_{j+k}| \leqslant M$ follows from (i) and (ii) in the hypothesis and the Schwarz inequality). The second expression on the right side of inequality (9) can be made uniformly small for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ by choosing N^* sufficiently large (by hypothesis, $\sum_{k=N^*+1}^{\infty} | g_k(\boldsymbol{\theta})|$ can be made uniformly small for all $\theta \in \Theta$). The first expression on the right side of (9) can be made uniformly small for all $\boldsymbol{\theta} \in \Theta$ by choosing $N_0 > N^*$ sufficiently large that each of the finitely many terms $|(1/N)\sum_{j=0}^{N-1-k} w_j v_{j+k}|$, $k = 0, 1,..., N^*$, is smaller than $\epsilon/2M$ for $N \geqslant N_0$ (because of hypothesis (i) and the fact that $\sum_{k=0}^{\infty} | g_k(\boldsymbol{\theta})| \leqslant M$ for all $\boldsymbol{\theta} \in \Theta$). Q.E.D.

THEOREM 1. *Suppose Assumptions* I, R, *and* S *hold. There is an event* $\Omega^0 \subseteq \Omega$ *with* $P(\Omega^0) = 1$ *such that for every* $\omega \in \Omega^0$ *and* $\epsilon > 0$, *there is an integer* $N_0 = N_0(\omega, \epsilon)$ *such that* $N \geqslant N_0$ *implies that*

$$| F(N, \boldsymbol{\theta}, \omega) - J(\boldsymbol{\theta})| < \epsilon$$

for all $\boldsymbol{\theta} \in \Theta$, *where*

$$J(\boldsymbol{\theta}) = \lim_{N\to\infty} E[F(N, \boldsymbol{\theta}, \cdot)].$$

Proof. Let Ω^0 be the set of all ω for which all of the following are true:

(i) $(1/N)\sum_{k=0}^{N-1} \xi_k{}^2(\omega) \to \sigma_1{}^2$,

(ii) $(1/N)\sum_{k=0}^{N-1} \eta_k{}^2(\omega) \to \sigma_2{}^2$,

(iii) $(1/N)\sum_{k=0}^{N-1} \xi_k(\omega)\, \xi_{k+j}(\omega) \to 0$ for $j = 1, 2,...$,

(iv) $(1/N)\sum_{k=0}^{N-1} \xi_k(\omega)\, \eta_{k+j}(\omega) \to 0$ for $j = 0, 1, 2,...$,

(v) $(1/N)\sum_{k=0}^{N-1} u_k \xi_{k+j}(\omega) \to 0$ for $j = 0, 1, 2,...$,

(vi) $(1/N)\sum_{k=0}^{N-1} u_{k+j} \xi_k(\omega) \to 0$ for $j = 0, 1, 2,...$,

(vii) $(1/N)\sum_{k=0}^{N-1} u_k \eta_{k+j}(\omega) \to 0$ for $j = 0, 1, 2,...$.

Kolmogorov's strong law of large numbers [8, p. 241] implies that each of the sequences in (i) to (iii), and (v) to (vii) converges for almost all ω. The sequence of random variables $\{\xi_k \eta_{k+j}\}$ is j-dependent, and so the strong law of large numbers holds (see [4, Theorem 3]). It follows that Ω^0 is a countable intersection of amost sure events, hence Ω^0 is an almost sure event.

Let $\omega \in \Omega^0$. Then

$$\begin{aligned} F(N, \boldsymbol{\theta}, \omega) = {} & (1/N)\langle Q_N(\boldsymbol{\theta})\,\mathbf{u}_N, Q_N(\boldsymbol{\theta})\,\mathbf{u}_N\rangle + (1/N)\langle P_N \boldsymbol{\xi}_N(\omega), P_N \boldsymbol{\xi}_N(\omega)\rangle \\ & + (2/N)\langle Q_N(\boldsymbol{\theta})\,\mathbf{u}_N, P_N \boldsymbol{\xi}_N(\omega)\rangle + (2/N)\langle Q_N(\boldsymbol{\theta})\,u_N, \boldsymbol{\eta}_N(\omega)\rangle \\ & + (2/N)\langle P_N \boldsymbol{\xi}_N(\omega), \boldsymbol{\eta}_N(\omega)\rangle. \end{aligned} \tag{10}$$

Lemma 2 yields the uniform convergence of the first, second, and fourth terms in (10), Lemma 3 implies that the fifth and sixth terms converge uniformly in $\boldsymbol{\theta}$, and condition (ii) above yields the convergence of the third term. Q.E.D.

THEOREM 2. *Suppose Assumptions* I, R, *and* S *hold. Then for almost all* ω, $\lim_{N\to\infty} \hat{\boldsymbol{\theta}}_N(\omega) = \boldsymbol{\theta}^0$.

