The Annals of Statistics
1975, Vol. 3, No. 5, 1109–1121

ASYMPTOTICALLY EFFICIENT ESTIMATORS FOR A CONSTANT REGRESSION WITH VECTOR-VALUED STATIONARY RESIDUALS

BY ROLF K. ADENSTEDT

Lehigh University

Estimation of linear functions of a vector parameter θ when an observed discrete- or continuous-time vector-valued stationary process has mean value $H\theta$, H a known matrix, is considered. Large-sample comparisons of best linear unbiased estimators and estimators based on the sample mean are made. Limits and rates of convergence of the variances of these estimators are obtained. It is shown that under general conditions there are asymptotically efficient estimators based on the sample mean, their form determined by the spectrum at the origin. Conditions under which all least squares estimators are asymptotically efficient are also given.

1. Introduction. Let $X(t)$ be an $n \times 1$ vector-valued wide-sense stationary random process, with possibly complex components, where the parameter t ranges over either the integers (referred to as the *discrete case*) or the real line (*continuous case*). The process has constant mean value vector $m = H\theta$, where H is a known $n \times p$ matrix of arbitrary rank r and θ is a $p \times 1$ vector of unknown regression coefficients. The $n \times n$ matrix-valued covariance function $R(t)$ of the process is assumed to be a (componentwise) continuous function in the continuous parameter case. All quantities are allowed to be complex.

Considered for this model is the estimation of linear combinations of the components of θ from the observations $X(t)$ for $t = 1, 2, \cdots, \tau$ in the discrete case and for $0 \leqq t \leqq \tau$ in the continuous case. Attention is restricted to unbiased estimators that are linear in the components of the observations. Of primary interest are large-sample ($\tau \to \infty$) comparisons of best linear unbiased estimators (BLUE's) with estimators based on the sample mean

$$\begin{aligned} \hat{m} &= \tau^{-1} \textstyle\sum_{t=1}^{\tau} X(t)\,, \quad \text{discrete case,} \\ &= \tau^{-1} \textstyle\int_0^{\tau} X(t)\, dt\,, \quad \text{continuous case,} \end{aligned}$$

which does not require knowledge of R in calculation.

Many authors have considered the fixed sample size regression model of a single observed random vector X with mean value $H\theta$. Such models arise in numerous ways; a prime example is in the regression approach to experimental design. In many situations, such as experimental design, the regression matrix H does not have full rank. This means that more parameters than can be

Received February 1974; revised February 1975.

AMS 1970 *subject classifications.* Primary 62M10; Secondary 62J05.

Key words and phrases. Linear estimation, regression, stationary vector-valued processes, asymptotic efficiency, spectral representations, Moore–Penrose pseudoinverse.

unbiasedly estimated have been introduced into the model. While a reparametrization may serve to transform the model to a full rank one, it disguises the role of the original parameters. Thus it is of interest to treat the case of nonfull rank H and to consider the estimation only of certain linear combinations of parameters.

One way in which the model being considered in this paper could arise is when repeated (and stationarily correlated) observations are made with a fixed sample size model as discussed above. Another situation that can be transformed to this model occurs when there are known linear constraints on the mean of an observed vector-valued stationary process. For example, let the n components $X_j(t)$ of $X(t)$ have means θ_j subject to $n - p$ linearly independent constraints of the form $a_{k1}\theta_1 + \cdots + a_{kn}\theta_n = 0$. Then some, though not necessarily every, set of p means may be treated as independent unknown parameters, with the remaining θ's expressed in terms of this set. If, say, $\theta_1, \cdots, \theta_p$ are the independent means then the model of this paper is obtained. Note that H has full rank in this example. As is shown in an example by Freiberger, Rosenblatt and Van Ness (1962), the least squares estimator (LSE) for the unknown part of the mean need not be asymptotically as good as the BLUE in the type of situation just described.

There is a substantial literature comparing LSE's with BLUE's in regression models. Zyskind (1967) and Kruskal (1968), among others, develop general conditions for coincidence of these estimators in the case of a fixed sample size. Watson (1967, 1972) also goes on to discuss the efficiency of LSE's relative to BLUE's.

Large-sample studies deal generally with finding conditions guaranteeing that LSE's are asymptotically efficient relative to the BLUE's for their expectations. For observations in discrete time, with stationary residuals, Grenander (1954) gives a thorough treatment for what has been called a stationary regression. Rosenblatt (1955) extends Grenander's results to the vector-valued case. It is to be noted that the constant regression $H\theta$ considered here is a special case in Rosenblatt's model only if time is discrete and if each component of θ appears in no more than one component of the mean value m. Various attempts have been made to obtain continuous parameter versions of Grenander's results; the most general of these is Kholevo (1969).

In the model here LSE's are generally not asymptotically efficient. When this is the case one typically tries to find other easily computed estimators that have optimal properties over as wide as possible a range of conditions. Vitale (1973) and Adenstedt (1974) prescribe such estimators for the mean of a stationary sequence. For the present model, under general conditions, it will be seen that asymptotically efficient estimators for linear functions of θ are to be found among linear combinations of components of the sample mean $\hat{m}$. As is to be expected, the form of the best combinations is determined by the spectrum at the origin.

In Section 2 the model is formalized and the notation is set. Section 3 contains the regularity conditions to be imposed on the spectrum of the process $X(t)$ and a statement of the main results, which are proved in Sections 4 through 8. Theorem 1 deals with limits of variances of estimators while Theorem 2 deals with the rate of convergence.