Proof. Let $\omega \in \Omega^0$, where Ω^0 is the set of probability one defined in the proof of Theorem 1. The parameter set is compact, so $\{\hat{\boldsymbol{\theta}}_N(\omega)\}$ has a convergent subsequence, say $\lim_{j\to\infty} \hat{\boldsymbol{\theta}}_{N_j}(\omega) = \boldsymbol{\theta}^*$. Suppose $\boldsymbol{\theta}^* \neq \boldsymbol{\theta}^0$. Then $J(\boldsymbol{\theta}^*) > J(\boldsymbol{\theta}^0)$ because J has a unique minimum (see [1, proof of Theorem 1]). Let $\epsilon = J(\boldsymbol{\theta}^*) - J(\boldsymbol{\theta}^0)$. Now, $F(N_j, \hat{\boldsymbol{\theta}}_{N_j}(\omega), \omega) \leqslant F(N_j, \boldsymbol{\theta}^0, \omega)$ for each j; hence $\lim_{j\to\infty} \sup F(N_j, \hat{\boldsymbol{\theta}}_{N_j}(\omega), \omega) \leqslant J(\boldsymbol{\theta}^0)$. For j sufficiently large it follows that $F(N_j, \hat{\boldsymbol{\theta}}_{N_j}(\omega), \omega) \leqslant J(\boldsymbol{\theta}^0) + (\epsilon/4)$. By Theorem 1 for fixed $\omega \in \Omega^0$, $\lim_{N\to\infty} F(N, \boldsymbol{\theta}, \omega) = J(\boldsymbol{\theta})$ and the convergence is uniform in $\boldsymbol{\theta}$. Since $F(N, \cdot, \omega)$ is continuous on $\boldsymbol{\Theta}$, it follows that $J(\cdot)$ is continuous on $\boldsymbol{\Theta}$. By the continuity of $J(\cdot)$ and the uniform convergence of $F(N, \cdot, \omega)$, it follows that for j sufficiently large,

$$\begin{aligned} & | J(\boldsymbol{\theta}^*) - F(N_j, \hat{\boldsymbol{\theta}}_{N_j}(\omega), \omega)| \\ & \qquad \leqslant | J(\boldsymbol{\theta}^*) - J(\hat{\boldsymbol{\theta}}_{N_j}(\omega))| + | J(\hat{\boldsymbol{\theta}}_{N_j}(\omega)) - F(N_j, \hat{\boldsymbol{\theta}}_{N_j}(\omega), \omega)| < (\epsilon/8) + (\epsilon/8). \end{aligned}$$

Then $J(\boldsymbol{\theta}^*) < F(N_j, \hat{\boldsymbol{\theta}}_{N_j}(\omega), \omega) + (\epsilon/4) < J(\boldsymbol{\theta}^0) + (\epsilon/2)$, from which it follows that $J(\boldsymbol{\theta}^*) - J(\boldsymbol{\theta}^0) < \epsilon/2$, contrary to the choice of ϵ.

4. ASYMPTOTIC NORMALITY OF THE ESTIMATES

Let $\boldsymbol{\phi}_N(\boldsymbol{\theta}) = \nabla_{\boldsymbol{\theta}} F(N, \boldsymbol{\theta}, \omega)$. In showing that $\hat{\boldsymbol{\theta}}_N$ is asymptotically normal, we will use the equation

$$0 = \boldsymbol{\phi}_N(\hat{\boldsymbol{\theta}}_N) = \boldsymbol{\phi}_N(\boldsymbol{\theta}^0) + (\nabla_{\boldsymbol{\theta}} \boldsymbol{\phi}_N(\boldsymbol{\theta}_N^*)) \cdot (\hat{\boldsymbol{\theta}}_N - \boldsymbol{\theta}^0), \tag{11}$$

for some θ_N^* satisfying $\| \theta_N^* - \theta^0 \| \leqslant \| \hat{\theta}_N - \theta^0 \|$. We will show that $N^{1/2}\phi_N(\theta^0)$ converges in distribution to a normal distribution and that $\nabla_\theta \phi_N(\theta_N{}^0)$ converges almost surely to a degenerate random matrix which is invertible.

In this section we assume, in addition to the previous hypotheses, that the sequences of random variables $\{\xi_k\}$ and $\{\eta_k\}$ have finite third moments.