The main tools employed in this paper are spectral representations and adaptation of a Hilbert space viewpoint developed in Adenstedt and Eisenberg (1974). Also drawn on are fixed sample size regression results. To reduce length of exposition, the discrete and continuous cases are treated simultaneously as much as possible, with the notation directed toward this end. For conciseness, many results are presented in terms of the Moore–Penrose pseudoinverse. Another approach might have been reduction to a canonical full rank form of the model, as carried out by Zyskind (1967) in the fixed sample size case. Lemma 3 and the proof in Section 8 also indicate how the pseudoinverse might be avoided.

2. Definitions and notation. The random $n \times 1$ vectors $X(t)$ are regarded as defined on some measurable space $(\Omega, \mathscr{B})$. If μ_θ is a probability measure corresponding to the regression vector θ, then the model may be described by

$$\int_\Omega X(t)\, d\mu_\theta = H\theta$$

and

$$\int_\Omega [X(t+s) - H\theta][X(s) - H\theta]'\, d\mu_\theta = R(t) .$$

The prime always denotes Hermitian transposition. This description of the model as a process governed by one of a family of measures is statistically more accurate than the usual view of the model as the sum of a regression $H\theta$ and a residual $X(t) - H\theta$ (see the discussion in Adenstedt and Eisenberg (1974) on this point). When t is a continuous parameter the function $R(t)$ is assumed continuous, so that the components of $X(t)$ are continuous in mean under each θ. Thus the sample mean $\hat{m}$ exists as a mean square integral.

The measure μ_0 corresponding to $\theta = 0$ will play a special role in the sequel. E will denote expectation with respect to μ_0, while equalities between and claims of uniqueness for random variables are understood to be modulo μ_0. $L^2(\mu_0)$ means $L^2(\Omega, \mathscr{B}, \mu_0)$ while the *observation period* is $\{1, 2, \cdots, \tau\}$ in the discrete case and $[0, \tau]$ in the continuous case. Dependence on τ is usually suppressed and, unless otherwise indicated, *limits are taken as* $\tau \to \infty$.

As in Rao (1965), a *linear parametric function* has the form $\psi = b'\theta$ with b a constant $p \times 1$ vector. ψ is *estimable* iff b is in the column space $\mathscr{C}(H')$ of H'. A linear unbiased estimator (LUE) for an estimable $\psi = b'\theta$ is defined as any random variable that either has the form

$$\hat{\psi} = \sum_{\nu=1}^{k} w_\nu' X(t_\nu) , \qquad H' \sum w_\nu = b \tag{1}$$

for some $n \times 1$ vectors w_ν and parameter values t_ν in the observation period, or that is the limit in $L^2(\mu_0)$ of such forms. In the latter case the same sequence coverges in $L^2(\mu_\theta)$ for each θ, so this definition makes intuitive sense. See Kuk and Petunin (1973) for a detailed discussion of how LUE's are defined in this type of model. Since

$$\operatorname{Var} \hat{\psi} = \int |\hat{\psi} - b'\theta|^2 \, d\mu_\theta = E|\hat{\psi}|^2$$

does not depend on θ for the estimator (1), it is natural to regard $E|\hat{\psi}|^2$ as the variance of an arbitrary LUE.

If $\psi = b'\theta$ is estimable then there is an $n \times 1$ vector w with $H'w = b$, and $w'\hat{m}$ is a LUE for ψ. The primary interest in this paper is in comparing performance of $w'\hat{m}$ with the BLUE $\hat{\psi}_{\mathrm{BLU}}$, i.e., the (unique) minimum variance LUE for ψ.

The LSE $\hat{\theta}_{\mathrm{LS}}$ for θ is a value that minimizes

$$\begin{aligned}&\textstyle\sum_{t=1}^{\tau} [X(t) - H\theta]'[X(t) - H\theta] \\ &\qquad = \textstyle\sum_{t=1}^{\tau} [X(t) - \hat{m}]'[X(t) - \hat{m}] + \tau(\hat{m} - H\theta)'(\hat{m} - H\theta)\end{aligned}$$

in the discrete case and

$$\begin{aligned}&\textstyle\int_0^{\tau} [X(t) - H\theta]'[X(t) - H\theta] \, dt \\ &\qquad = \textstyle\int_0^{\tau} [X(t) - \hat{m}]'[X(t) - \hat{m}] \, dt + \tau(\hat{m} - H\theta)'(\hat{m} - H\theta)\end{aligned}$$

in the continuous case, thus minimizes $(\hat{m} - H\theta)'(\hat{m} - H\theta)$. The LSE for an estimable $\psi = b'\theta$ is then $\hat{\psi}_{\mathrm{LS}} = b'\hat{\theta}_{\mathrm{LS}}$. Implementation of the criterion leads, as in Watson (1967) to $\hat{\psi}_{\mathrm{LS}} = w'\hat{m}$ where w is in $\mathscr{C}(H)$ and $H'w = b$.

Results will be stated, and analysis carried out, in terms of well-known spectral representations. Thus

$$R(t) = \int e^{it\lambda} \, dF(\lambda) , \tag{2}$$

where the $n \times n$ matrix-valued spectral distribution function has Hermitian nonnegative (meaning positive semi-definite) increments $dF(\lambda)$. The integral in (2) is understood to be over $[-\pi, \pi]$ in the discrete case and over $(-\infty, \infty)$ in the continuous case, *a convention used throughout* for integrals with unspecified limits. Similarly

$$X(t) = \int e^{it\lambda} \, dZ(\lambda) ,$$

where each component of the random $n \times 1$ vector $Z(\lambda)$ is in the $L^2(\mu_0)$-linear span of the corresponding component of $X(t)$ and (cf. Rozanov (1967))

$$\begin{aligned} E\, dZ(\lambda)\, dZ(\lambda_1)' &= dF(\lambda) \qquad \text{for} \quad \lambda = \lambda_1 , \\ &= 0 \qquad\qquad \text{for} \quad \lambda \neq \lambda_1 . \end{aligned}$$

Some of the results are described in terms of Moore–Penrose pseudoinverses of matrices. This generalized inverse is descriptively defined for any rectangular matrix A by

$$A^+ = \lim_{\delta\to 0} (A'A + \delta^2 I)^{-1}A' = \lim_{\delta\to 0} A'(AA' + \delta^2 I)^{-1} ,$$

where I, as throughout, represents the identity matrix of suitable dimension. An alternate definition, useful for proving identities, states that A^+ is the unique matrix satisfying the *Penrose* (1955) *conditions*: $(AA^+)' = AA^+$, $(A^+A)' = A^+A$, $AA^+A = A$ and $A^+AA^+ = A^+$. A thorough reference on properties and applications of the pseudoinverse is Albert (1972). One pleasing use is in the possibility of explicitly representing the LSE of an estimable $\psi = b'\theta$, namely by $\hat{\psi}_{\mathrm{LS}} = b'H^+\hat{m}$. Among other easily proven facts used are: $(A')^+ = (A^+)' \equiv A'^+$; $A^+ = (A'A)^+A'$; AA^+, A^+A, $I - A^+A$ and $I - AA^+$ are the projection matrices respectively onto $\mathscr{C}(A)$, $\mathscr{C}(A')$, the null space $\mathscr{N}(A)$ of A and $\mathscr{N}(A')$; and for a projection matrix P, $P(AP)^+ = (AP)^+$ and $P(PA'AP)^+ = (PA'AP)^+P = (PA'AP)^+$.

3. Assumptions and results. As is well known, the $n \times n$ matrix-valued spectral measure $dF(\lambda)$ in (2) may be decomposed by

$$dF(\lambda) = dF_S(\lambda) + (2\pi)^{-1}f(\lambda)\,d\lambda\,.$$

Here the $n \times n$ matrix $f(\lambda)$ is Hermitian nonnegative for all λ while (every entry of) the spectral measure $dF_S(\lambda)$ is singular with respect to $d\lambda$. $dF_S(\lambda)$ encompasses both discrete and singular parts of the spectrum.

The following regularity conditions are imposed on the above decomposition:

(C1) $f(\lambda)$ is continuous and nonsingular at $\lambda = 0$;

(C2) $F_S(\lambda) - F_S(-\lambda)$ is constant $(= dF(0))$ in $0 < \lambda < \varepsilon$ for some $\varepsilon > 0$;

(C3) (discrete case) $f(\lambda)$ is nonsingular for almost all λ and $|q(\lambda)|^2 f(\lambda)^{-1}$ is integrable over $[-\pi, \pi]$ for some trigonometric polynomial $q(\lambda)$;

(C3) (continuous case) $f(\lambda)$ is nonsingular for almost all λ and $(1 + \lambda^2)^{-d}|q(\lambda)|^2 f(\lambda)^{-1}$ is integrable over $(-\infty, \infty)$ for some positive integer d and function of the form $q(\lambda) = \sum_{\nu=1}^{k} c_\nu e^{-it_\nu \lambda}$.

These assumptions are the matrix equivalents of similar assumptions made in Adenstedt and Eisenberg (1974). One may easily show that (C3) implies that $X(t)$ is a nondeterministic process. It should be noted that none of the conditions is required for Theorem 1 (below) and that all are used only in the proof of part (b) of Theorem 2.

When the spectrum contains a singular part, there may be estimable functions that cannot be consistently estimated. The first theorem concerns the limits of the variances of LUE's of interest.

THEOREM 1. (a) *For every $n \times 1$ vector w,*

$$\lim_{\tau\to\infty} \operatorname{Var} w'\hat{m} = w'\,dF(0)w \equiv \gamma(w)\,. \tag{3}$$

(b) *For any estimable $\psi = b'\theta$,*

$$\lim_{\tau\to\infty} \operatorname{Var} \hat{\psi}_{\mathrm{BLU}} = \inf\{\gamma(w) : H'w = b\} \equiv \gamma_b\,. \tag{4}$$

(c) *For every b in $\mathscr{C}(H')$ there exists a w satisfying $H'w = b$ and $\gamma(w) = \gamma_b$. For $H'w = b$, $\gamma(w) = \gamma_b$* iff $dF(0)w \in \mathscr{C}(H)$.

(d) $\lim \operatorname{Var} \hat{\psi}_{\mathrm{LS}} = \gamma_b$ *for every estimable $\psi = b'\theta$* iff $\mathscr{C}(H)$ *is invariant under* $dF(0)$.

It is shown in Section 4 that w satisfies $H'w = b$ and $\gamma(w) = \gamma_b$ iff it has the form

$$w = \{I - [P_1\, dF(0)P_1]^+\, dF(0)\}H'^+b + P_2 v\,, \tag{5}$$

where

$$P_1 = I - HH^+$$

is the projection onto $\mathscr{N}(H')$,

$$P_2 = P_1\{I - [P_1\, dF(0)P_1]^+ P_1\, dF(0)P_1\} \tag{6}$$

is the projection onto $\mathscr{N}(H') \cap \mathscr{N}(dF(0))$, and v is an arbitrary $n \times 1$ vector. Also

$$\gamma_b = b'H^+MH'^+b\,, \tag{7}$$

where

$$M = dF(0) - dF(0)[P_1\, dF(0)P_1]^+\, dF(0)\,. \tag{8}$$

The second result concerns the rate of convergence of estimator variances, which should be of interest whether or not the estimators are consistent.