LEMMA 3. *Let α_{mk} and β_{mk}, for $k = 0, 1,..., m-1$ and $m = 1, 2,...$, denote uniformly bounded real numbers. If*

$$S_m = \sum_{k=0}^{m-1} (\alpha_{mk}\xi_k + \beta_{mk}\eta_k)$$

and

$$\lim_{m\to\infty} \left[\sum_{k=0}^{m} \beta_{mk}^2 + \sum_{k=0}^{m} \alpha_{mk}^2 \right] = \infty,$$

Then $S_m/[\mathrm{var}(S_m)]^{1/2}$ converges in distribution to the standard normal distribution.

Proof. For the proof it suffices to show that

$$\lim_{m\to\infty} \frac{\sum_{k=0}^{m-1} E(|\alpha_{mk}\xi_k|)^3 + \sum_{k=0}^{m-1} E(|\beta_{mk}\eta_k|)^3}{[\mathrm{var}(S_m)]^{3/2}} = 0$$

and then apply the Basic Lemma in [8, p. 277].

Let us consider the case where both

$$\lim_{m\to\infty} \sum_{k=0}^{m-1} \alpha_{mk}^2 = \infty \qquad \text{and} \qquad \lim_{m\to\infty} \sum_{k=0}^{m-1} \beta_{mk}^2 = \infty.$$

The proofs in the other cases are analogous. Let M denote a bound for the α_{mk}'s and β_{mk}'s. We have

$$\frac{\sum_{k=0}^{m-1} E(|\alpha_{mk}\xi_k|)^3 + \sum_{k=0}^{m-1} E(|\beta_{mk}\eta_k|)^3}{[\mathrm{var}(S_m)]^{3/2}}$$

$$= \frac{E(|\xi_0|^3)\sum_{k=0}^{m-1} |\alpha_{mk}|^3 + E(|\eta_0|^3)\sum_{k=0}^{m-1} |\beta_{mk}|^3}{[\sigma_1{}^2 \sum_{k=0}^{m-1} \alpha_{mk}^2 + \sigma_2{}^2 \sum_{k=0}^{m-1} \beta_{mk}^2]^{3/2}}$$

$$\leqslant \frac{E(|\xi_0|^3)\sum_{k=0}^{m-1} |\alpha_{mk}|^3}{[\sigma_1{}^2 \sum_{k=0}^{m-1} \alpha_{mk}^2]^{3/2}} + \frac{E(|\eta_0|^3)\sum_{k=0}^{m-1} |\beta_{mk}|^3}{[\sigma_2{}^2 \sum_{k=0}^{m-1} \beta_{mk}^2]^{3/2}}$$

$$= \frac{E(|\xi_0|^3)\sum_{k=0}^{m-1} |\alpha_{mk}/M|^3}{[\sigma_1{}^2 \sum_{k=0}^{m-1} (\alpha_{mk}/M)^2]^{3/2}} + \frac{E(|\eta_0|^3)\sum_{k=0}^{m-1} |\beta_{mk}/M|^3}{[\sigma_2{}^2 \sum_{k=0}^{m-1} (\beta_{mk}/M)^2]^{3/2}}$$

$$\leqslant \frac{E(|\xi_0|)^3 M}{\sigma_1{}^3(\sum_{k=0}^{m-1} \alpha_{mk}^2)^{1/2}} + \frac{E(|\eta_0|^3)\cdot M}{\sigma_2{}^3(\sum_{k=0}^{m-1} \beta_{mk}^2)^{1/2}}.$$

Now it is easily seen that both of the expressions on the right side of the inequality go to zero as $m \to \infty$. Q.E.D.

Let $\phi_{N,i}(\theta^0)$ denote the ith component of $\boldsymbol{\phi}_N(\theta^0)$. Using matrix calculus, we find that for $i = 1, 2, \dots, n$,

$$\phi_{N,i}(\theta^0) = (2/N)\langle P_N \xi_N + \eta_N , (A_N{}^0)^{-2} B_N{}^0 S_N{}^i \mathbf{u}_N \rangle$$

and

$$\phi_{N,i+n}(\theta^0) = -(2/N)\langle P_N \xi_N + \eta_N , (A_N{}^0)^{-1} S_N{}^i \mathbf{u}_N \rangle.$$

LEMMA 4. *For each* $i = 1, 2, \dots, 2n$, $\phi_{N,i}(\theta^0)$ *is asymptotically normal.*

Proof. We will give the proof only for $i = 1, 2, \dots, n$ since the proof for $i = n + 1, \dots, 2n$ is analogous.