THEOREM 2. (a) *For every $n \times 1$ vector w,*

$$\lim_{\tau\to\infty} \tau[\mathrm{Var}\, w'\hat{m} - \gamma(w)] = w'f(0)w \equiv \rho(w)\,. \tag{9}$$

(b) *For every estimable $\psi = b'\theta$,*

$$\lim_{\tau\to\infty} \tau(\mathrm{Var}\, \hat{\psi}_{\mathrm{BLU}} - \gamma_b) = \inf\{\rho(w) : H'w = b, \gamma(w) = \gamma_b\} \equiv \rho_b\,. \tag{10}$$

(c) *For every b in $\mathscr{C}(H')$ there is a unique w satisfying $H'w = b$, $\gamma(w) = \gamma_b$ and $\rho(w) = \rho_b$.*

(d) $\lim \tau(\mathrm{Var}\, \hat{\psi}_{\mathrm{LS}} - \gamma_b) = \rho_b$ *for all estimable $\psi = b'\theta$* iff *$\mathscr{C}(H)$ is invariant under both $dF(0)$ and $Nf(0)$, where*

$$N = I - S^+S$$

is the projection onto the null space of

$$S = dF(0) - M = dF(0)[P_1\, dF(0)P_1]^+\, dF(0)\,. \tag{11}$$

It seems natural to call the LUE $w'\hat{m}$ for $\psi = b'\theta$ *asymptotically efficient* (relative to $\hat{\psi}_{\mathrm{BLU}}$) if $\gamma(w) = \gamma_b$ and $\rho(w) = \rho_b$. This differs somewhat from the usual definition, that $\mathrm{Var}\, \hat{\psi}_{\mathrm{BLU}}/\mathrm{Var}\, w'\hat{m} \to 1$, but reduces to the same thing when $dF(0)$ is nonsingular or vanishes. It is shown that $\rho_b > 0$ for $b \neq 0$, with

$$\rho_b = b'\{H'[Nf(0)N]^+H\}^+b\,, \tag{12}$$

and that $w'\hat{m}$ is an asymptotically efficient LUE for $b'\theta$ iff

$$w = [Nf(0)N]^+H\{H'[Nf(0)N]^+H\}^+b\,. \tag{13}$$

The right side of (12) reduces to the familiar $b'[H'f(0)^{-1}H]^+b$ when $dF(0) = 0$.

Obviously the matrix S in (11), or more precisely its null space, plays an important role in the model. As seen from (8), S is a measure of the performance of the sample mean $\hat{m}$ for estimating the mean m, for $dF(0)$ is the limit of the covariance matrix of $\hat{m}$ while M is the limit covariance matrix of $\hat{m}_{\mathrm{BLU}}$.

The above results are proved in the remainder of the paper as follows: Theorems 1 and 2(a) in Section 4; Theorem 2(c) in Section 5, Theorems 2(b) and (d) respectively in Sections 7 and 8. Section 6 contains some lemmas needed in the proofs. All quantities so far defined will have the same meaning in the sequel.

4. Proof of Theorems 1 and 2(a). In the discrete case

$$\operatorname{Var} w'\hat{m} = w' \int \tau^{-2}|\textstyle\sum_{t=1}^{\tau} e^{it\lambda}|^2 \, dF(\lambda) w$$

and application of the dominated convergence theorem, with use of properties of Féjer's kernel, yields (3). By condition (C2),

$$\gamma(w) = w' \int_{|\lambda|<\varepsilon} \tau^{-2}|\textstyle\sum_{t=1}^{\tau} e^{it\lambda}|^2 \, dF_S(\lambda) w$$

for some $\varepsilon > 0$, so that

$$\begin{aligned}\tau[\operatorname{Var} w'\hat{m} - \gamma(w)] = {} & w' \int_{|\lambda|\geqq\varepsilon} \tau^{-1}|\textstyle\sum_{t=1}^{\tau} e^{it\lambda}|^2 \, dF_S(\lambda) w \\ & + w' \int (2\pi\tau)^{-1}|\textstyle\sum_{1}^{\tau} e^{it\lambda}|^2 f(\lambda)\, d\lambda w\,.\end{aligned}$$

(9) then follows by dominated convergence, properties of the Féjer kernel, and assumption (C1). The proof of (3) and (9) for the continuous case are identical to the above with $\sum_1^\tau e^{it\lambda}$ replaced by $\int_0^\tau e^{it\lambda}\,dt$.

Now $\gamma(w) = w'\,dF(0)w$ is just the variance of $w'Y$ in the fixed sample size regression model $Y = H\theta + dZ(0)$, so γ_b is the variance of the BLUE for $\psi = b'\theta$ in this model. Parts (c) and (d) of Theorem 1 are therefore simply restatements of results given by Zyskind (1967) in his Theorem 3 and Theorem 2 (part 8).

Albert (1972) in his Chapter VI shows that w satisfies $H'w = b$ and minimizes $\gamma(w)$ iff it has the form

$$w = [I - (VP_1)^+V]H'^+b + P_2 v\,,$$

where $V = dF(0)^{\frac{1}{2}}$ (the unique Hermitian nonnegative square root), $P_1 = I - HH^+$, $P_2 = P_1[I - (VP_1)^+VP_1]$, and v is an arbitrary $n \times 1$ vector. As the projections P_1 and $I - (VP_1)^+VP_1$ commute, P_2 is also a projection, namely onto $\mathscr{N}(H') \cap \mathscr{N}(VP_1) = \mathscr{N}(H') \cap \mathscr{N}(V)$. The representations given in (5) and (6) follow from the above since $(VP_1)^+ = (P_1V^2P_1)^+P_1V = (P_1V^2P_1)^+V$. Calculation of $\gamma(w)$ with w given by (5) also yields (7) after some use of pseudoinverse properties. The right side of (7) may be reduced to $b'[H'\,dF(0)^{-1}H]^+b$ when $dF(0)$ is nonsingular, but some facility with pseudoinverses is required to prove this.