Consider

$$\langle P_N \xi_N + \eta_N , (A_N{}^0)^{-2} B_N{}^0 S_N{}^i \mathbf{u}_N \rangle = \sum_{k=0}^{N-1} \langle S_N{}^k \mathbf{p}_N , (A_N{}^0)^{-2} B_N{}^0 S_N{}^i \mathbf{u}_N \rangle \cdot \xi_k$$

$$+ \sum_{k=0}^{N-1} \langle \mathbf{e}_{N,k} , (A_N{}^0)^{-2} B_N{}^0 S_N{}^i \mathbf{u}_N \rangle \, \eta_k ,$$

where $\mathbf{p}_N = (p_0 , p_1 , \dots, p_{N-1})^{\mathrm{T}}$ and $e_{N,k} = (\delta_{0,k} , \dots, \delta_{N-1,k})^{\mathrm{T}}$ with $\delta_{k,k} = 1$ and $\delta_{j,k} = 0$ for $j = k$. We now apply Lemma 3 with $\alpha_{N,k} = \langle S_N{}^k \mathbf{p}_N , (A_N{}^0)^{-2} B_N{}^0 S_N{}^i \mathbf{u}_N \rangle$ and $\beta_{N,k} = \langle \mathbf{e}_{N,k} , (A_N{}^0)^{-2} B_N{}^0 S_N{}^i \mathbf{u}_N \rangle$. We only need to show that $\sum_{k=0}^{N-1} \beta_{N,k}^2 \to \infty$ as $N \to \infty$. But

$$\sum_{k=0}^{N-1} \beta_{N,k}^2 = \| (A_N{}^0)^{-2} B_N{}^0 S_N{}^i \mathbf{u}_N \|^2$$

$$= \| B_N{}^0 S_N{}^i \mathbf{u}_N \|_0^2 ,$$

where $\| \cdot \|_0$ denotes the norm defined by the positive definite symmetric matrix $((A_N{}^0)^2 (A_N{}^0)^{2^T})^{-1}$. Since there are positive mumbers ρ_1 , ρ_2 such that $\rho_1 I_N \leqslant (A_N{}^0)^2 (A_N{}^0)^{2^T} \leqslant \rho_2 I_N$, where I_N is the $N \times N$ identity matrix, it follows that $\| B_N{}^0 S_N{}^i \mathbf{u}_N \|^2 \to \infty$ as $N \to \infty$ if and only if $\| B_N{}^0 S_N{}^i \mathbf{u}_N \|^2 \to \infty$ as $N \to \infty$. But $\lim_{N\to\infty} (1/N) \cdot \| B_N{}^0 S_N{}^i \mathbf{u}_N \|^2$ exists and is positive since

$$(1/N) \| B_N{}^0 S_N{}^i \mathbf{u}_N \|^2 = (1/N) \left\langle \sum_{k=1}^{n} b_k S_N{}^k S_N{}^i \mathbf{u}_N , \sum_{j=1}^{n} b_j S_N{}^j S_N{}^i \mathbf{u}_N \right\rangle$$

$$= \sum_{k=1}^{n} \sum_{j=1}^{n} b_k b_j (1/N) \langle S_N^{k+i} \mathbf{u}_N , S_N^{j+i} \mathbf{u}_N \rangle,$$

which converges to $\sum_{k=1}^{n} \sum_{k=1}^{n} b_k b_j \tilde{u}_{k+i,j+i}$ as $N \to \infty$ and the matrix $[\tilde{u}_{k+i,j+i}]_{k,j=1}^{n}$

is a positive definite submatrix of the positive definite matrix $\tilde{U}$. It now follows that $\| B_{N0} S_N{}^i \mathbf{u}_N \| \to \infty$ and consequently $\| B_N{}^0 S_N{}^i \mathbf{u}_N \|^2 \to \infty$ as $N \to \infty$. Q.E.D.

LEMMA 5. *$N^{1/2} \boldsymbol{\phi}_N(\boldsymbol{\theta}^0)$ is asymptotically normal, the covariance matrix of the limit distribution is $M_0 = \lim_{N\to\infty}(N \cdot M_N)$ where M_N is the covariance matrix of $\boldsymbol{\phi}_N(\boldsymbol{\theta}^0)$, and $M_0 > 0$.*

Proof. First we show that $\lim_{N\to\infty}(N \cdot M_N)$ exists. Let $M_{N,i,j}$ denote the (i,j)th entry of M_N. For $i, j = 1, 2,..., n$,

$$M_{N,i,j} = (4\sigma_1{}^2/N^2) \cdot \langle P_N{}^{\mathrm{T}}(A_N{}^0)^{-2} B_N{}^0 S_N{}^i \mathbf{u}_N , P_N{}^{\mathrm{T}}(A_N{}^0)^{-2} B_N{}^0 S_N{}^j \mathbf{u}_N \rangle + (4\sigma_2{}^2/N^2) \cdot \langle (A_N{}^0)^{-2} B_N{}^0 S_N{}^i \mathbf{u}_N , (A_N{}^0)^{-2} B_N{}^0 S_N{}^j \mathbf{u}_N \rangle;$$

$$M_{N,i,j+n} = (-4\sigma_1{}^2/N^2) \cdot \langle P_N{}^{\mathrm{T}}(A_N{}^0)^{-2} B_N{}^0 S_N{}^i \mathbf{u}_N , P_N{}^{\mathrm{T}}(A_N{}^0)^{-1} S_N{}^j \mathbf{u}_N \rangle + (-4\sigma_2{}^2/N^2) \cdot \langle (A_N{}^0)^{-2} B_N{}^0 S_N{}^i \mathbf{u}_N , (A_N{}^0)^{-1} S_N{}^j \mathbf{u}_N \rangle;$$

and

$$M_{N,i+n,j+n} = (4\sigma_1{}^2/N^2) \cdot \langle P_N{}^{\mathrm{T}}(A_N{}^0)^{-1} S_N{}^i \mathbf{u}_N , P_N{}^{\mathrm{T}}(A_N{}^0)^{-1} S_N{}^j \mathbf{u}_N \rangle + (4\sigma_2{}^2/N^2) \cdot \langle (A_N{}^0)^{-1} S_N{}^i \mathbf{u}_N , (A_N{}^0)^{-1} S_N{}^j \mathbf{u}_N \rangle.$$

If we let $H_N = [S_N \mathbf{x}_N , S_N{}^2 \mathbf{x}_N ,..., S_N{}^n \mathbf{x}_N , -S_N \mathbf{u}_N ,..., -S_N{}^n \mathbf{u}_N]$, where $\mathbf{x}_N = (A_N{}^0)^{-1} B_N{}^0 \mathbf{u}_N$, then we can express $N \cdot M_N$ as

$$N \cdot M_N = (4\sigma_1{}^2/N) H_N{}^{\mathrm{T}} K_N H_N + (4\sigma_2{}^2/N) H_N{}^{\mathrm{T}} L_N H_N ,$$

where $K_N = (A_N{}^0)^{-1\mathrm{T}} P_N P_N{}^{\mathrm{T}}(A_N{}^0)^{-1}$ and $L_N = (A_N{}^0)^{-1\mathrm{T}} (A_N{}^0)^{-1}$. It was shown in [1] that the second matrix in the sum above has a limit that is positive definite. The first matrix is clearly nonnegative definite, and so we only need to show that the limit of the first matrix exists.

Let us consider the (i,j)th entry for $i, j = 1, 2,..., n$:

$$(4\sigma_1{}^2/N)\langle P_N{}^{\mathrm{T}}(A_N{}^0)^{-2} B_N{}^0 S_N{}^i \mathbf{u}_N , P_N{}^{\mathrm{T}}(A_N{}^0)^{-2} B_N{}^0 S_N{}^j \mathbf{u}_N \rangle$$

$$= (4\sigma_1{}^2/N) \sum_{k=0}^{N-1} \sum_{m=0}^{N-1} p_k p_m \langle S_N^{k\mathrm{T}}(A_N{}^0)^{-2} B_N{}^0 S_N{}^i \mathbf{u}_N , S_N^{m\mathrm{T}}(A_N{}^0)^{-2} B_N{}^0 S_N{}^j \mathbf{u}_N \rangle. \tag{11}$$

By a technique similar to the proof of Lemma 1, it can be shown that the limit of (11) exists as $N \to \infty$ and likewise for the other entries of the matrix.

From Lemma 4 we know that for each $i = 1, 2,... 2n$,

$$\phi_{N,i}(\boldsymbol{\theta}^0)/V_{N,i}^{1/2} \xrightarrow{d} N(0, 1), \qquad \text{where} \quad V_{N,i} = \mathrm{Var}(\phi_{N,i}(\boldsymbol{\theta}^0)).$$

Then also $N^{1/2}\phi_{N,i}(\boldsymbol{\theta}^0)/(N \cdot V_{N,i})^{1/2} \to^d N(0, 1)$. Since $N \cdot V_{N,i} = \mathrm{Var}(N^{1/2}\phi_{N,i}(\boldsymbol{\theta}^0))$ and since $\lim_{N\to\infty} N \cdot V_{N,i} = M_{0,i,i}$, the (i,i)th entry of M_0, it follows that

$N^{1/2}\phi_{N,i}(\theta^0)$ converges in distribution to a normal distribution with zero mean and variance $M_{0,i,i}$.