The following almost obvious lemma is needed in the proof of part (b) of Theorem 1.

LEMMA 1. *Let* $b \in \mathscr{C}(H')$. *If* w *is given by* (5) *and* $H'w_1 = b$, *then* $w_1'\,dF(0)w = \gamma_b$.

PROOF. Since $dF(0)P_2 = 0$, it follows from (5) and (8) that $dF(0)w = Mw$. Also $MP_1 = dF(0)P_2 = 0$, whence

$$M = HH^+MHH^+ = HH^+MH'^+H' . \tag{14}$$

But $b = H'w = H'w_1$, so from (14) and (7)

$$w_1' \, dF(0)w = w_1'Mw = b'H^+MH'^+b = \gamma_b .$$

Let $\psi = b'\theta$ be estimable. Now (3) implies that $\limsup \operatorname{Var} \hat{\psi}_{\mathrm{BLU}} \leqq \gamma_b$, so to establish (4) it suffices to show that $\operatorname{Var} \hat{\psi}_{\mathrm{BLU}} \geqq \gamma_b$. In fact, it is enough to show that $\operatorname{Var} \hat{\psi} \geqq \gamma_b$ for any LUE of the form (1). Since this is obviously true if γ_b vanishes, take $\gamma_b > 0$. For w given by (5) and the random variable $\phi = w' \, dZ(0)/\gamma_b$ in $L^2(\mu_0)$ it is seen that $E|\phi|^2 = 1/\gamma_b$ and that $EX(t)\bar{\phi} = dF(0)w/\gamma_b$ for all t. From Lemma 1, therefore $E\hat{\psi}\bar{\phi} = 1$ when $\hat{\psi}$ is given by (1). The Schwartz inequality then yields $E|\hat{\psi}|^2 \geqq 1/E|\phi|^2 = \gamma_b$, as desired.

5. **Proof of Theorem 2(c).** The fundamental theorem of least squares states that, for a given vector z and matrix A of suitable dimensions, there exists a vector x_0 minimizing $||z - Ax||^2 = (z - Ax)'(z - Ax)$. Moreover, x_0 is unique up to an additive vector in $\mathscr{N}(A)$.

As noted, $H'w = b$ and $\gamma(w) = \gamma_b$ iff w has the form (5). Therefore

$$\rho_b = \inf_v ||W\{I - [P_1 \, dF(0)P_1]^+ \, dF(0)\}H'^+b + WP_2 v||^2 ,$$

where $W = f(0)^{\frac{1}{2}}$. The least squares theorem cited above implies that the infimum is attained for some v_0 unique up to an additive $y \in \mathscr{N}(WP_2)$. But W is nonsingular by (C1), so that $P_2 y = 0$ for such a y. It follows therefore that the w given by (5) with $v = v_0$ is the unique vector sought in Theorem 2(c).

While it is possible to write this "best" w in terms of pseudoinverses, the expressions are lengthy and reduction to (13) is tedious. Therefore proof of (12) and (13), in another way, is deferred to Section 7.

6. **Four lemmas.** The results proved in this section are needed to establish the remainder of Theorem 2.

LEMMA 2. (a) *The Hermitian matrix S in* (11) *is nonnegative.* (b) *For $H'w = b$, $\gamma(w) = \gamma_b$ iff $Sw = 0$.* (c) *$s \equiv \dim \mathscr{N}(S) \geqq r = \operatorname{rank} H$.* (d) *$\operatorname{rank} H'K = r$ for any $n \times s$ matrix K whose columns span $\mathscr{N}(S)$.*

PROOF. (a) Follows from (11) and the easily proven fact that Hermitian nonnegative matrices have Hermitian nonnegative pseudoinverses. (b) From (7) and (14), $w'Sw = \gamma(w) - \gamma_b$ when $H'w = b$. (c) Choose linearly independent vectors $b_1, \cdots, b_r$ in $\mathscr{C}(H')$ and then $n \times 1$ vectors $w_1, \cdots, w_r$ such that $H'w_j = b_j$ and $\gamma(w_j) = \gamma_{b_j}$ for all j. The w_j are in $\mathscr{N}(S)$ and are easily seen to be linearly independent. (d) Let $b_1, \cdots, b_r$ and $w_1, \cdots, w_r$ be as in the proof of (c). Then clearly $\operatorname{rank} H'K_1 = r$, where the columns of the $n \times s$ matrix K_1 include $w_1, \cdots, w_r$ and form a basis for $\mathscr{N}(S)$. The statement follows because $K = K_1 B$ for some nonsingular $s \times s$ matrix B.

LEMMA 3. *Let K be an $n \times s$ matrix whose columns form an orthonormal basis for $\mathscr{N}(S)$ and let Q be an $n \times r$ matrix whose columns form an orthonormal basis for $\mathscr{C}(H)$.* (a) *$H = QJ$ where the $r \times p$ matrix J has* rank *r and $JJ^+ = I = J'^+J'$.* (b) *$H^+ = J^+Q'$.* (c) *If A is a square matrix for which $Q'AQ$ is nonsingular then $(H'AH)^+ = J^+(Q'AQ)^{-1}J'^+$.* (d) *If $K'AK$ is nonsingular then $(NAN)^+ = K(K'AK)^{-1}K'$.* (e) rank *$Q'K = r$.*

PROOF. (a) Clearly $H = QJ$ where $J = Q'H$, and it is easy to see that rank $J = r$. Therefore $J^+ = J'(JJ')^{-1}$ and $JJ^+ = I$. (b) and (c) may be verified by the Penrose conditions with $Q'Q = I$, as may be (d) with $N = KK'$. (e) Follows obviously from Lemma 2(d).