Let $\gamma_1, \gamma_2, \ldots, \gamma_{2n}$ be arbitrary real numbers for which $\sum_{j=1}^{2n} \gamma_j^2 \neq 0$ and let $Y_N = N^{1/2}(\gamma_1 \phi_{N,1}(\theta^0) + \cdots + \gamma_{2n}\phi_{N,2n}(\theta^0))$. From Lemma 3 it follows that $Y_N/[\mathrm{Var}(Y_N)]^{1/2} \to^d N(0, 1)$. Also $\mathrm{Var}(Y_N) = N(\gamma^T M_N \gamma)$, where $\gamma = (\gamma_1, \ldots, \gamma_{2n})^T$ and $\lim_{N\to\infty} \mathrm{Var}(Y_N) = \gamma^T M_0 \gamma$. Hence $y_N \to^d N(0, \gamma^T M_0 \gamma)$. By the Cramer–Wold Lemma [3], it follows that $N^{1/2}\boldsymbol{\phi}_N(\theta^0) \to^d N(0, M_0)$. Q.E.D.

LEMMA 6. *Let* $R(\theta) = \lim_{N\to\infty} E(\nabla_\theta \boldsymbol{\phi}_N(\theta))$. *Then*

(a) $\lim_{N\to\infty} \nabla_\theta \boldsymbol{\phi}_N(\theta) = R(\theta)$ *almost surely, and*

(b) $\lim_{N\to\infty} \nabla_\theta \boldsymbol{\phi}_N(\theta_N^*) = R(\theta^0)$ *almost surely.*

Proof. Using matrix calculus, we find that the (i, j)th component of $\nabla_\theta \boldsymbol{\phi}_N(\theta)$ is as follows: for $i, j = 1, 2, \ldots, n$,

$$\partial^2 F/\partial a_i\, \partial a_j$$

$$= (2/N)\langle A_N^{-2} B_N S_N^{\,j} \mathbf{u}_N, A_N^{-2} B_N S_N^{\,i} \mathbf{u}_N\rangle$$

$$- (4/N)\langle P_N \boldsymbol{\xi}_N + \boldsymbol{\eta}_N + (A_N^0)^{-1} B_N^0 \mathbf{u}_N - A_N^{-1} B_N \mathbf{u}_N, A_N^{-3} B_N S_N^{i+j} \mathbf{u}_N\rangle;$$

for $i = 1, \ldots, n$ and $j = k + n$, $k = 1, \ldots, n$,

$$\partial^2 F/\partial a_i\, \partial b_k$$

$$= -(2/N)\langle A_N^{-2} B_N S_N^{\,i} \mathbf{u}_N, A_N^{-1} S_N^{\,k} \mathbf{u}_N\rangle$$

$$+ (2/N)\langle P_N \boldsymbol{\xi}_N + \boldsymbol{\eta}_N + (A_N^0)^{-1} B_N^0 \mathbf{u}_N - A_N^{-1} B_N \mathbf{u}_N, A_N^{-2} S_N^{i+k} \mathbf{u}_N\rangle;$$

and for $i, j = n + 1, \ldots, 2n$, $k = j - n$, $m = i - n$,

$$\partial^2 F/\partial b_k\, \partial b_m = (2/N)\langle A_N^{-1} S_N^{\,i} \mathbf{u}_N, A_N^{-1} S_N^{\,j} \mathbf{u}_N\rangle.$$

Part (a) is established by using Lemma 1 and Kolmogorov's sufficient conditions for the strong law of large numbers just as in the proof of Theorem 1.

To prove (b) we first will show that

for every $\epsilon > 0$ there are a $\delta > 0$ and an integer N_1 such that $\| \nabla_\theta \boldsymbol{\phi}_N(\theta) - R(\theta)\| < \epsilon$ for all θ such that $\| \theta - \theta^0 \| < \delta$ and $N \geqslant N_1$. (12)

A typical expression in the expansion of the $(i\ j)$th term of $\nabla_\theta \boldsymbol{\phi}_N(\theta)$ is of the form $(1/N)\langle G_N(\theta)\, \mathbf{v}_N, F_N(\theta)\, \mathbf{w}_N\rangle$, where $G_N(\theta) = G(S_N; \theta)$ and $F_N = F(S_N; \theta)$ for some functions $G(z; \theta) = \sum_{i=0}^{\infty} g_i(\theta)\, z^i$ and $F(z; \theta) = \sum_{i=0}^{\infty} f_i(\theta)\, z^i$ which are analytic on $|z| \leqslant 1$ and for which $G(\cdot; \theta) \to G(\cdot; \theta^0)$ and $F(\cdot; \theta) \to F(\cdot; \theta^0)$ uniformly on compact sets as $\theta \to \theta^0$. Also, the sequences

$\{v_N\}_{N=0}^{\infty}$ and $\{w_N\}_{N=0}^{\infty}$ will satisfy hypotheses (i)–(iv) of Lemma 1. Given $\epsilon > 0$ there are $\delta \to 0$ and an integer N_0 such that

$$\sum_{i=N_0+1}^{\infty} |f_i(\boldsymbol{\theta})| < \epsilon, \qquad \sum_{i=N_0+1}^{\infty} |g_i(\boldsymbol{\theta})| < \epsilon,$$

$$\sum_{i=0}^{\infty} |f_i(\boldsymbol{\theta})| \leqslant \sum_{i=0}^{\infty} |f_i(\boldsymbol{\theta}^0)| + 1, \quad \text{and} \quad \sum_{i=0}^{\infty} |g_i(\boldsymbol{\theta})| \leqslant \sum_{i=0}^{\infty} |g_i(\boldsymbol{\theta}^0)| + 1 \tag{13}$$

for all $\boldsymbol{\theta}$ satisfying $\| \boldsymbol{\theta} - \boldsymbol{\theta}^0 \| \leqslant \delta$. (For proof see [10, Lemma 1].)

We have

$$\Big| (1/N)\langle G_N(\boldsymbol{\theta})\, \mathbf{v}_N , F_N(\boldsymbol{\theta})\, \mathbf{w}_N\rangle - \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} g_i(\boldsymbol{\theta})\, f_j(\boldsymbol{\theta})(wv)_{i,j} \Big|$$

$$\leqslant \Big| \sum_{i=0}^{N_0}\sum_{j=0}^{N_0} g_i(\boldsymbol{\theta})\, f_j(\boldsymbol{\theta})[(1/N)\langle S_N{}^i \mathbf{v}_N , S_N{}^j \mathbf{w}_N\rangle - (wv)_{i,j}] \Big|$$

$$+ 2\sum_{i=0}^{\infty}\sum_{j=N_0+1}^{\infty} | g_i(\boldsymbol{\theta})\, f_j(\boldsymbol{\theta})| \cdot M + 2 \sum_{i=N_0+1}^{\infty}\sum_{j=0}^{\infty} | g_i(\boldsymbol{\theta})\, f_j(\boldsymbol{\theta})| \cdot M,$$

where M is a bound for all the sequences of averages as in Lemma 1. From (13) we see that $\delta > 0$ and N_0 can be found so that the last two terms are small. Having fixed N_0, we can choose $N_1 \geqslant N_0$ large enough that the first term is small. Thus we have (12).

Next we note that $R(\cdot)$ is continuous at $\boldsymbol{\theta}^0$. The (i, j)th term of $R(\boldsymbol{\theta})$ is a finite sum of expression of the form $\sum_{i=0}^{\infty}\sum_{i=0}^{\infty} g_i(\theta)\, f_j(\theta)(vw)_{i,j}$. It suffices to show that these vary continuously with $\boldsymbol{\theta}$ and this can be done by using (13) and the fact that $g_i(\cdot)$ and $f_j(\cdot)$ are continuous functions of $\boldsymbol{\theta}$.

The continuity of $R(\cdot)$ together with (12) and the fact that $\boldsymbol{\theta}_N{}^*(\omega) \to \boldsymbol{\theta}^0$ for almost all ω yields part (b) of the lemma. Q.E.D.

Remark. The matrix $R(\boldsymbol{\theta}^0)$ can be expressed as

$$R(\boldsymbol{\theta}^0) = \lim_{N\to\infty} (2/N)\, H_N{}^{\mathrm{T}} L_N H_N\,.$$

It was noted in the proof of Lemma 5 that this matrix is positive definite.

LEMMA 7. *For each $N = 1, 2 \ldots$ let C_N be a random $(m \times k)$ matrix and $\mathbf{X}_N$ a random k-vector. If $C_N \to^p C$, a degenerate random matrix, and $\mathbf{X}_N \to^d \mathbf{X}$ as $N \to \infty$, then $C_N\mathbf{X}_N \to^d C\mathbf{X}$ as $N \to \infty$.*

Proof. Since $\mathbf{X}_N \to^d \mathbf{X}$, $E(\exp[i\langle \mathbf{t}, \mathbf{X}_N\rangle]) \to E(\exp[i\langle \mathbf{t}, \mathbf{X}\rangle])$ for all $\mathbf{t} \in R^k$. Thus

$$E(\exp[i\langle \boldsymbol{\tau}, C\mathbf{X}_N\rangle]) = E(\exp[i\langle C^T\boldsymbol{\tau}, \mathbf{X}_N\rangle]) \to E(\exp[i\langle C^T\boldsymbol{\tau}, \mathbf{X}\rangle])$$
$$= E(\exp[i\langle \boldsymbol{\tau}, C\mathbf{X}\rangle])$$

for all $\boldsymbol{\tau} \in R^m$. By the continuity theorem for characteristic functions [8, p. 191], $C\mathbf{X}_N \to^d C\mathbf{X}$.