LEMMA 4. *Let A_τ be a Hermitian positive definite matrix that converges to $[Nf(0)N]^+$ as $\tau \to \infty$. Then*

$$\lim_{\tau\to\infty} (H'A_\tau H)^+ = \{H'[Nf(0)N]^+H\}^+ .$$

PROOF. By (C1) and Lemma 3(e), both $K'f(0)K$ and $Q'K[K'f(0)K]^{-1}K'Q$ are nonsingular. K and Q here are as in the previous lemma. Therefore

$$[Nf(0)N]^+ = K[K'f(0)K]^{-1}K' \tag{15}$$

and

$$\{H'[Nf(0)N]^+H\}^+ = J^+\{Q'K[K'f(0)K]^{-1}K'Q\}^{-1}J'^+ \tag{16}$$

by Lemmas 3(c) and (d). Also

$$(H'A_\tau H)^+ = J^+(Q'A_\tau Q)^{-1}J'^+ , \tag{17}$$

and it is not difficult to see that the right side of (17) converges to the right side of (16).

LEMMA 5. $\lim_{\tau\to\infty} [f(0) + \tau S]^{-1} = [Nf(0)N]^+$.

PROOF. Note $f(0) + \tau S$ is nonsingular because $f(0)$ is. A direct proof may be based on diagonalizing S. More concisely, use of a perturbation theorem in Albert (1972), page 50, yields

$$[f(0) + \tau S]^{-1} = [f(0) + \tau S]^+ = [Pf(0)P]^+ + O(\tau^{-1})$$

as $\tau \to \infty$, where $P = I - (S^{\frac{1}{2}})^+S^{\frac{1}{2}} = N$.

7. Proof of Theorem 2(b). The proof of (10) will require a number of steps. At the same time (12) and the representation (13) of the "best" w will be established. In this section $b \in \mathscr{C}(H')$ and $\psi = b'\theta$ are regarded as fixed. For the present define

$$\beta_b = b'\{H'[Nf(0)N]^+H\}^+b ,$$

the quantity on the right of (12).

Let w be given by (13). With aid of (15), (16) and part (a) of Lemma 3 one obtains $H'w = b$. Also $Nw = w$, i.e., $Sw = 0$, so $\gamma(w) = \gamma_b$ by Lemma 2(b). A straightforward calculation shows that $\rho(w) = w'f(0)w = \beta_b$, so one can conclude that $\beta_b \geqq \rho_b$. Since Theorems 1 (c) and 2 (a) imply that

$$\limsup_{\tau\to\infty} \tau(\operatorname{Var} \hat{\phi}_{\mathrm{BLU}} - \gamma_b) \leqq \rho_b ,$$

the desiderata will follow from

$$\liminf_{\tau\to\infty} \tau(\operatorname{Var} \hat{\phi}_{\mathrm{BLU}} - \gamma_b) \geqq \beta_b , \tag{18}$$

Note that in the notation of Lemma 3, (12) can be written as

$$\rho_b = w'Q\{Q'K[K'f(0)K]^{-1}K'Q\}^{-1}Q'w \tag{19}$$

when $H'w = b$. This form vanishes iff $w \in \mathscr{N}(Q') = \mathscr{N}(H')$, hence iff $b = 0$.

Let $\hat{\phi}$ be the LUE given by (1). Using (7), (11), (14) and assumption (C2), one finds that

$$\textstyle\sum_{\nu,\mu} w_\nu' \int_{|\lambda|<\varepsilon} e^{i(t_\nu - t_\mu)\lambda}\, dF_S(\lambda) w_\mu = \sum_{\nu,\mu} w_\nu' S w_\mu + \gamma_b$$

for some $\varepsilon > 0$, and hence that

$$\operatorname{Var} \hat{\phi} - \gamma_b \geqq \textstyle\sum_{\nu,\mu} w_\nu'[S + (2\pi)^{-1} \int e^{i(t_\nu - t_\mu)\lambda} f(\lambda)\, d\lambda] w_\mu . \tag{20}$$

Used here is the fact that $\int_{|\lambda|\geqq\varepsilon} dF_S(\lambda)$ is nonnegative. The right side of (20) is just the variance of $\hat{\phi}$ calculated under the hypothesis that

$$\begin{aligned} F_S(\lambda) &= 0 \qquad \text{for} \quad \lambda < 0 , \\ &= S \qquad \text{for} \quad \lambda \geqq 0 . \end{aligned} \tag{21}$$

A little reflection now shows that, in order to prove (18), it suffices to assume (21) and prove that then

$$\liminf_{\tau\to\infty} \tau \operatorname{Var} \hat{\phi}_{\mathrm{BLU}} \geqq \beta_b . \tag{22}$$

Therefore let $F_S(\lambda)$ be given by (21) in the rest of this section. This is equivalent to assuming that $M = 0$, hence that all estimable functions are consistently estimable.

Suppose that $Y = Y_\tau$ is a random $p \times 1$ vector with components in $L^2(\mu_0)$ that satisfies

$$EX(t)Y' = H \tag{23}$$

for all t in the observation period. Then $E\hat{\phi}Y' = b'$ for every LUE of the form (1), hence for every general LUE. If $\sigma_\tau = EYY'$ is the covariance matrix of Y, then

$$0 \leqq E|\hat{\phi}_{\mathrm{BLU}} - b'\sigma_\tau^{+}Y|^2 = E|\hat{\phi}_{\mathrm{BLU}}|^2 - b'\sigma_\tau^{+}b ,$$

or $\operatorname{Var} \hat{\phi}_{\mathrm{BLU}} \geqq b'\sigma_\tau^{+}b$. If Y can be constructed so that $\sigma_\tau = \tau H'A_\tau H$, where the A_τ are Hermitian positive definite matrices that converge to $[Nf(0)N]^{+}$ as $\tau \to \infty$, then (22) will follow from Lemma 4.