By [3, Corollary 1 of Theorem 5.1], $\| \mathbf{X}_N \| \to^d \| \mathbf{X} \|$. Also $\| C - C_N \| \to^p 0$ so by [3, Problem 1, p. 28] or [5, Theorem 4.4.6] it follows that $\| \mathbf{X}_N \| \cdot \| C - C_N \| \to^p 0$. This implies that $\| C\mathbf{X}_N - C_N\mathbf{X}_N \| \to^p 0$ since $\| C\mathbf{X}_N - C_N\mathbf{X}_N \| \leqslant \| \mathbf{X}_N \| \cdot \| C - C_N \|$. Therefore, by [3, Theorem 4.1], $C_N\mathbf{X}_N \to^d C\mathbf{X}$.

THEOREM 3. *$N^{1/2}(\hat{\boldsymbol{\theta}}_N - \boldsymbol{\theta}^0)$ is asymptotically normal.*

Proof. We observed above that $N^{1/2}(\hat{\boldsymbol{\theta}}_N - \boldsymbol{\theta}^0) = (\nabla_{\boldsymbol{\theta}}\boldsymbol{\phi}_N(\boldsymbol{\theta}_N^*)^{-1}(N^{1/2}\boldsymbol{\phi}_N(\boldsymbol{\theta}^0))$, for some $\boldsymbol{\theta}_N^*$ such that $\| \boldsymbol{\theta}^0 - \boldsymbol{\theta}_N^* \| \leqslant \| \boldsymbol{\theta}^0 - \boldsymbol{\theta}_N \|$. By Lemma 6, $\nabla_{\boldsymbol{\theta}}\boldsymbol{\phi}_N(\boldsymbol{\theta}_N^*) \to R(\boldsymbol{\theta}^0)$ almost surely so $(\nabla_{\boldsymbol{\theta}}\boldsymbol{\phi}_N(\boldsymbol{\theta}_N^*))^{-1} \to (R(\boldsymbol{\theta}^0))^{-1}$ almost surely. By Lemma 5, $N^{1/2}\boldsymbol{\phi}_N(\boldsymbol{\theta}^0) \to^d N(\mathbf{O}, M_0)$. Hence by Lemma 7, we have that $N^{1/2}(\hat{\boldsymbol{\theta}}_N - \boldsymbol{\theta}^0) \to^d N(\mathbf{O}, R^{-1}M_0R^{-1^T})$ where $R_0 = R(\boldsymbol{\theta}^0)$.

REFERENCES

1. M. AOKI AND P. C. YUE, On certain convergence questions in system identification, *SIAM J. Control* **8** (1970), 239–256.
2. K. T. ASTROM, T. BOHLIN, AND S. WENSMARK, "Automatic Construction of Linear Stochastic Dynamic Models for Stationary Industrial Processes with Random Disturbances using Operating Records," IBM Nordic Laboratory TP18.150, 1965.
3. P. BILLINGSLEY, "Convergence of Probability Measures," Wiley, New York, 1968.
4. J. R. BLUM, D. L. HANSON, AND L. H. KOOPMANS, On the strong law of large numbers for a class of stochastic processes, *Z. Warscheinlichkeitstheorie und Verw. Gebiete*, **2** (1963), 1–11.
5. K. L. CHUNG, "A Course in Probability Theory," 2nd ed. Academic Press, New York, 1974.
6. R. L. KASHYAP, Maximum likelihood identification of stochastic linear systems, *IEEE Trans. Automatic Control* **AC-15** (1970), 25–34.
7. M. LEVIN, Estimation of system pulse transfer function in the presence of noise, *IEEE Trans. Automatic Control* **AC-9** (1964), 239–235.
8. M. LOÈVE, "Probability Theory," Van Nostrand, New York, 1955.
9. H. B. MANN AND A. WALD, On the statistical treatment of linear stochastic difference equations, *Econometrica* **11** (1943), 173–220.
10. D. O. NORRIS AND L. E. SNYDER, Consistency of least squares estimates used in linear systems identification, *SIAM J. Control* **13** (1975), 1183–1193.
11. K. J. ASTROM AND P. EYKHOFF, System identification—a survey, *Automatica* **7** (1971), 123–162.

12. K. Y. Wong and E. Polak, Identification of linear discrete time systems using the instrument variable method, *IEEE Trans. Automatic Control* **AC-12** (1967), 707–718.
13. D. O. Norris and L. E. Snyder, "Statistical Properties of a Least Squares Identification Algorithm Used in Modeling Human Operator Performance," Aerospace Medical Research Laboratory Technical Report AMRL-TR-75-29, Wright–Patterson Air Force Base, Ohio, May 1975.