Following Rozanov (1967), denote by $L^2(dF)$ the class of $1 \times n$ row vector-valued functions $u(\lambda)$ with $\int u(\lambda)\, dF(\lambda) u(\lambda)' < \infty$. For such a $u(\lambda)$, $\int u(\lambda)\, dZ(\lambda)$ exists and is in $L^2(\mu_0)$. The random vector Y referred to above will be constructed in the form $Y = H' \int U(\lambda)\, dZ(\lambda)$, where the $n \times n$ matrix function

$U(\lambda) = U_\tau(\lambda)$ has rows in $L^2(dF)$. (23) will certainly be satisfied if

$$\int e^{-it\lambda}U(\lambda)\,dF(\lambda) = U(0)S + (2\pi)^{-1}\int e^{-it\lambda}U(\lambda)f(\lambda)\,d\lambda = I \tag{24}$$

for all t in the observation period. Note (21) is used here. Also EYY' then has the required form $\tau H'A_\tau H$, where

$$\begin{aligned} A_\tau &= \tau^{-1}\int U(\lambda)\,dF(\lambda)U(\lambda)' \\ &= \tau^{-1}U(0)SU(0)' + (2\pi\tau)^{-1}\int U(\lambda)f(\lambda)U(\lambda)'\,d\lambda\,. \end{aligned} \tag{25}$$

It will be necessary to show that this expression is nonsingular and approaches $[Nf(0)N]^+$.

Because of assumption (C1), the matrix (25) will certainly be nonsingular if $U(\lambda)$ is continuous and nonsingular at $\lambda = 0$. This will easily be seen to be true for the $U(\lambda)$ displayed below.

In the following let $C_\tau = [f(0) + \tau S]^{-1}$. To exhibit $U(\lambda)$ the discrete and continuous cases must be considered separately. First consider the discrete case. There is no loss of generality in writing the trigonometric polynomial in (C3) as $q(\lambda) = \sum_{\nu=0}^{\alpha} c_\nu e^{-i\nu\lambda}$ and, because of (C1), in assuming that $q(0) \neq 0$. Now let

$$U(\lambda) = q(0)^{-1}q(\lambda)\sum_{t=1}^{\tau+\alpha} e^{it\lambda}C_{\tau+\alpha}f(0)f(\lambda)^{-1}\,.$$

Clearly this matrix has rows in $L^2(dF)$ and it satisfies (24) for $t = 1, 2, \cdots, \tau$, as direct calculation verifies. With this $U(\lambda)$ in (25),

$$\begin{aligned} \tau(\tau+\alpha)^{-1}A_\tau &= (\tau+\alpha)C_{\tau+\alpha}SC_{\tau+\alpha} \\ &\quad + C_{\tau+\alpha}f(0)\int K_{\tau+\alpha}(\lambda)\left|\frac{q(\lambda)}{q(0)}\right|^2 f(\lambda)^{-1}\,d\lambda\, f(0)C_{\tau+\alpha}\,, \end{aligned} \tag{26}$$

where $K_\tau(\lambda) = (2\pi\tau)^{-1}|\sum_{t=1}^{\tau} e^{it\lambda}|^2$ is Féjer's kernel. By assumption (C1), properties of Féjer's kernel, and Lemma 5 the second term on the right side of (26) approaches $[Nf(0)N]^+f(0)[Nf(0)N]^+ = [Nf(0)N]^+$. The first term on the right of (26) is identical to $C_{\tau+\alpha} - C_{\tau+\alpha}f(0)C_{\tau+\alpha}$ and approaches zero. Thus indeed $\lim A_\tau = [Nf(0)N]^+$.

The continuous case is similar. The constants t_ν in the function $q(\lambda) = \sum_{\nu=1}^{k} c_\nu e^{-it_\nu\lambda}$ may be taken nonnegative, with $q(0) \neq 0$. Defining $\alpha = \max\{t_1, \cdots, t_k\}$, let

$$U(\lambda) = q(0)^{-1}q(\lambda)g_{\tau+\alpha}(\lambda)(1+\lambda^2)^{-d}C_{\tau+\alpha+2d}f(0)f(\lambda)^{-1}\,,$$

where d is the integer in (C3) and

$$g_\tau(\lambda) = \int_0^\tau e^{it\lambda}\,dt + \sum_{\nu=0}^{d-1}\binom{d}{\nu+1}(i\lambda)^\nu[(-1)^\nu + e^{i\tau\lambda}]\,.$$

Clearly $U(\lambda)$ has rows in $L^2(dF)$. (24) for $0 \leqq t \leqq \tau$ now follows from the fact that

$$(2\pi)^{-1}\int e^{-it\lambda}g_\tau(\lambda)(1+\lambda^2)^{-d}\,d\lambda = 1\,, \quad 0 \leqq t \leqq \tau\,,$$

as shown in Adenstedt and Eisenberg (1974). These authors also show that

$$\lim_{\tau\to\infty}(2\pi\tau)^{-1}\int |g_\tau(\lambda)|^2(1+\lambda^2)^{-d}h(\lambda)\,d\lambda = h(0) \tag{27}$$

for any function $h(\lambda)$ that is continuous at $\lambda = 0$ and integrable over $(-\infty, \infty)$. (25) now becomes

$$A_\tau = \tau^{-1}(\tau + \alpha + 2d)^2 C_{\tau+\alpha+2d} S C_{\tau+\alpha+2d} + C_{\tau+\alpha+2d} f(0) \int L_\tau(\lambda) \left|\frac{q(\lambda)}{q(0)}\right|^2 f(\lambda)^{-1}\, d\lambda\, f(0) C_{\tau+\alpha+2d}\,,$$

where $L_\tau(\lambda) = (2\pi\tau)^{-1}|g_{\tau+\alpha}(\lambda)|^2(1 + \lambda^2)^{-2d}$. By an argument as for (26), with use of (27) and Lemma 5, it follows that $\lim A_\tau = [Nf(0)N]^+$.

8. Proof of Theorem 2(d). As noted in Section 2, $b'H^+\hat{m}$ is the LSE for an estimable $\psi = b'\theta$. Clearly $\tau(\operatorname{Var}\hat{\psi}_{\mathrm{LS}} - \gamma_b)$ converges iff $\operatorname{Var}\hat{\psi}_{\mathrm{LS}} \to \gamma_b$, and by Theorem 2(a) the limit is the $\rho(H'^+b)$. Because of Theorem 1(d) it suffices now to assume that $\mathscr{C}(H)$ is invariant under $dF(0)$ and then to show that $\rho(H'^+b) = \rho_b$ for every $b \in \mathscr{C}(H')$ iff $\mathscr{C}(H)$ is invariant under $Nf(0)$.

The set of all LSE's $\hat{\psi}_{\mathrm{LS}}$ coincides with the set of $w'HH^+\hat{m}$ for w an $n \times 1$ vector. Since $\mathscr{C}(H)$ is now assumed invariant under $dF(0)$, by Lemma 2(b) $SHH^+w = 0$ for every w, hence $SHH^+ = 0$, $SH = 0$, and $NH = H$. Thus invariance of $\mathscr{C}(H)$ under $Nf(0)$ is the same now as invariance under $Nf(0)N$.

In the notation and representations in Lemmas 3 and 4,

$$\rho(H'^+H'w) = w'HH^+Nf(0)NH'^+H'w = w'QQ'KK'f(0)KK'QQ'w\,.$$

Since $Q'Q = I$, this expression coincides with $\rho_{H'w}$, given by (19), for all w iff

$$Q'KK'f(0)KK'Q = \{Q'K[K'f(0)K]^{-1}K'Q\}^{-1}\,. \tag{28}$$

Consider now the full-rank fixed sample size regression model in which an observed random $s \times 1$ vector has mean value $K'Q\delta$, with δ an $r \times 1$ parameter, and covariance matrix $K'f(0)K$. The right side of (28) is just the covariance matrix of the BLUE $\hat{\delta}_{\mathrm{BLU}}$ for δ in this model while, since $Q'KK'Q = Q'NQ = I$, the left side is the covariance matrix of the LSE $\hat{\delta}_{\mathrm{LS}}$. Applying a result of Zyskind (1967), therefore (28) holds iff $\mathscr{C}(K'Q)$ is invariant under $K'f(0)K$. Since $N = KK'$ and $K'K = I$, the last is seen easily to be equivalent to invariance of $\mathscr{C}(NQ) = \mathscr{C}(Q) = \mathscr{C}(H)$ under $Nf(0)$.

REFERENCES

[1] ADENSTEDT, R. K. (1974). On large-sample estimation for the mean of a stationary random sequence. *Ann. Statist.* **2** 1095–1107.

[2] ADENSTEDT, R. K. and EISENBERG, B. (1974). Linear estimation of regression coefficients. *Quart. Appl. Math.* **32** 317–327.

[3] ALBERT, A. (1972). *Regression and the Moore-Penrose Pseudoinverse.* Academic Press, New York.

[4] FREIBERGER, W., ROSENBLATT, M. and VAN NESS, J. (1962). Regression analysis of vector-valued random processes. *J. Soc. Indust. Appl. Math.* **10** 89–102.

[5] GRENANDER, U. (1954). On the estimation of regression coefficients in the case of an autocorrelated disturbance. *Ann. Math. Statist.* **25** 252–272.

[6] KHOLEVO, A. S. (1969). On estimates of regression coefficients. *Theor. Probability Appl.* **14** 79-104.

[7] KRUSKAL, W. (1968). When are Gauss-Markov and least-squares estimators identical? A coordinate-free approach. *Ann. Math. Statist.* **39** 70-75.

[8] KUK, J. V. and PETUNIN, J. I. (1973). Observable linear estimates of the mathematical expectation of a random process. *Soviet Math. Dokl.* **14** 323-326.

[9] PENROSE, R. (1955). A generalized inverse for matrices. *Proc. Cambridge Philos. Soc.* **51** 406-413.

[10] RAO, C. R. (1965). *Linear Statisticat Inference and Its Applications.* Wiley, New York.

[11] ROSENBLATT, M. (1955). On the estimation of regression coefficients of a vector-valued time series with a stationary residual. *Ann. Math. Statist.* **26** 99-121.

[12] ROZANOV, Y. A. (1967). *Stationary Random Processes.* Holden-Day, San Francisco.

[13] VITALE, R. S. (1973). An asymtotically efficient estimate in time series analysis. *Quart. Appl. Math.* **30** 421-440.

[14] WATSON, G. S. (1967). Linear least squares regression. *Ann. Math. Statist.* **38** 1679-1699.

[15] WATSON, G. S. (1972). Prediction and the efficiency of least squares. *Biometrika* **59** 91-98.

[16] ZYSKIND, G. (1967). On canonical forms, nonnegative covariance matrices and best and simple least squares linear estimators in linear models. *Ann. Math. Statist.* **38** 1092-1109.

CENTER FOR THE APPLICATION OF MATHEMATICS
LEHIGH UNIVERSITY
BETHLEHEM, PENNSYLVANIA 18015